Abstract—In this paper, we develop a new cross-layer protocol, called as SmSCTP protocol, for session mobility over WLAN and 3G UMTS heterogeneous networks. Two key issues investigated in this work are the connection broken problem and fleeting-location-collapse problem. The connection broken problem indicates that a connection is broken when a mobile device is roaming to a different network. The connection broken problem incurs the long handoff delay time and cannot provide the seamlessly handoff result for session mobility. Traditional session mobility schemes, such as SIP (layer-7 solution) and MSCTP (layer-4 solution) protocols, only provide non-real time registration when third party user is calling up mobile user. The non-real time registration incurs the well-known fleeting-location-collapse problem and causes additional location update overhead. This paper presents a SIP-based MSCTP (SmSCTP) protocol which is a cross-layer, combination of layer-4 and layer-7, approach. The SmSCTP protocol utilizes the multi-homing mechanism to reduce hand-off delay time and to provide the more seamless handoff scheme. In our SmSCTP protocol, two new SIP signalings for our SmSCTP protocol are designed to simplify the initial handoff procedure and solve fleeting-location-collapse problem. Finally, simulation results are conducted to illustrate the performance achievements of our proposed SmSCTP protocol by imporving the signaling cost, the hand-off delay time.

I. INTRODUCTION

As the evolution of personal communication, many wireless network technologies coexist around us. The fourth-generation (4G)[16][17][18] personal communication service (PCS) is a heterogeneous network. Many kinds of wireless technologies are adopted to provide multimedia service through multiple mode mobile device. Difference function and difference capability wireless network technologies also push 4G wireless network to make a ubiquitous wireless environment. This paper considers a session mobility scheme in 4G environment which is formed with Universal Mobile Telecommunications System (UMTS)[1][2][3] and IEEE 802.11 wireless LAN (WLAN). In a heterogeneous network, a moving mobile device can pre-connect to serval favorite wireless networks. If the signal strength of the current wireless network is too weak and the service quality is not acceptable, the mobile device can switch to another appropriate wireless network automatically. This procedure is called handoff. In some literatures, the soft handoff can provide some advantageous, for instance, low handoff delay and low packet loss. In heterogeneous network, multiple connections can help the mobile device seamlessly handoff to appropriate wireless network.

In addition, SIP location update (REGISTER method) is executed after the session terminated. If the mobile device is handing off to other network, any message incoming from other node would be forwarded to the original IP address by the SIP server, because of the location information in database is incurrent. This phenomenon is termed as fleeting-location-collapse. This phenomenon also incurs location update overhead. In this paper, we propose a SmSCTP scheme that allows mobile device to quickly establish multiple connections before handoff occurs and uses new designed SIP method to solve the fleeting-location-collapse problem. Based on cross layer design, our algorithm uses two new SIP methods to control handoff procedure and uses MSCTP to provide soft handoff ability. Two new SIP methods also can reduce the signaling cost in SmSCTP scheme. Lower signaling cost can help system to serve more clients. By combining above advantages, SmSCTP scheme can provide a SIP session more flexibility, robustness and fast reaction. Simulation results show that the propose scheme can achieve low signaling cost, low hand-off delay time.

II. RELATED WORK

This section describes two popular solutions of session mobility. One is SIP re-INVITE mechanism and other is MSCTP. MSCTP protocol provides soft handoff architecture and works on layer-4. The SIP re-INVITE and MSCTP ASCONF are illustrated in the following.

A. SIP re-INVITE Mechanism

In 3GPP specification document TS 23.228 V7.0.0[7], SIP has been selected to be the main signaling protocol in UMTS IP multimedia core network subsystem (IMS), which can
support IP multimedia for provision multimedia service. A main component in IMS system, Call Session Control Function (CSCF), is defined to control call session. As well as SIP proxy server and registrar server which are defined in RFC-3261[8][9], the CSCF server executes call admission and forwards SIP signaling to appropriate destinations. Two kinds of user agent are defined, user agent server (UAS) and user agent client (UAC). The UAS accepts the SIP requests from the UAC and generates an accept, reject or re-direct response on the behalf of the user. Similar to the Hyper Text Transfer Protocol (HTTP), SIP is also a text-based protocol[10] and continues using the request-response model, and much of HTTP's syntax, header fields and semantics. In session mobility, SIP uses re-INVITE method to re-establish connection in new network environment. The re-INVITE message contains the new IP address of the new wireless network. After the re-INVITE procedure completes, the streaming data can be re-directed to new IP address. However, the UAC must register its new SIP URI to CSCF server when the call terminated. On the other hand, single IP architecture means that the SIP re-INVITE message need be delayed because of mobile node needs several seconds to obtain the new IP address. This long delay may break connection.

B. Mobile SCTP: MSCTP

A new non-mobility transport layer protocol called the Session Control Transport Protocol (SCTP)[11] recently has been developed to transport wide application data. Each SCTP end-point has one primary path and multiple backup paths. The concept of multi-homing of SCTP allows a transport layer connection to be defined between a set of local IP addresses and a set of remote IP addresses. A pair of local and remote IP addresses, such as \((IP_{local}, IP_{remote})\), can be set as the primary association. If the primary association crashes, the association will be seamlessly switched to another backup path. Multi-homing is a useful feature for a handoff process since a mobile station can prepare a new path before the breakdown of current path. In order to support mobility, mobile SCTP (MSCTP)[12][13][14][15] has been proposed which supports the Dynamic Address Reconfiguration that reconfigures IP address information on an existing association. Conceptually, dynamic address reconfiguration allows two end-points to add or delete IP address dynamically during handoff procedure. If association is crashed on the primary path, the \(ASCONF\_set\_primary\) message will be sent to the remote end-point for resetting parameter of the primary path. The heartbeat message is used to test the status of the original path. If no heartbeat\_ACK message received after three times (or more), the association handoffs to other backup path. However, the MSCTP mechanism also needs to register its new SIP URI after the call hang up because that the SIP can not know how the MSCTP processes the handoff procedure.

III. SYSTEM MODEL AND BASIC IDEA

This section first describes the system model. The following session describes the SmSCTP approach and compares it with SIP re-INVITE and MSCTP.

A. System Model

The Fig. 1 shows the heterogeneous wireless network architecture for SmSCTP scheme. The proposed architecture includes two kinds of wireless network, one is WLAN[19][20],
and other is UMTS. The dual mode UE can move from WLAN network to UMTS network and vice versa. The UMTS[6] is a third-generation mobile system, see Fig. 1, which extends 2G GSM and 2.5G General Packet Radio System (GPRS) to provide higher data transfer rate and more available application. In the 3GPP R99 specification [4][5], the UMTS network can support packet-switched (PS) service. The PS service can provide mobile device to get digital data service. The PS domain consists of two components to connect to external Packet Data Network (PDN) through Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN)[6], see Fig. 1(a). The user’s current position information is stored in SGSN so that an incoming data packet can be routed to user immediately. The GGSN plays a similar role as gateway which is between PDN and UMTS core network. In the mobility management part, the GGSN will tunnel to SGSN using GPRS Tunnel Protocol (GTP) protocol.

The CSCF, see fig. 1(c), is a SIP server. It has responsibility for interacting with HSS (Home Subscribe System) for mobility. The CSCF can be divided into three main components, Proxy-CSCF (P-CSCF), Serving-CSCF (S-CSCF) and Interrogating-CSCF (I-CSCF). The main function of the P-CSCF is to forward SIP messages to S-CSCF or I-CSCF, depending on the type of message or procedure. As SIP proxy server, the S-CSCF can forward SIP messages to appropriate node. The S-CSCF also supports registrar procedure. Similar to doorkeeper, the I-CSCF element supports charging, routing and forward SIP messages to appropriate node. In Fig. 1(c), we ignore I-CSCF for simple illustration. The WLAN uses WAG (Wireless Access Gateway; see Fig. 1(d))[15][16] element to interwork with UMTS network. The WAG connects more than one wireless access point and UE receives data from Internet or UMTS PS domain.

B. Basic Idea and Challenge

The SmSCTP scheme combines SIP protocol at layer 7 and MSCTP protocol at layer 4. Utilizing elasticity of SIP, the SmSCTP scheme uses new designed messages to notify CSCF and CN about the location information of the UE. In the CSCF server, new messages which come from the UE can make the CSCF to update the location information of the UE, and then forward messages to the CN. In the CN, new messages also can let CN to update location information and what new IP address should be switched to.

At layer-7, we design two new SIP messages for control session mobility. One is ANNEXIP and other is SWITCHSESSION. The ANNEXIP is used to notify the CSCF and the CN to add the location information of the UE. This message can be sent more than once if the UE gets multiple IP address. The function of SWITCHSESSION is to confirm the real location of the UE. The UE sends this signaling to the CSCF and the CN for informing the final IP address which the UE switches to. The lifetimes of ANNEXIP and SWITCHSESSION are the same as the call. At layer-4, the SIP directly orders MSCTP protocol to add IP address into IP pool or switch to other IP address. We cancel MSCTP ASCONF message between the UE and the CN. The Fig. 2 shows the simple model of SmSCTP. A moving UE scans and gets an existing network profile. In order to gets IP address, the UE executes first PDP content procedure in UMTS or attach in WLAN. Then, the ANNEXIP message is sent from UE to CN through CSCF server. The backup path will be established when CN gets the ANNEXIP message. Finally, the UE would depend on the trigger information to select a better network for transmission data. If UE determines switching to other network, it would send SWITCHSESSION message to CN.

The advantage of the signaling cost is displayed in Fig. 3 and Fig. 4. The SIP re-INVITE mechanism, see 3(a), costs nine messages in first cycle. Oppositely, the MSCTP (see Fig. 3(b)) and the SmSCTP (see Fig. 4) are based on multi-IP architecture. Address can be dynamically added or deleted. But, the MSCTP mechanism is based on the heartbeat message to decide the status of association is ACTIVE or INACTIVE. Finally, the SIP REGISTER message is also sent to the CSCF for update location information. Therefore, the MSCTP costs
eleven messages in first cycle. On the other hand, the SmSCTP uses SWITCHSESSION message to replace the ASCONF set primary path message and SIP REGISTER message. The SmSCTP adopts actively handoff mechanism. The SmSCTP costs eight signalings in first cycle.

IV. SmSCTP:SIP-BASED MSCTP PROTOCOL

This section describes the SmSCTP approach. Fig. 5 illustrates the three stages of the SmSCTP procedure. Some parameters are defined below.

- \( P_{\text{Lratio}} \): Packet loss ratio
- \( W_{\text{threshold}} \): Threshold in wireless location network
- \( S \): Sampling unit
- \( HO_{\text{threshold}} \): Handoff threshold

Fig. 5(a) initiates the PDP context procedure. In this stage, the UMTS resource is assigned to UE when PDP context finish. In this stage, the main path is \(<\text{UE} \xrightarrow{AP} \text{WAG} \xrightarrow{PDN} \text{CN}>\).

Step 1 The UE sends Activate PDP Context Request message to SGSN if the UE detects \( P_{\text{Lratio}} > W_{\text{threshold}} \). The SGSN negotiates with the RNS to allocate the resource for UE.

Step 2 The SGSN sends Create PDP Context Request message to the GGSN for establishing a GTP tunnel between the SGSN and the GGSN. At this moment, the GGSN allocates an IP address for the UE from DHCP server IP pool.

Step 3 The GGSN replies Create PDP Context response message to SGSN, and assigned UMTS IP address to the UE.

Step 4 The SGSN returns the Activate PDP Context Accept message which included the UMTS IP address to the UE. The UE adds assigned IP address into local IP pool.

Fig. 5(b) executes the ANNEXIP procedure. The function of this procedure is to inform the CN the secondary (or third) IP address of the UE. Conceptually, the backup path is established, but does not work. The data is transmitted precedentially through primary path.

Step 5 The UE sends the ANNEXIP signaling which included the UMTS IP address information to the CSCF server. The CSCF server sets the timer for this ANNEXIP signaling to maintain session state.

Step 6 The CSCF forwards the ANNEXIP signaling to the CN. Then, the CN commands SCTP protocol to add IP information in peer IP pool. At this moments, a new path, which is \(<\text{UE} \xrightarrow{RNS} \text{SGSN} \xrightarrow{GGSN} \text{PDG} \xrightarrow{PDN} \text{CN}>\), is established between the UE and the CN.

Step 7 The CN replies the 200 OK message to the CSCF if the CN finishes activation of adding IP. The CSCF cancels the timer.

Step 8 The CSCF forwards the 200 OK message to UE.

V. PERFORMANCE ANALYSIS

In this section, we evaluate the performance of SmSCTP through mathematical analysis and our developed testbed. In performance analysis, we focus on two factors, handoff latency and signaling cost. All of these factors will be analysed in following section.

A. Mathematical analysis

1) Signaling Cost Analysis: According to Fig. 3 and Fig. 4, we develop the formula of signaling cost. We focus on how many handoff times in one session. Some parameters are defined in following.

- \( CMR \): Call Mobility to Rate
- \( C_{\text{req}} \): Request cost
- \( C_{\text{res}} \): Response cost
Fig. 7. A simple model for delay analysis of SmSCTP.

- $C_{hb}$: Heartbeat cost
- $n$: Number of handoff times
- $\Psi_{SIP}$: Total signaling cost of SIP
- $\Psi_{MSCTP}$: Total signaling cost of MSCTP
- $\Psi_{SmSCTP}$: Total signaling cost of SmSCTP

The range of CMR is from 0.1 to $\infty$. The system model for SIP re-INVITE is shown in Fig. 3(a). The signaling cost for the SIP is described below:

$$\Psi_{SIP} \approx nCMR[5C_{req} + 4C_{res}]$$  \hspace{1cm} (1)

The MSCTP sends heartbeat message to make sure the network status. We assume heartbeat is sent three times. Thus, the signaling cost for MSCTP is:

$$\Psi_{MSCTP} \approx nCMR[7C_{req} + C_{res} + 3cost_{hb}]$$  \hspace{1cm} (2)

In SmSCTP scheme, see Fig. 3(c), the SWITCHSESSION can solve the fleeting-location-collapse problem. Therefore, we can omit the register cost. The formula is shown below:

$$\Psi_{SmSCTP} = 2CMR \ast (n + 1) \ast (C_{req} + C_{res})$$  \hspace{1cm} (3)

2) Handoff Latency: According to [7][9], we develop the model of SIP re-INVITE in Fig. 6(a). We also deliver MSCTP model in Fig. 6(b) based on [13][14]. The distance of between GGSN and UE is equal to between GGSN and CSCF. The network parameters are given following:

- $T_i$: Delay in every routing
- $T_{DHCP}$: Delay for obtains IP address
- $BW_{w,UMTS}$: BW of the UMTS access network
- $BW_{w,UMTS}$: BW of the UMTS core network
- $S$: Average packet size
- $L$: Link latency

The formula of $T_i$ is $T_i = \frac{S}{BW} + L$. Let $D_{PDP}$ be the time for PDP context activation procedure. Thus, $D_{PDP}$ is equal to $3T_1 + 3T_2 + T_{DHCP}$. Let $D_{SIP}$ be the delay time of SIP re-INVITE procedure.

$$D_{SIP} \approx nCMR[4T_1 + 4T_2 + 4T_3 + 3T_5 + D_{PDP}]$$  \hspace{1cm} (4)

For MSCTP, $D_{MSCTP}$ is

$$D_{MSCTP} \approx nCMR[4T_1 + 4T_2 + 4T_3]$$  \hspace{1cm} (5)

$$D_{MSCTP} = nCMR(2T_1 + 2T_2 + 4T_3)$$  \hspace{1cm} (6)

B. Simulation result

In this section we first describe the simulation topologies and configuration that have been used to compare the performance of SmSCTP, SIP and MSCTP. All of these are really implemented in our testbed system. In the 3G UMTS system, we use Chunghwa telecom (CHT) system in Taiwan. On the other hand, we use IEEE 802.11b network to simulate the WLAN network. In the UE, we use the Kphone from http://www.wirlab.net/kphone/ and it runs on Linux Fedora core 4. However, we modify the Kphone source code for supporting SmSCTP scheme. We also modify layer-4 SCTP socket and replace the UDP socket.

In our simulation, we are interest in the number of handoff during the call. The UE can move as its will. The CMR parameter is denoted the mobility frequency of the UE.

1) Signaling cost: Fig. 8 plots the impact call mobility ratio, which ranges from 0.5 to 1.5, on the signaling cost. Higher CMR value means the UE moves frequently.

In CMR = 0.5 and $n = 1$, the signaling cost of SmSCTP is lower than SIP re-INVITE and MSCTP. Because of SmSCTP just uses four signaling to setup handoff procedure and REGISTER procedure, but SIP should use nine and MSCTP should use eleven. In $n > 1$, the SmSCTP only uses two signaling to complete handoff procedure. Therefore, the slope of the SmSCTP is also lower than SIP and MSCTP. In CMR >= 1, the advantage of SmSCTP is more obvious. The implementation result and average CMR are presented in Fig.
8(d). The result shows that our implementation outcomes are very close the curve of the average CMR.

2) Handoff latency: Fig. 8 plots the impact call mobility ratio, which ranges from 0.5 to 1.5, on the delay time. We assume all of the moving directions of the UE are from WLAN to UMTS. In SIP re-INVITE procedure, because of the PDP context procedure is executed before handoff occurs and SIP needs more time to re-direct session to new location more than other schemes. In any CMR value, the delay time of SIP is always greater than MSCTP and SmSCTP. In MSCTP, the heartbeat message is used to check the session status. The exchange of some heartbeat messages lead to the session to idle a span. Finally, SmSCTP adopts the trigger information which from layer-2, so delay time can be controlled efficiently. Therefore, the degree of handoff latency is SIP > MSCTP > SmSCTP. The blue curve line show the trend of SmSCTP when CMR = 0.5. If the user moves infrequently (n < 3), the delay time of SmSCTP is unable to perceive. When n > 3, SmSCTP also can provide good handoff time, even CMR value is more than 1.

Fig. 9(d) shows the actual delay time on SIP, MSCTP and SmSCTP and average CMR. We assume the PDP context procedure needs 2 seconds. In general, the SmSCTP can complete the handoff procedure by using round-trip time of two signalings. In every handoff, thus, the SmSCTP can finish all of the procedure within about 200 milliseconds. The average CMR value can denote the characteristic of mobility of most human. Our implementation outcomes close the average CMR curve. This means our scheme is good for a heterogeneous network.

VI. CONCLUSION

In this paper, we introduce our cross layer scheme, SmSCTP, over WLAN and 3G UMTS heterogeneous network. Our SmSCTP scheme is a cross layer, layer-4 and layer-7, approach. Two new SIP messages are designed to process SmSCTP handoff procedure. Benefiting from these SIP signalings, the UE can ready backup path setup before handoff and the CSCF also can get the information which is the real location of the UE. Thus, the SmSCTP is not only supporting soft handoff mechanism but also solving fleeting-location-collapse problem. In the mathematical analysis, we verify that our SmSCTP scheme offers a better latency and signaling cost than SIP re-INVITE and MSCTP. Finally, the implementation result also illustrates that SmSCTP scheme actually achieves the performance improvements in handoff delay time.

REFERENCES