Mobile SCTP
 Transport Layer Mobility Management for the Internet

Maximilian Riegel
Siemens AG, Munich, Germany
E-mail: maximilian.riegel@icn.siemens.de

Dr. Michael Tüxen
Siemens AG, Munich, Germany
E-mail: michael.tuexen@icn.siemens.de

Abstract: **Transport layer mobility management is proposed as an alternative to network layer methods like Mobile IP for realising seamless mobility in the Internet. Mobility management in the transport layer is purely accomplished by use of SCTP (Stream Control Transmission Protocol) and some of its currently proposed extensions. With Mobile SCTP seamless hand over even crossing physical, link and network layer technologies is fully realised inside the mobile client at the end of the network not needing any additional provisions in the network. Mobile services can be provided to such clients by mobile SCTP enabled servers from everywhere in the Internet to anywhere in the Internet.**

Client mobility management based on mobile SCTP seems not to require any new protocol development. It is a particular application of SCTP eventually solving the requirements of mobility in the Internet.

INTRODUCTION

Traditionally mobility in networks is provided by particular functions in the core network allowing terminals to be moved from one place to another steadily being reachable by the same address. The translation between the static assigned address identifying the terminal in a global scope and a temporary address representing the current location of the terminal in the network is accomplished by functions belonging to the network layer of communication networks, the Internet as well as 2G/3G mobile networks.

NETWORK LAYER MOBILITY

There are commonalities and architectural similarities between mobility management of 2G/3G mobile networks and Mobile IP, which is up to now aimed to be the mobility solution for the Internet.

Both solutions are based on a functional unit at the home location taking care of the current position of the mobile terminal and forwarding all the information aimed for the mobile terminal to the location where the terminal currently stays. To allow a terminal to move from one place in the network to another while maintaining reachability by the same global address two different network addresses are used for each terminal; one address specifying the current location in the network, the other uniquely identifying of the terminal.

Mobility Management for 2G/3G Mobile Networks

In 2G/3G Mobile Networks circuit switched applications (telephony) and packet switched applications are handled by different network elements in the core network with a common Radio Access Network (RAN).

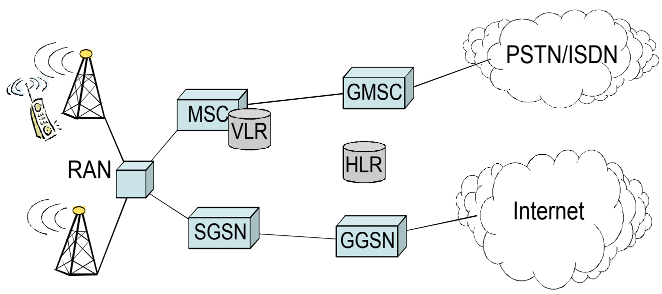

Figure 1 - Architecture of 2G/3G mobile networks

Packet data from the Internet or other public data networks is forwarded from the GGSN (Gateway GPRS Serving Node) via SGSN (Serving GPRS Serving Node) to the RAN and thereby to the mobile terminal. Circuit switched calls coming from the PSTN are routed by the GMSC (Gateway Mobile service Switching Center) via the MSC (Mobile service Switching Center) of the RAN the mobile terminal stays in to the terminal. The association between the Mobile Station ISDN (MSISDN) number providing the identity of the communication endpoint and the Mobile Subscriber Roaming Number (MSRN) specifying the current location of the terminal is maintained in a hierarchical, distributed database consisting of a centralised Home Location Register

(HLR) and a distributed Visited Location Register (VLR).

To establish a connection to the mobile terminal the GGSN/GMSC retrieves from the HLR/VLR the current MSRN of the mobile terminal and the address of the responsible SGSN/MSC taking care of the particular MSRN. With this information the GGSN/GMSC can establish via the responsible SGSN/MSC the communication path.

During movement of the mobile terminal the established communication path is switch over to other base stations and even other MSCs/SGSNs when becoming necessary together with updates of the information stored in the VLR as well as the HLR.

Mobile IP

Mobility in the Internet is accomplished by making sure the moving host is reachable by its originally assigned IP address even when the address leaves the network area the address belongs to. To keep reach ability by an address outside its assigned area the protocol Mobile IP [3| can be used installing an agent in the home area taking care of all packets sent to a mobile host staying outside its native network area.

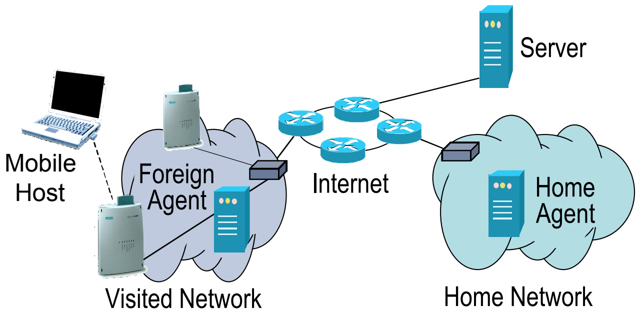

Figure 2 - Architecture of Mobile IP

The home agent knows about the foreign location of the mobile host and forwards all packets addressed to it to an agent in the foreign location which finally delivers the packets to the mobile host. Home agent and foreign agent are connected by a tunnel making the mobility enabled network layer 'circuit switched'. The foreign agent may be collocated with the mobile host, reducing the requirements of visited networks.

Architectural considerations for the Internet

Mobile IP as well as 2G/3G mobile networks are maintaining databases in the core network where the associations between identity address and location address are stored for each mobile host/terminal and are using circuit-like tunnels for forwarding the data to the related position in the network. While storage and maintenance of ‘connection states’ is quite natural for traditional switched networks, the Internet does not know about connection states in the network. Thus Mobile IP does not fully harmonise with the nature of the Internet.

TRANSPORT LAYER MOBILITY

Transport layer mobility management keeps the circuit-less nature of the network layer of the Internet untouched and implements the whole functionality for providing mobility to hosts in the transport layer entities at both ends of the network. The transport layer of the Internet is the first layer going up the networking stack which provides end-to-end control.

Transport Layer functions

A client host accesses a particular service over the Internet by establishing a transport layer connection to a server host providing such service. This connection is mostly made reliable by an appropriate transport control protocol and carries the application protocol elements and all the user data of a particular service between the hosts over the Internet. Applications needing a reliable transport may use the Transmission Control Protocol (TCP) which provides a reliable, duplicate free and in-sequence delivery of user data.

The transport layer makes use of transport layer addresses. For the Transmission Control Protocol this is a pair consisting on an IP-address and a port number. A TCP connection is established between two TCP endpoints, each of the TCP endpoints being identified by one transport layer address of TCP. It is important that the two TCP endpoints of a TCP connection can not change during the lifetime of a TCP connection.

When a host changes its IP address, for example by attaching to a new network, existing TCP connections can not use this new address because the TCP endpoint can not be changed.

This is one of the reasons why today mobile IP is used to provide the mobile host with a constant IP address which is used for communication with the peer.

Transport protocols supporting multihoming

A host is called multihomed if it has multiple network layer addresses. In case of IP networks this means that the host has multiple IP-addresses. This does not necessarily means that the host has also multiple link layer interfaces. Multiple IP addresses can be configured on a single link layer interface.

A transport protocol supports multihoming if the endpoints can have more than one transport layer addresses addresses. These transport layer addresses are considered as logically different paths of the peer towards the endpoint having the multiple transport addresses.

One of the examples of a transport protocol which supports multihoming is the Stream Control Transmission

Protocol currently being defined in [4]. An SCTP transport address is a pair of an IP-address and a port number like in the case of TCP. But a SCTP endpoint can be identified by a sequence of SCTP transport addresses all sharing the same port number.

Mobility enabled transport protocols

Transport layer protocols allowing the modification of the endpoints during the lifetime of a connection are called mobility enabled transport protocols.

A mobility enabled transport protocol allows for the change of the IP address of the network layer while keeping the end-to-end connection intact. If the transport protocol supports multihoming and the host can attach to multiple networks the transport protocol can make use of the simultaneous connection.

SCTP with the extension described in [2] is an example of a mobility enabled transport protocol supporting multi-homing. Another protocol currently being discussed in the IETF is the Datagram Congestion Control Protocol (DCCP) is a mobility enabled transport protocol not supporting multihoming.

Transport layer mobility

The mobility enabled transport layer shields the application not only from the actual network beneath and provides virtual circuits end to end through the Internet but also hides the change of underlying network addresses. Most application protocols, except those using IP addresses in messages of their own will continue to work when being ported to an mobility enabled transport layer.

Since the mobility is now handled by the endpoints which reside in the hosts and not in the network the transport layer mobility connect harmonises with fully with the nature of the Internet.

TRANSPORT LAYER MOBILITY BY EXAMPLE

Figure 3 illustrates the concept of transport layer mobility. This example is based on a mobility enabled transport protocol supporting multihoming.

Initial connection to the Internet

A mobile client connects to the Internet by some wireless technology and gets assigned an IP address from the local address space at location A [e.g. 2.0.0.2]. This can be accomplished by any of the techniques currently known for dynamic address assignment, like PPP or DHCP.

The mobile client being now reachable over the Internet establishes a transport layer connection to a server anywhere 'in' the Internet [e.g. 8.0.0.4] and starts using the provided service.

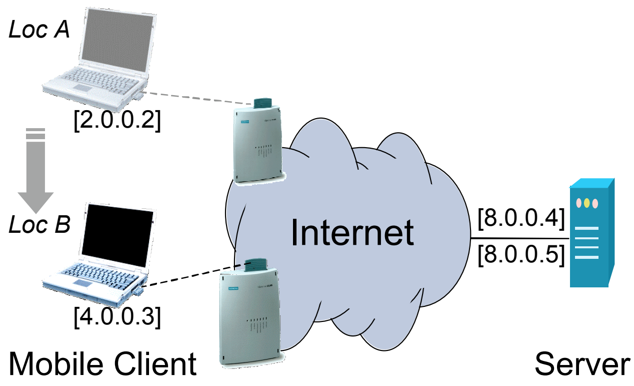

Figure 3 - Transport layer mobility

Soft hand over

The mobile client moves from location A towards location B and gets knowledge of reaching the coverage of another network by information from the physical layer of its NIC. In addition to the link already existing the mobile client establishes a link to the network at location B and gets assigned an IP address of the network at location B on its second network interface. Thus the mobile client becomes multihomed and is now reachable by the way of two different networks.

The mobile client tells the corresponding server using the established transport layer connection that it is now reachable by at a second IP address. Technically speaking, it adds the newly assigned IP address to the association identifying the connection to the server. To enable easy distinction of the two links at the mobile client several IP addresses should be assigned to the network interface of the server. This allows to represent different links by different entries in the routing table of the mobile client.

By reaching location B the mobile client may leave the coverage of the access point at location A and may loose the link for the first IP address of the mobile client. The data stream between server and mobile client gets interrupted and the reliability behaviour of the transport protocol ensures that all data is sent over the second link in case of permanent failure of the first link.

If the mobile client has access to information about the strength of the wireless signal the hand over to the second link will be initiated before severe packet loss occurs, making the hand over more soft.

Tear down of the initial link

When the mobile client has proved by information from the physical layer that the failure of the first link is permanent, it will inform its peer that it is now not reachable any more by the first IP address and removes this IP address from the association.

The procedure continues...

When the mobile client moves on, it may reach the coverage of another wireless network. It will repeat the procedure described above gaining seamless mobility while keeping running applications working.

MOBILE SCTP,
THE MOBILITY ENABLED PROFILE OF SCTP

It is not necessary to develop a new protocol for transport layer mobility. All the functionality described above providing transport layer mobility already exists in SCTP [4] when the proposed extension for dynamic addition of IP addresses is added as being described in [2].

A further extension, currently being described in [7], to the SCTP protocol provides the service of partial reliability. This makes SCTP also applicable for applications where the application data is not valid indefinitely.

Mobility enabling functions of SCTP

The thrill using SCTP for mobility is due to two features provided by SCTP:

- an end-point can use multiple IP-addresses for one connection
- an end-point can change these multiple addresses dynamically without affecting the established association.

An association is a connection between two SCTP endpoints.

Support of multi homing

An SCTP end-point can use multiple IP-addresses for an association. These are exchanged during the initiation of the association. The multiple addresses of the peer are considered as different paths towards that peer.

This means that a server must use multiple IP-addresses to provide the mobile client with multiple paths. These will be used while moving between locations.

It should be mentioned that this path-concept is used only for redundancy, not for load sharing. Therefore one path is used for normal transmission of user data. It is called the primary path.

Dynamic addition and deletion of IP-addresses

Using the multi homing feature provides multiple paths through the network. This does not solve the problem of having different IP-addresses at different locations. Therefore the IP addresses of an SCTP end-point have to be modified without affecting an established association. It is also helpful to tell the peer what path to use as the primary.

An extension to SCTP providing exactly these features is described in [2].

CONSIDERATIONS FOR
MOBILE SCTP ENABLED HOSTS

The only general requirement is that the transport protocol must be SCTP with the extensions described in [2].

Requirements for SCTP mobility enabled mobile clients

To motivate the requirements for the mobile client one has to consider the situation where the client has connections to multiple access points. The following figure shows this scenario with two access points.

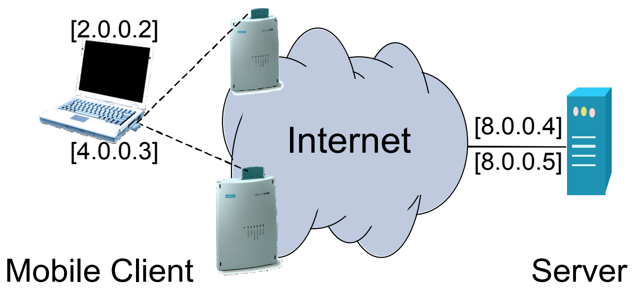

Figure 4 - Multihomed mobile client

During the time where the mobile client is reachable via two different access networks it has to make sure that it uses both links. Thus, for example, the forwarding of the mobile client has to be set up in a way that the traffic towards 8.0.0.4 uses the upper link (interface 2.0.0.2) and the traffic towards 8.0.0.5 uses the lower link (interface 4.0.0.3).

The mobile client also knows the quality of the two links and can make sure that it uses the better one whenever appropriate. Using the ability to request the server to modify its primary path it is also possible that the mobile client makes sure that the traffic from the server towards the mobile client uses the better link.

It should be mentioned that this link hand over has to be done carefully to avoid oscillation and frequent switching.

Summarising this, the mobile client must use an implementation of SCTP with the extension ADDIP. Furthermore the forwarding table of the mobile client has to be modified according the connectivity state.

Requirements for SCTP mobility supporting servers

From the discussion in the previous section it is clear that the server has also to support SCTP with the ADDIP extension. Additionally it must be multihomed but it is not necessary that the server uses multiple link layer interfaces. This is one of the examples where having multiple IP addresses on one interface makes sense.

COMBINATION OF LINK LAYER MOBILITY AND TRANSPORT LAYER MOBILITY

Some radio technologies like IEEE802.11 wireless LAN provide mobility management functionalities in the link layer. Link layer hand over is mostly restricted to micro mobility but can be favourably combined with transport layer mobility management reducing the processing requirements at the server side for handling all the hand overs.

CONCLUSION

This paper compared the the concepts of network layer mobility and transport layer mobility. It was shown that transport layer mobility is an alternative to Mobile IP for the Mobile Internet considerably reducing the complexity of the network by realising the mobility functions within the hosts at the end of the networks. Thus transport layer mobility harmonises well with the architectural principles of the Internet.

A detailed example was given describing how transport layer mobility works. Finally one example of a mobility enabled transport protocol was given.

Future work needs to address also the mobility of the servers and the mobility in a peer-to-peer network. Also the rules when to add/delete transport layer addresses to/from the endpoints must be studied.

REFERENCES

[1] Ong, L. and J. Yoakum, "An Introduction to the SCTP", draft-ong-sigtran-sctpover-02 (work in progress), January 2002.

[2] Stewart, R., "Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration", draft-ietf-tsvwg-addip-sctp- 05 (work in progress), May 2002.

[3] Perkins, C., "IP Mobility Support", RFC 2002, October 1996.

[4] Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L. and V. Paxson, "Stream Control Transmission Protocol", RFC 2960, October 2000.

[5] Tuexen, M., Xie, Q., Stewart, R., Shore, M., Ong, L., Loughney, J. and M. Stillman, "Requirements for Reliable Server Pooling", RFC 3237, January 2002.

[6] Coene, L., "Stream Control Transmission Protocol Applicability Statement", RFC 3257, April 2002.

[7] Stewart, R., Ramalho, M., Xie, Q., Tuexen M., Conrad, P, “SCTP Partial reliability Extension”, draft-stewart-prsctp-00.txt (work in progress), May 2002.