Application Layer Mobility Management Scheme for Wireless Internet

Ashutosh Dutta, Faramak Vakil, Jyh-Cheng Chen, Miriam Tauil
Telcordia Technologies, 445 South Street, Morristown, NJ 07960

Shinichi Baba, Nobuyasu Nakajima
Toashiba America Research, 445 South Street, Morristown, NJ 07960

Henning Schulzrinne
Computer Science Department, Columbia University, New York, NY 10027

Abstract—

This paper discusses an application layer framework that uses a set of standard IETF protocols for supporting real-time and non-real-time multimedia applications on mobile terminals of next generation wireless [3G/4G] networks. It describes a 3G-IP architecture and brings out the requirements and issues which are essential for supporting roaming users in the wireless Internet in a ubiquitous way. In addition it highlights ways of achieving end-to-end mobility by means of signaling instead of using any agent in the middle of the network as often done currently.

I. Introduction

Rapid growth of Internet and increased demand for ubiquitous wireless access are the driving forces behind intense activities towards the design of all IP wireless networks in different standard and technical forums. Several forums and standard bodies e.g., MWIF, 3GPP, and IETF are actively working on specifications of all IP wireless networks that allow roaming users to access integrated data, voice and multimedia services of the Internet via their wireless IP terminals and appliances. Furthermore, the existing circuit switched network and 1G/2G (i.e., 1G and 2G) wireless systems will eventually evolve and merge into an end-to-end IP platform that provides ubiquitous real-time as well as non-real-time services. In a nutshell, it is envisioned that an end-to-end wireless/wireline IP platform comprising 3G wireless access networks and a wireline IP backbone will support real-time and non-real-time multimedia services in the future.

In principle mobility management scheme in the wireless Internet provides means of terminal, session and service and personal mobility. Currently there are Mobile IP and many of its variants such as MIP with route optimization provide inter-domain mobility solution, on the other hand there are several other evolving approaches such as Cellular IP[8], HAWAII[7], IDMP[9] which provide intra-domain mobility by using source based routing or dynamic home agent. All these approaches use network layer solution while providing continuous media support when nodes move around and provide means of terminal mobility as well as dynamic address assignment using network access identifier (NAI). However Mobile IP by itself does not provide means of device independent personal mobility and location service. QoS requirements of real-time services are also not well defined. In addition to impact of trinagular routing which adds delay in the network in classical Mobile IP case, these require modification of the host IP protocol stacks for IP in IP encapsulation which also adds to the overhead for bandwidth constrained wireless link. [6] provides an end-to-end approach to host mobility but it does not address achieving mobility as part of end to end signaling, and does not address personal or service mobility.

On the other hand, Session Initiation Protocol (SIP) [2] is rapidly gaining wide-spread acceptance as the signaling protocol of choice for Internet multimedia and telephony services which can be used for Wireless Internet as well. SIP supports personal mobility as part of its signaling mechanism, its feature set can also be extended to provide adequate means of terminal, and service mobility. It has already been accepted as a means for signaling for session management in many of the forum and standard bodies such as MWIF, 3GPP who are designing the architecture for all IP wireless network. Because of some of the shortcomings associated with Mobile IP or its variants as outlined above, it may not be the best approach for providing mobility management in the wireless Internet.

One alternative is to use SIP only for personal mobility, mobile IP or one of its route-optimized variants for terminal mobility, and either extend mobile IP to support service mobility too or develop another protocol for service mobility. Another alternative as propsed in [1] is to use SIP‘s capability for real-time communication and use Mobile IP or its variant for non-real-time communication by using a policy table.

The proposed approach attempts to develop an end-to-end mobility management framework that exclusively uses SIP messaging/signaling to support terminal, sesion and service as well as personal mobility for both real-time and non-real-time communication. It utilizes as well as extends SIP to provide various means of roaming capabilities within an administrative domain and between the domains, so that users can access the network and obtain variety of services from any location using their own mobile terminal with minimum disruption. SIP works with other set of IETF protocols such as DHCP, AAA, Diffserv to provide different functionality such as address assignment, security, accounting and QoS. It also offers another advantage so that it can spoof constant endpoints which supports mobile TCP applications in a SIP environment without any changes to TCP.

Primary motivation for this proposed approach are mentioned below:

- It allows the users to depend upon their appliances rather than the network for supporting mobility on an end-to-end basis without reliance on and knowledge about abilities of network elements for packet interception and forwarding.

- It provides a means of route optimization and improved performance for real-time services via SIP signaling messages for address binding, registration etc.

- Allows dealing with mobility at a semantic level above IP terminals, by using appplication layer identifier such as SIP URI.

- It provides easier interaction with other IETF protocols such as DIAMETER, DHCP, SMTP, DNS.

II. A 3G-IP Architecture

A 3G-IP network comprises all IP wireless access networks and a packet switched IP backbone network. The IP backbone network is an end-to-end wire-line infrastructure that will comprise regional providers' wireline IP networks that are connected through the Internet, besides mobile stations/terminals, a wireless access network also consists of a radio access network (RAN), and an Edge Router and Controller. In order to support the needs of its users, a wireless access network interacts with the network control entities which are shown as Domain Control Agent. Figure 1 depicts the skeleton of one such 3G-IP network. Following are the description of some of the elements described there.

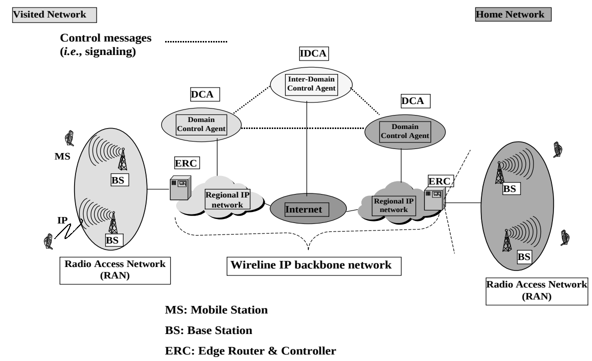

Fig. 1. A 3G Architecture

Mobile Station (MS) : It is the user mobile terminal that allows users to communicate, and also provides a means of interaction and control between users and the network.

Radio Access Network: The Radio Access Network represents the wireless and back-haul infrastructure that provides MSs with wireless access to the wireline infrastructure. In IMT-2000, RANs use programmable software radios to provide the flexibility across frequency bands at the MS and across RAN.

Edge Router Controller (ERC): An ERC is a routing and control system that connects a wireless access network to a regional wireline IP network. Each ERC may support several RANs. An ERC comprises two functional entities, an edge router (ER) and an edge control agent (ECA). The ER functions as an IP router , while the ECA is an intelligent agent that interacts with the Domain Control Agent (DCA) to control the RANs as well as support necessary network-wide control tasks. Control Agent functionality can be distributed.

Domain Control Agent: (DCA)

The domain control agent provides session management as well as the means of interaction between users and network control system and interaction among network control entities. DCA also supports 1) Mobility Management, 2) Authentication, Authorization and Accounting and 3) QoS management. It is assumed that each autonomous system would have components of domain control agent distributed within it.

Each domain control agent would interact with each other directly or via an Inter-Domain control agent using proper authentication mechanism.

III. Mobility Management Requirements and Issues

In general, mobility management scheme of wireless IP networks satisfies the following requirements:

1. It supports means of personal, service and terminal mobility, i.e., it allows users to access network services anywhere, as well as to continue their ongoing communication and to access network services anywhere using one's own mobile terminal.

2. It supports global roaming, i.e., it should remain independent of the underlying wireless technology. This would ensure that the mobility management scheme can be transported over all members of IMT-2000 family which comprises several wireless technologies such as W-CDMA, TDMA etc.

3. It supports both real-time and non-real-time multimedia services such as mobile telephony, mobile web access, and mobile data services in a way that their prices and performances are comparable. In order to achieve this mobility management scheme should interact effectively with the QoS management, and AAA schemes to verify the user's identities and rights, as well as to ensure that the QoS requirements and applications are satisfied and maintained as users roam around.

4. It transparently supports TCP based application. It should support TCP as is without requiring any changes to TCP or TCP based application

5. It should support multicast and anycast trees efficiently as mobile stations/users move around.

Mobility management scheme for the wireless IP network should take care of the following issues such as registration, configuration, dynamic binding, and location management on need basis.

A. Hand-off

Hand-off, often known as hand-over is a process that allows an established call/session to continue when an MS moves from one cell to another without interruptions in the call/session. This hand off process can be either hard or soft. In the hard hand-off case the mobile receives and accepts only one radio signal from a radio channel or base station within a single cell, as the mobile moves into a new cell, its signal is abruptly handed

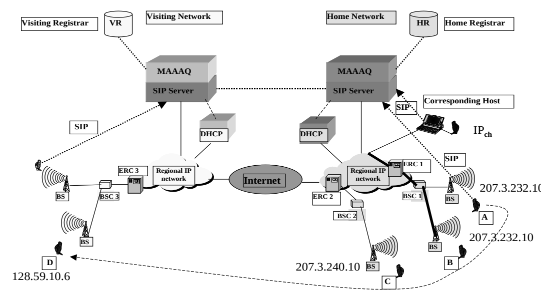

Fig. 2. Internet Roaming

over from its current cell (or base station) to the new one rapidly in few seconds. With soft hand-off [14] the MS continues to receive and accept radio signals from the base stations within its previous as well as its new cell for a limited period of time. In order to ensure the layer two independence requirement of mobility management scheme, a maximum acceptable hand-off time (MAHT) of 2-3 seconds is required. In end-to-end wireless IP paradigm, three logical levels of hand-off procedure can be defined:

i. Cell Hand-off : It allows an MS to move from a cell to another in a subnet within an administrative domain. One subnet may consist of multiple cells. IP address of the mobile host remains same in this case.

ii. Subnet hand-off : It allows an MS to move from a cell within a subnet to an adjacent cell within another subnet that belongs to the same administrative domain.

iii. Domain hand-off : It allows an MS to move from one subnet within an administrative domain to another in a different administrative domain.

The hand-off process is built upon the registration, configuration, dynamic address binding, and location management functions. Hand-off process is transparent to users and should satisfy the following requirements so that it would ensure the integrity, privacy, and confidentiality of user's location, perform the necessary AAA process to verify users' identities. It should ensure the service mobility as the MS roams around by making sure that it maintains the QoS of the ongoing sessions through minimizing the loss of transient data during the hand-off, as well as satisfying the delay requirements of real-time applications.

Figure 2 illustrates an Internet roaming as a mobile user keeps moving from its home base in the Internet, and performs different kinds of handoffs as it moves along. it also shows various functional elements which are part of the control agents within a domain. Each of these functionality can be distributed within the domain agent.

B. Registration

Registration is a process by which a network is made aware of the existence and location of an MS and its associated user. When an MS becomes active or roams into a network for the first-time, it would register with the network. This process comprises sending a registration request from the MS to the network and performing an AAA (Accounting, Authorization, Auditing) process by the network, and sending appropriate responses to the MS as well as location management entities to ensure that the network is aware of MS's current location. Depending upon the extent of registration, it can be categorized as complete or expedited/partial registration. Complete registration usually takes more time than the expedited registration. There are variants of AAA protocols to take care of security association between the mobile station and home AAA server when a client moves between the subnets within a domain. Home AAA server or an intermediate broker agent SIP Central Point of Contact[12] is contacted when the user moves into a new domain for the first time to establish the credentials. It is important to complete the registration process in a timely manner during the handoff process.

C. Configuration

Configuration is a process by which an MS updates its IP address as it roams between subnets either within the same administrative domain or in different administrative domains. As an MS moves between subnets, it needs to re-configure itself by acquiring new IP address, new default gateway, subnet mask. The key requirements of the re-configuration process include that, it should not take more than a few hundred milliseconds to complete and the DNS should get updated automatically to reflect the current name to address mapping and vice-versa only if the mobile is a server.

D. Dynamic Address Binding

Dynamic address binding is a process for allowing an MS to maintain a constant identifier regardless of its point of attachment to the network after the session is established. There are many existing solutions for providing dynamic address binding when MS's network attachment changes. It is essential to provide dynamic binding in such a way so that there will be minimum packet loss and less overhead associated with the media when the session is in progress during a hand-off. This dynamic binding can take place at the end-points or by some agents in the middle of the network. There are many ways of providing dynamic address binding, the proposed framework provides an elegant application layer approach for both real-time and non-real-time application which is defined later in the paper.

E. Location Management

Location management is a process by which the network updates the location database, and supports location/redirect services to authorized users and authorities. Location service is essential for new inbound sessions, before a session is established. It is essential for a user to be accessible independent of the device one is logged in.

IV. Mobility Tiers

Proposed mobility management scheme needs to provide means of terminal, session and service, and personal mobility. Each of these types of mobility is outlined below.

A. Terminal mobility

Multimedia traffic is categorized as real-time or non-real-time traffic and is mostly characterized by delay and loss factors. Different transport mechanism is used to carry each type of traffic separately. Most of the real-time traffic is carried over RTP/UDP where as non-real-time traffic is carried over TCP. Terminal mobility provide means of cell, subnet and domain hand-off while the session is in progress. In most cases it is taken care of by network layer mechanism. The proposed scheme provides a different approach to achieve terminal mobility by means of application layer signaling unlike traditional Mobile IP approach for both type of traffic.

B. Terminal Mobility for Real-Time application

For multimedia application which are typically RTP/UDP based, delay and loss are of primary concern, hence it is paramount to decrease the latency as much as possible. So it is advisable to avoid the triangular routing and any kind of encapsulation mechanism. [1] illustrates such a mechanism to provide mobility for real-time traffic.

Handoff is an important factor for terminal mobility for multimedia calls which would determine the latency. Upon every cell hand-off the DHCP client in MS may initiate reconfiguration process to obtain a new IP address from the DHCP server depending upon the detection mechanism as to if it is in the same subnet or a different one.

SIP is used to support subnet and domain handoff, while cell hand-off is taken care by the link layer. When the MS moves from one subnet to another within the same administrative domain, SIP would support subnet hand-off during the session as described below. It uses a combination of DHCP, and an AAA protocol (e.g., DIAMETER). This process can be summarized as follows.

- MS interacts with DHCP or one of its fast variants [11]to reconfigure itself, this process takes a round trip MS-DHCP-MS propagation delay.

- MS re-invites the corresponding host to its new temporary address. The identifier of the outbound proxy in the visited network should be inserted in the Record-Router field of this SIP INVITE messages.

- MS also initiates an expedited registration may local AAA protocol to update its location information in the home registrar.

- In case of domain hand-off a complete registration takes place.

Details of the call flows are not shown in this extended abstract version.

C. Non-Real-time application

The proposed framework also uses SIP signaling to provide means of terminal mobility for mobile TCP applications. In this proposed approach the mobile station is equipped with a SIP_eye agent that maintains a record of its ongoing TCP connections, upon hand-off the MS sends information to the corresponding host using INFO [13]method. Within the body of the INFO method it sends the binding message, requesting binding of the MS's old address to its new one. MS and CH use IP encapsulation maintain constant endpoints for MS's ongoing TCP connections. This is needed since long lived TCP connections would die if the destination address changes in the middle of a continuing session.

The SIP_eye is either integrated with the SIP UA or kept as a separate agent that interacts with the SIP UA as necessary. SIP_eye operates as follows:It examines headers of TCP packets to monitor the birth and death of TCP connections as well as identify their end-points. It maintains a current record of Mobile's ongoing TCP connections identifiers within the MS. SIP_eye records a state comprising 3 integers, (original MS IP address, current MS IP address, original CH IP address), per TCP connection. The original IP address is the IP address at the beginning of the TCP session. This approach can also be used to forward any transiet data from a router.

The CH and the MS use IP encapsulation either within IP or minimal to forward the TCP information to each other.

D. Service Mobility

Service mobility refers to the end user's ability to maintain ongoing sessions and obtain services in a transparent manner regardless of the end user's point of attachment. The service mobility includes the ability of the service home provider to either maintain control of the services it provides to the user in the visited network or transfer their control to the visited network. In order to support the service mobility one strives to maintain the QoS of ongoing sessions as the user (MS) roams around, and ensure that the MS has access to all of its subscribed network services and features regardless of its point of attachment. The QoS can be maintained through appropriate resource allocation during the hand-off and SIP registration [5]] along with the necessary AAA functions can be used to ensure users access to subscribed services.

E. Personal Mobility

Personal mobility refers to the ability of end users to originate and receive calls and access the subscribed network services on any terminal in any location in a transparent manner, and the ability of the network to identify end users as they move across administrative domains. This is achieved by personal mobility feature inherent in SIP. SIP's URI scheme and registration mechanism [5] are some of the main components which would provide personal mobility. A roaming user can be accessible independent of the device one is logged on. It works in conjunction with DNS and other database servers such as LDAP, and whois. As proposed in [14] mobile servers could use dynamic DNS so as to be able to be availabe to the client as it roams.

V. Architecture Realization

With a view to realizing different protocols and components needed for supporting mobile streaming multimedia such as wireless IP telephony, and delivering multimedia streaming content in a 3G wireless IP environment, a test-bed has been designed and implemented where the proof of concept for different entity of the said architecture has been demonstrated. Figure 3 shows the testbed architecture where different functionality of application layer mobility management scheme such as registration, signaling, binding, security and QoS features have been prototyped including different types of mobility such as micro, macro and domain mobility. This testbed can also be used to provide support for billing, network management and roaming across different domains. Here we would present some initial measurements from SIP based terminal mobility for real-time traffic.

As part of the experiment, SIP based terminal mobility for real-time traffic was tried using a modified and customized version of Columbia University's sipc (SIP client) and sipd (SIP server). Following provides some performance measurement during subnet mobility. A Multimedia SIP session for real-time traffic was established between two clients in a IEEE 802.11 environment . The mobile host moves away from the CH while the session is in progress. As soon as the mobile host discovers that it is in a new subnet it gets a new IP address using DRCP [11] a light-weight version of DHCP, sends a Re-Invite to the corressponding host and sends a new registration message to the registrar. Packet sizes for different messages are also noted. For initial call set up a typical INVITE message was 455 UDP bytes, ringing was 223 bytes, OK was 381 bytes, ACK was 216 bytes, REGISTER and its OK messages were about 370 bytes and 412 bytes. Subsequent de-registration and re-registration messages were of 372 and 425 bytes respectively followed by OK messages which were of size 510 and 410 bytes. A typical Re-Invite after subnet change and respective OK messages were 450 and 380 UDP bytes respectively. It takes about 100 msec for the processing time on Linux based MH between consecutive messages (e.g., between receiving an OK and sending the ACK which is strongly OS dependent). It took about 5 msec to forward the INVITE packet to traverse between MH and CH when MH is at home. As the MH moves away to a foreign network with more network routers in between, Re-INVITE takes about 70 msec because of the queueing delay at the routers enroute. The SIP registrar was placed in the middle of the network between the two domains, it took about 150 msec to complete the whole re-registration process, however it takes more processing time to delete the old registration than updating a new one at the registrar. These figures would vary depending upon the operating system, and network topology. It would be desirable to reduce the Re-Invite time, re-registration time in a wireless roaming environment for faster handoff and avoiding loss of data.

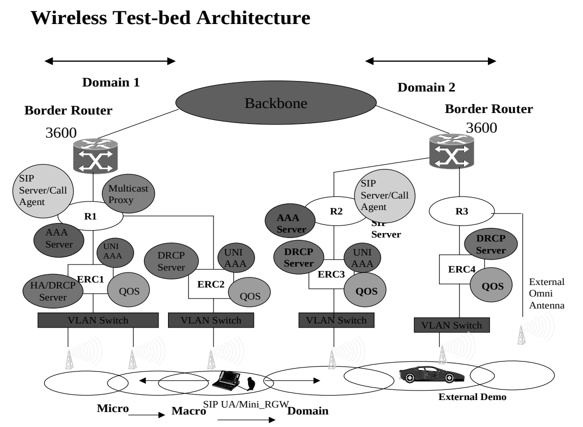

Fig. 3. Snapshot of the testbed

VI. Conclusion and Open Issues

This paper outlines the requirements and issues involved with mobility management in a 3G-wireless Internet environment. It describes an application layer solution to variety of mobility that would be needed to provide ubiquitous access to a roaming user by using SIP signaling scheme. This approach highlights some of the issues involved with application layer mobility support in an Internet roaming environment.The current architecture with all the functional components has been realized in a test-bed emulating a wireless Internet in a wavelan environment. Simulations and experiments are being carried out to measure the efficiency of SIP based mobility scheme compared to other mobility management schemes currently in practice, some of it have been presented here.

VII. Acknowledgement

Authors would like to acknowledge other members of the ITSUMO (Internet Technologies Supporting Universal Mobile Operations) team for their effort in realizing different components of application layer mobility management architecture. Authors would also like to acknowledge Xiaotao Wu and Jonathan Lennox for answering several questions towards sipc and sipd respectively.

References

[1] Elin Wedlund, H. Schulzrinne , " Mobility Suport using SIP" in *IEEE/ACM Multimedia conference WOWMOM 1999*

[2] Mark Handley, Eve Schooler, H. Schulzrinne, Jonathan Rosenberg, "Session Initiation Protocol", RFC 2543

[3] Faramak Vakil, A. Dutta, J.C Chen, S. Baba, Y. Sobatake, Henning Schulzrinne"Mobility Requirements in SIP environment", draft-itsumo-sip-mobility-req-02.txt, IETF Work in progress

[4] F. Vakil et. al, "Host Mobility Management Protocol for 3G network ". Internet Draft Work in progress

[5] Henning Schulzrinne, "SIP registration", draft-schulzrinne-sip-register-00.txt, IETF Draft Work in Progress

[6] Alex C. Snoeren and Hari Balakrishnan"An End-to-End Approach to Host Mobility". Proceedings of Mobicom 2000

[7] R. Ramjee et al "IP Micro-mobility support using HAWAII", draft-ietf-mobileip-hawaii-01.txt, IETF, July 2000, work in Progress

[8] A. Campbell, J. Gomez et al,"Cellular IP" draft-valko-cellularip-00.txt, IETF WOrk in Progress

[9] Archan Misra, S. Das, A.Dutta, Anthony Mcauley, "IDMP: An Intra-domain Mobility Management Protocol using mobility agents", draft-ietf-misra-idmp-00.txt, IETF, July 2000 work in progress

[10] C. Perkins, "IP mobility support", RFC 2002, IETF October 1996"

[11] A. Mcauley, S. Das, S. Baba et al, "Dynamic Registration and Configuration Protocol (DRCP)" draft-itsumo-drcp-01.txt IETF July 2000 work in progress

[12] Pat Calhoun et al, "AAA requirements for IP Telephony/Multimedia" SIP WG, IETF, Work in Progress

[13] Steve Donavan, "INFO method" SIP WG, IETF, work in progress

[14] Yi-Bing Lin, Al-Chun Pang, "Comparing Soft and Hard Handoffs" IEEE transcations on Vehicular Technology

[15] J.C Chen, S. Baba et al. "QoS Architecture based on Differentiated Services for Next Generation Network." IETF Work in Progress.

[16] www.mwif.org

[17] www.3gpp.org

[18] www.ietf.org