Performance evaluation of SIP based handover in heterogeneous access networks

ROK LIBNIK*, ALES SVIGELJ**, GORAZD KANDUS**
* Network and Services Development Department, Telekom Slovenije d.d.
*Cigaletova 15, 1000 Ljubljana
** Department of Communication Systems, Jozef Stefan Institute
**Jamova 39, 1000 Ljubljana
SLOVENIA
rok.libnik@telekom.si ales.svigelj@ijs.si gorazd.kandus@ijs.si http://www.ijs.si/

Abstract: Wireless technologies evolve very fast in last years. In the future operators will need to enable users to use communication's services independent from access technologies, thus they will need to support handovers between different networks. In this paper we presented simulation results for handover from WLAN network to HSDPA network and vice versa. We used IP telephony application with SIP signalization. We proposed new approach, which is more applicable to real operator environment as assures proper billing during handover. The handover was executed using simple decision function, based on received signal measurement. To achieve more credible results we used characteristics of real operator's networks. Simulation results show that both handovers i.e. from WLAN to HSDPA network and from HSDPA to WLAN network have some limitations which can affect the quality of service especially for real-time application, like VoIP. With result we also shoved that characteristics of today's networks are not good enough to offer seamless handover using SIP protocol and that mobile packet wireless networks will need to improve to offer lower latencies.

Key-Words: Heterogeneous networks, SIP, Seamless handover, Mobility management, Performance evaluation, Real-time applications

1 Introduction

In the filed of wireless communications we are witnessing nowadays rapid development of new services. The most popular wireless technologies today are wireless LAN, which is used within buildings and GSM/UMTS outside the buildings. High adoption of WLAN networks happened due to the low cost of equipment and the ease of deployment in home/business environment. Furthermore, there are also other technologies which evolve in today's market like WiMAX, HSPA (High Speed Packet Access) family and pre-4G. The number of mobile phone users is increasing very fast, while the usage of wire line telephony is decreasing. Thus, the equipment manufactures started to offer dual mode handsets; while operators are starting to offer fixed mobile converged services. Some of them (BT as the first [1]) already offer services which enable users to handover seamlessly from GSM to WLAN network. However, user demands are increasing very fast and the operators will need to offer seamless handover between different wireless access technologies. Operators will need to decide which protocol they will use for handover. This can be very challenging as some modifications may need to be done to their existing infrastructure.

1.1 Overview of mobility management

Mobility management techniques are used to support user movements within the same and between different networks. In general, handover can be performed in three phases, which are (i) network discovery, (ii) handover decision and (iii) handover execution.

Network discovery is the process in which mobile terminal finds/detects new network. To detect a new network as fast as possible all network interfaces need to be enabled all the time. This can have huge impact on energy consumption (i.e. draining out batteries fast). By selecting the proper time intervals, in which network interfaces are active, the power consumption can be reduced, resulting in later discovery of the new network.

In the second phase the decision for handover is made. This phase represents the most important and critical part of handover process. If the decision is not made at appropriate time, based on improper inputs users can notice high degradation of service level.

In the last phase handover of traffic to new network is done. It means that all the traffic uses the new connection and the connection with an old network is terminated. It is worth noting that all three phases contribute to overall handover time, which is of paramount importance for the user experience.

From the network architecture point of view handover can be performed at different OSI layers.

IP protocol, which is the mostly used protocol at network layer, defines that a node's IP address uniquely identifies the node's point of attachment to the network. A node can not change its point of attachment without losing its ability to communicate. A new, scalable, mechanism is required for accommodating node mobility within the network. In [2] mobile IP (MIP) protocol is defined. The document specifies protocol enhancements that allow transparent routing of IP datagrams to mobile nodes in the Internet. Each mobile node is always identified by its home address, regardless of its current point of attachment to the Internet. While situated away from its home, a mobile node is also associated with a care-of address, which provides information about its current point of attachment to the Internet. The protocol provides for registering the care-of address with a home agent. The home agent sends datagrams destined for the mobile node through a tunnel to the care-of address. After arriving at the end of the tunnel, each datagram is then delivered to the mobile node. As described above MIP enables nodes to change their point of attachment to the Internet without changing their IP address. In certain cases, the latency involved in handover can be above the threshold required for the support of real-time services. To shorten the handover delay authors in [3] proposed two methods called Post Registration handover technique and Pre Registration handover technique. The described techniques allow greater support for real-time services on a Mobile IPv4 network by minimizing the period of time when a Mobile Node is unable to send or receive IPv4 packets due to the delay in the Mobile IPv4 Registration process.

The most known protocols on transport layer are TCP and UDP. Both of them have some limitations for offering real-time services and mobility. To overcome those limitations new protocol called SCTP (Stream Control Transmission Protocol) was introduced [4]. SCTP can be viewed as a layer between the SCTP user application and a connectionless packet network service such as IP. The SCTP enables endpoint that during association start up provides the other endpoint with a list of transport addresses (i.e., multiple IP addresses in combination with an SCTP port). Through selected transport address the endpoint will receive and originate SCTP packets. However the SCTP protocol is not capable of changing IP address when the session already started. Thus, new solution called mSCTP (mobile SCTP) was developed, which enables to add, delete and change the IP addresses during active SCTP association [5]. The SCTP with the ADDIP extension (or mSCTP) would provide seamless handover for the mobile host without support of routers or agents in the networks. For location management, the mSCTP could be used along with Mobile IP, SIP or Reliable Server Pooling.

The SIP protocol was proposed by IETF as a general multimedia signaling protocol which enables peers to set up voice or any multimedia sessions between them [6]. SIP is an application layer protocol and runs on top of several different transport protocols and is today's widely used protocol for IP telephony.

With minor modifications it can support four types of mobility:

- Terminal mobility: enables devices that can move between subnets and being reachable to other hosts and continuing any ongoing session when they move.
- Session mobility: enables users that can maintain a session while moving from one terminal to another.
- Personal mobility: enables users that can use the same set of services even when the user is changing devices or network attachment points.
- Service mobility: enables users to be identified by the same logical address, even if the user is at different terminals.

For terminal mobility management two approaches were defined (i) pre-call mobility and (ii) mid-call mobility.

In order to compare approaches (i.e. MIP, mSCTP and SIP) on different OSI layers, we summarized main characteristics in Table 1. We selected four parameters:

- Impact on network: If operator decides to support mobility management, some modifications of the network may need to be done in terms of new network elements or additional functionality.
- Impact on applications: Supporting handover can have effect on applications which may need to be customized.
- Network involvement: For handover execution some resources needs to be allocated. This parameter describes how big network involvement in handover resource allocation is.
- Adoption in real operator environment: The possibility of adoption of protocols in real

Layer	Impact on network	Impact on application	Network involvement	Adoption in real operator environment
Network layer (MIP)	Big (home/foreign agents need to be installed)	None	High	Medium
Transport layer (mSCTP)	Little or none (MNs must support mSCTP	None	Low (just MNs)	Little or none
Application layer (SIP)	Little or none (SIP proxy is needed)	Big	Medium (only proxy)	High

Table 1: Comparison of mobility approaches at different OSI layers

operator environment due to complexity of the solution and/or price.

When using MIP for handover new elements such are home and foreign agents need to be installed in the network. Because MIP is network layer protocol which is used just for transportation the application which is in use is not aware of the handover process, however network involvement in handover execution is big as handover is executed in the network it self. The limitation of using MIP for handover is that it is not so widely used protocol and only few operators have implemented it to their network.

Performing handover on transport layer has little impact on the network. If MN supports mSCTP then operators do not need to change their network. If handover is executed on transport layer again similar to network layer, application is not aware of the handover execution and can be used as it is. The biggest drawback of the mSCTP protocol is that is rarely used in real operator environment.

The advantage of using SIP protocol for handover execution is that SIP is application layer protocol and thus agnostic to lower layers. On the other hand application usually needs to be improved to support handover. SIPs transport independence does not require high network involvement, but the biggest advantage of SIP protocol is high adoption in real operators environments as almost all operators that are offering VoIP services uses SIP for signalization.

As it can be seen all approaches have some advantages and drawbacks. Thus, authors in [7] suggest combining MIP and SIP, while authors in [8] are suggesting that mobility management should be done based on application which is in use. The combinations are always possible. However, our main aim is to focus on solutions, which can be easily deployed in real operator’s environment. It is a fact that in most of today’s networks IP protocol is used and that SIP protocol was selected as primary signaling protocol in IMS (IP Multimedia Subsystem) networks. Thus, we decided to focus on mobility management using SIP. To achieve more credible results we used characteristics of real operator networks.

2 SIP mobility

SIP protocol is an application layer signaling protocol for Internet multimedia session establishment, modification, and termination. These sessions include Internet telephone calls, multimedia distribution, and multimedia conferences. SIP invitations used to create sessions carry session descriptions that allow participants to agree on a set of compatible media types. SIP makes use of elements called proxy servers to help route requests to the user's current location, authenticate and authorize users for services, implement provider call-routing policies, and provide features to users. SIP also provides a registration function that allows users to upload their current locations for use by proxy servers.

The SIP framework consists of the following main network elements [6]:

- User Agent Client (UAC): A user agent client is a logical entity that creates a new request, and then uses the client transaction state machinery to send it.
- User Agent Server (UAS): A user agent server is a logical entity that generates a response to a SIP request.
- User Agent (UA): A logical entity that can act as both a user agent client and user agent server.
- Proxy Server: An intermediary entity that acts as both a server and a client for the purpose of making requests on behalf of other clients.
- Registrar: A registrar is a server that accepts register requests and places the information it receives in those requests into the location service for the domain it handles.

In this paper we focused on terminal mobility for which two types of mobility management approaches were defined (i) pre-call mobility and (ii) mid-call mobility.

In the following sections both of them are presented shortly.

2.1 Pre-call mobility

The scenario of pre-call mobility is shown in Fig. 1.

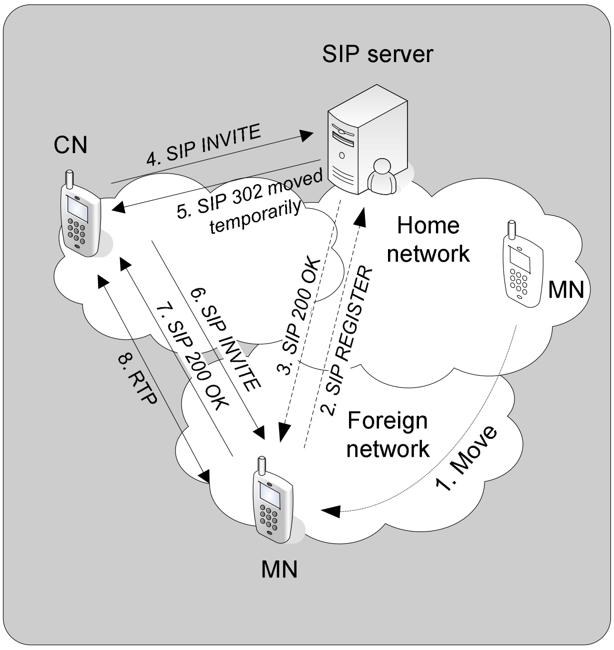

Fig. 1: Pre-call mobility

When a mobile node (MN) moves to another network (step 1) makes new registration at the SIP server (steps 2, 3). When the correspondent node (CN) makes a call to a MN, SIP INVITE message is first send to the SIP server (step 4), which informs the CN that the MN has changed the location (step 5). CN then sends the SIP INVITE message to the MN (steps 6, 7) and the RTP session is established (step 8).

When the MN moves to another network gets the new IP address. In the session description of SIP REGISTER message, the MN informs the SIP server about the new IP address. Usually the MN gets new IP address form the DHCP server, which is located in the network. The message exchange is presented in Fig. 2.

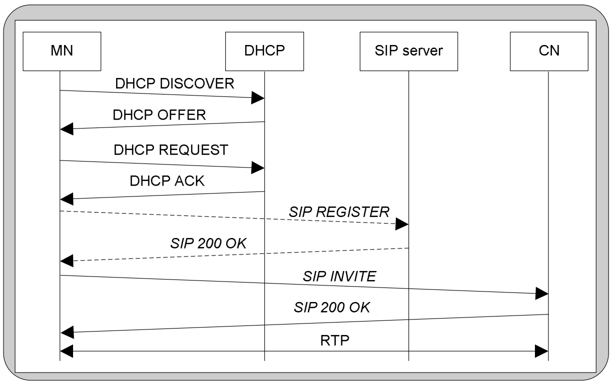

Fig. 2: Message exchange – pre-call mobility

Because the new IP address acquisition is done prior to the call this operation does not affect the quality of service. The only issue of pre-call mobility scenario is new IP address acquisition. This can be solved by a process in which the MN makes re-registration (step 2, 3 in Fig. 1) every few seconds.

2.2 Mid-call mobility

In the mid-call mobility scenario first a call is established between the CN and the MN, which is in the home network (steps 1, 2, 3, 4, 5). When MN moves to another network (step 6), sends the SIP re-INVITE message to CN (steps 7, 8) and informs it about the location change. The new RTP session is then established (step 9). The mid-call scenario is presented in Fig. 3.

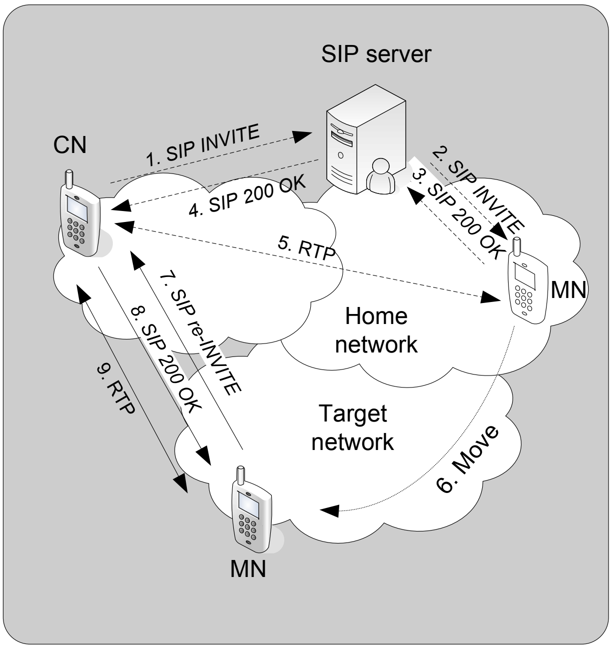

Fig. 3: Mid-call mobility

Similar to the pre-call scenario, the MN gets new IP address in the new network. In the SIP re-INVITE message the session description includes also the new IP address.

In Fig. 4 message exchange is presented. The new IP acquisition needs to be done in the middle of the call. This process increases the overall delay. Another limitation of this approach is that SIP server is not informed about the location change.

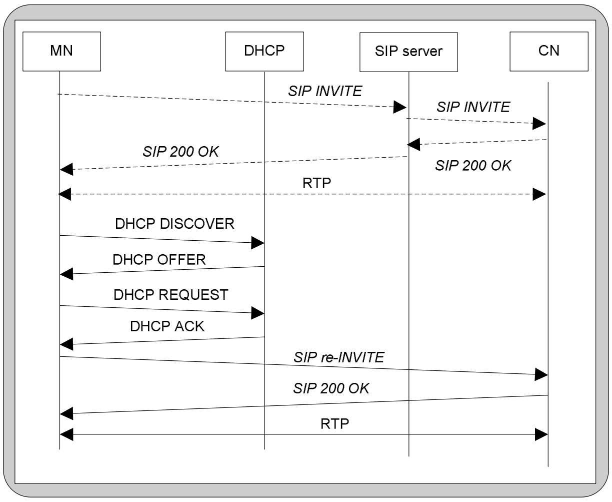

Fig. 4: Message exchange – mid-call mobility

2.3 Enhanced mid-call mobility

In this paper we focused on solutions which are likely to apply in real operator environment. In literature, some solutions are presented in which the MN informs the SIP server after sending SIP re-INVITE message [7]. However, in real operator's environment information about location change needs to be send to SIP server prior to starting the new SIP session between mobile node and correspondent node. This should be done to support proper charging as prices in two different networks can be different.

Thus, we proposed enchased mid-call mobility scenario, which is presented in Fig. 5.

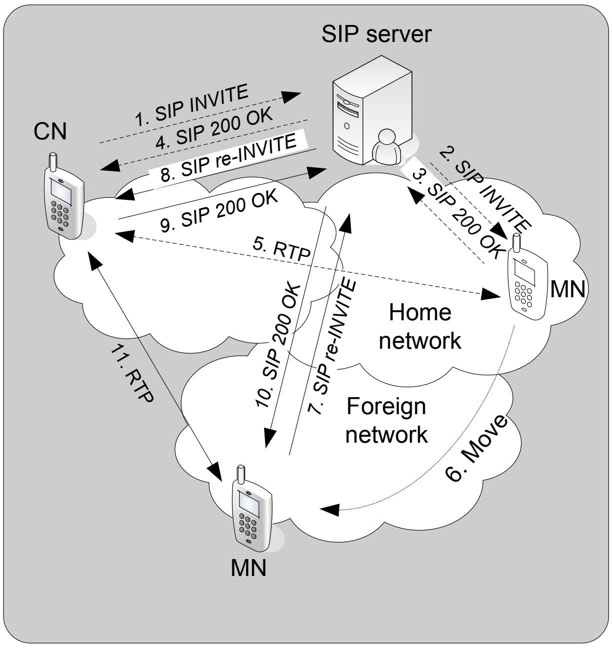

Fig. 5: Enhanced mid-call mobility

First five steps are the same as in mid-call mobility scenario. After the move to another network (step 6) the MN sends SIP re-INVITE message to the SIP server (step 7) to inform it about location change. SIP server than sends SIP re-INVITE message to the CN (step 8). After the acknowledgement (steps 9, 10) the new RTP session is established (step 11).

Like in both presented types of mobility the MN gets new IP address when it moves to new network. Fig. 6 presents message exchange for enchased mobility management scenario.

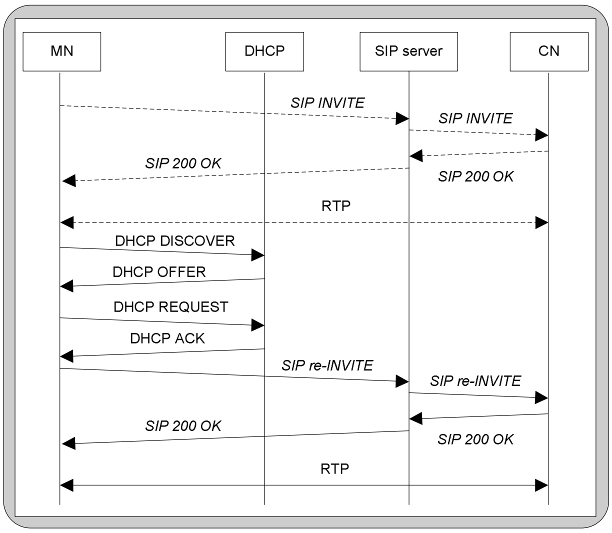

Fig. 6: Message exchange – enchased mid-call mobility

3 Handover performance evaluation

Referring to previous chapters, handover is done in three phases. All of those phases need some time to finish. The sum of all time intervals gives us the overall time needed for handover process to finish.

As the overall handover time influences the end-to-end delay of the application, which is in use, it needs to be small. Especially for time critical applications, like VoIP, end-to-end delay can have high impact on quality of service, which can be noticeable to a user as service degradation.

Table 2 shows delay and its impact on different applications as defined by ITU [9]. During handover there is usually some period of silence. We can use ITU recommendations when evaluating impact of handover time to the quality of service.

Another parameter that can have impact on user experience is packet loss, which is usually caused by congestions in the network. During the handover some packets are also lost. For real-time applications packet loss that is bigger than 20% is critical [9].

Delay by direction	Comments
0 - 150 ms	Acceptable for most conversations; only some highly interactive tasks may experience degradation.
150 - 300 ms	Acceptable for low-interactivity calls (satellite 250 ms per hop).
300 – 700 ms	Practically a half-duplex call.
above 700 ms	Unusable unless the callers are well-versed in the art of half-duplex conversation (as used in the military).

Table 2: Transition delay

3.1 Simulation model

In order to analyze the proposed solution presented in chapter 2.3 we have developed simulation model of telecommunication system, which comprise two networks, WLAN and HSDPA. We have also defined two mobile nodes (MN1 and MN2) that are using IP telephony as an application. In simulation MN1 will move from HSDPA network to WLAN network and then back to HSDPA network. The real example of such a case could be scenario in which user is in the middle of the call, which was established via HSDPA network, and then user moves to WLAN network coverage (e.g. congress center, his home, etc.) and performs handover to the new network. Than still in the call it moves back to HSDPA network, looses WLAN network connectivity and handover is made back to HSDPA.

To achieve results which are as close as possible to real operator environment, we collected information about wireless network characteristics. PING command was used for calculating the delay in WLAN and HSDPA networks. The target for PING was proxy server of real service provider. For testing the HSDPA network we used laptop computer with HSDPA card, while for WLAN we used WLAN interface of the laptop. The WLAN router was connected to Internet via leased line. We sent 10 packets with packet length of 200 bytes, which is approximately the size of the packet when using VoIP with G.711 codec and packetization time 20 ms. We test this for both interfaces. Because PING packets can travel using different paths, we calculated the average end-to-end delay. The results are presented in Table 3.

Network	Average end-to-end delay (one way)
WLAN	10 ms
HSDPA	100 ms

Table 3: Average end-to-end delay

The above values are used for modeling the appropriate core networks.

In this paper we focused on last two phases of handover, i.e. decision for handover and handover execution.

The following assumptions were made in simulation scenario:

- The HSDPA network is always available and is more expensive.
- The WLAN network has limited coverage (e.g. congress centers) and is cheaper.
- WLAN network is prioritized, which means that handover will be executed to WLAN network always when such a network will be availible and power of received signal will be above predefined threshold.
- MN1 is a dual mode handset capable of sending RTP packets and SIP INVITE messages at the same time via different interfaces.

In general, handover decision can be made based on different triggers:

- S/N ratio (user is using the network with the best available signal);
- QoS parameters in the network;
- bandwidth of the target network;
- economic price (user is always using cheapest network);
- combinations of the above triggers.

In this paper we focused on S/N ratio of received signal from WLAN network. This trigger is based on received power and it is usually the prerequisite for handover (i.e. we can not handover to network without or limited signal coverage).

We defined simple decision function, based on which MN1 decides for handover process to start.

The decision is made as in (1). First we need to determine in which network MN1 is. If at the beginning of the simulation S/N_{wlan} is above $T_{S/N}$ that means that user is in WLAN network, thus we set `wlan_network = 1`. If S/N_{wlan} is below $T_{S/N}$ that means that user is in HSDPA network and we set `hsdpa_network = 1`. After that, MN1 constantly measures S/N ratio of WLAN network and writes the value in S/N_{wlan} attribute. For every measurement we make a comparison of S/N_{wlan} to $T_{S/N}$. If S/N_{wlan} is above $T_{S/N}$ and user is in HSDPA network the MN1 will perform handover to WLAN network. If S/N_{wlan} is below predefined threshold $T_{S/N}$ and the user is in WLAN network the MN1 will perform handover to HSDPA network.

```
wlan_network = 0;                              (1)
hsdpa_network = 0;

S/Nwlan = S/N_ratio_of_WLAN_network;

if (S/Nwlan > TS/N)
  {
  wlan_network = 1;
  }
else
  {
  hsdpa_network = 1;
  }

repeat
{
S/Nwlan = S/N_ratio_of_WLAN_network;

if (S/Nwlan > TS/N &&
hsdpa_network==1)
 {
   execute_handover to WLAN;
   wlan_network = 1;
   hsdpa_network = 0;
 }

if (S/Nwlan < TS/N &&
wlan_network==1)
 {
   execute_handover to HSDPA;
   hsdpa_network = 1;
   wlan_network = 0;
 }
} until sim_time < defined sim_time
```

The decision for handover starts the handover process. MN1 then sends SIP INVITE message via new network to SIP proxy server, which forwards it to MN2. The MN2 than send acknowledge message SIP OK back to MN1 via SIP proxy server as described in chapter 2.3. During SIP message exchange for RTP traffic transmission still old network is used. With such approach we lower the time interval in which user is experiencing degradation of service due to packet loss.

The simulation model of communication system was developed using discrete event object-oriented modeling simulation tool OPNET Modeler [10]. It has open source code of commonly used protocols, which is very convenient for performance evaluation of user developed / enchased mobility management mechanisms [11]. OPNET is the network modeling and simulation tool for designing new protocols and technologies, and performance evaluation of existing and newly developed optimized protocols and applications.

It has hierarchical modeling environment consisting of three levels: (i) project level, (ii) node level and (iii) process level.

The project level graphically represents the topology of a communications network. It allows users to create node and link objects to represent network topology elements and configure them quickly.

The node level captures the architecture of a network device or system by depicting the flow of data between functional elements, which typically represent network protocols or algorithms and are assigned process models to achieve any required behavior.

The process level uses a powerful finite state machine (FSM) approach to support detailed specification of protocols, resources, applications, algorithms, and queuing policies. FSMs are dynamic and can be spawned during simulation in response to specific events. The C/C++ code that governs each state of a process model can be rapidly customized. OPNET Kernel Procedure APIs exist to facilitate development and support common communications mechanisms, such as packets, queues, and traffic.

As we decided for handover on application layer, which is not supported by OPNET we customized some pre-defined process models that incorporate SIP procedures. Fig. 7 shows added FSM states for SIP based handover of parent process for VoIP communications. When handover is initiated, simulation execution is moved to state handoff. If handover process is successfully finished, simulation execution is moved to state open and new connection is established, otherwise the handover process is terminated.

For better control of the simulation we defined new SIP message called INVITE_HANDOVER that is sent to proxy server when handover occurs. The message has the same structure as SIP INVITE message; we define it just for the differentiation from the standard SIP INVITE message. After the message is sent the new RTP session is established via the second network and the handover occurs. We also defined dual mode terminal capable of connecting to two different networks.

Fig. 8 shows simulation network configuration. Mobile node MN1 is dual mode terminal, capable of connect to WLAN (representing fixed operator) and to HSDPA network (representing mobile operator). Both networks are connected to Router 3 to which the SIP proxy server is also connected. Mobile node MN2 is connected to the access point AP2.

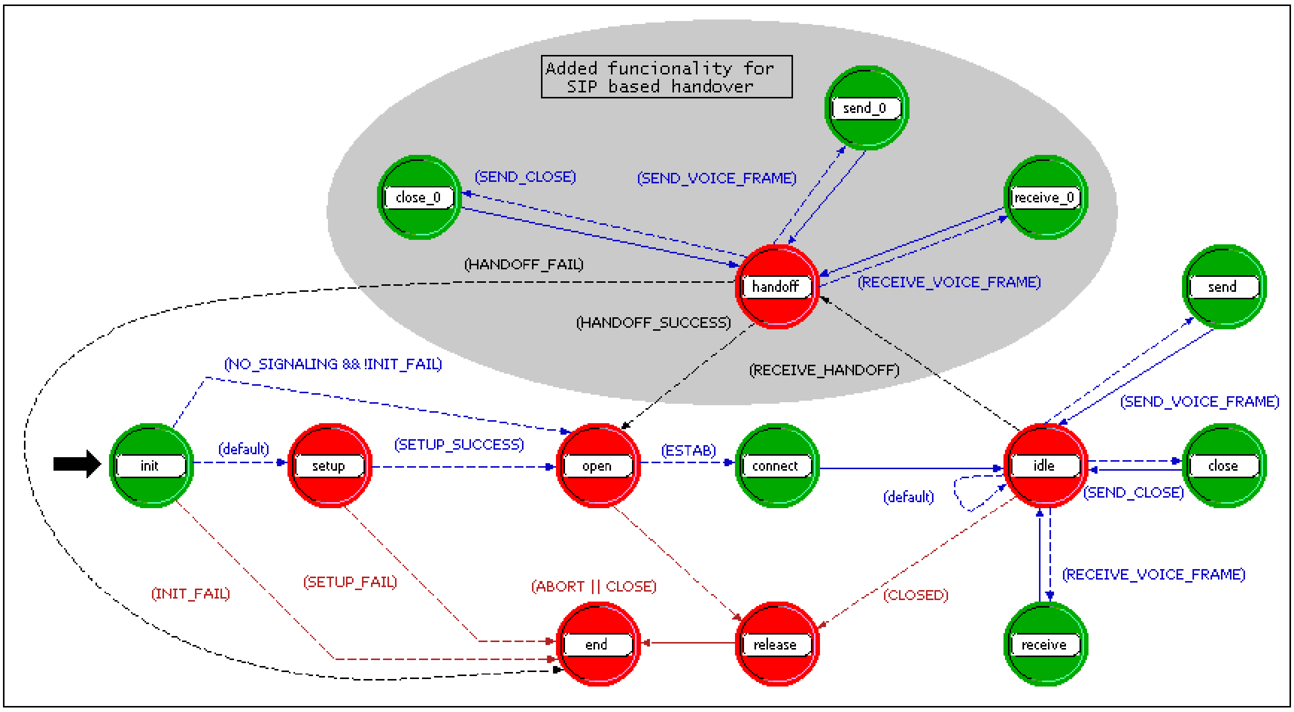

Fig. 7: Parent process for VoIP communications

For better control of the simulation we used static IP addresses on mobile nodes, proxy server and access points. On all other elements RIP protocol was used. In IP clouds we set latencies which were measured in HSDPA network (delay 100ms) and WLAN network (delay 10 ms).

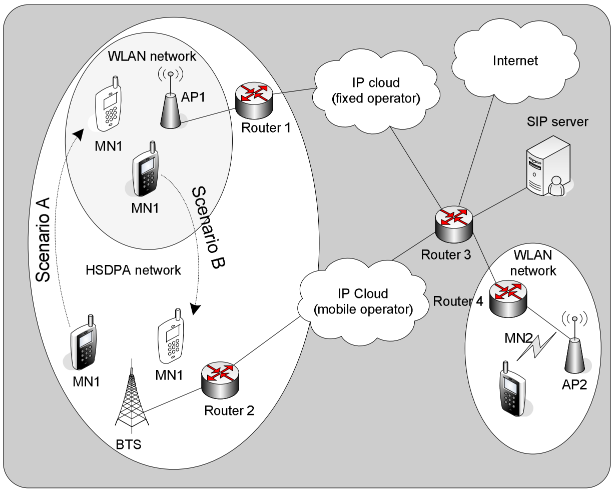

Fig. 8: Simulation network configuration

The simulation was 150 sec long. The first 112 sec were left for RIP massages to exchange and network routing to set automatically.

Between MN1 and MN2 the IP telephony application with SIP signalization was used. For VoIP the G.711 codec was used, while compression and decompression delay was set to 20 ms. The conversation time between MN1 and MN2 was set to 30 sec. During that time MN1 will move to the WLAN network coverage and then back out of WLAN network. The conversation profile of both MNs was configured so that we are simulating constant voice from/to both MNs with no silence in between. With such approach we could have better control of the simulation i.e. constant RTP traffic from both MNs for the whole period of simulation.

As in simulation MN1 will perform handover from HSDPA to WLAN network and back, we defined two scenarios:

- Scenario A: MN1 moves from HSDPA to WLAN network and performs handover.
- Scenario B: MN1 moves from WLAN to HSDPA network and performs handover.

At the beginning of the voice application simulation MN1 is connected to HSDPA network and establish a SIP session to MN2 via SIP server. During the call MN1 starts to move to WLAN network coverage area (scenario A). The S/N ratio of WLAN network starts to increase. When S/N ratio exceeds predefined threshold $T_{S/N}$, the new SIP session is established and handover is executed. Still in the middle of the call user starts to move back from WLAN network coverage area (scenario B), thus S/N ratio starts to decrease and when it falls below $T_{S/N}$ handover is performed back to HSDPA network.

In Fig. 9 S/N ratio of WLAN network, which is measured on MN1 during voice conversation, is presented for scenario A and scenario B. Due to constant speed of MN1 and open space the signal strength increase and decrease linearly.

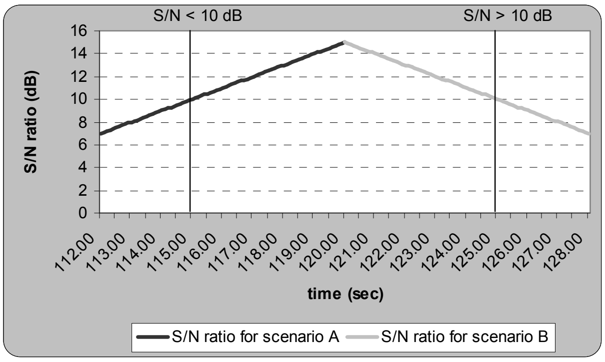

Fig. 9: S/N ratio of WLAN network on MN1 for scenario A and scenario B

For better control of simulation we set cck 11 modulation, which is also used in real WLAN networks. Based on modulation curve we can define BER of the radio channel. The BER values at different S/N ratios for cck 11 modulation are presented in Table 4:

S/N	(+0dB)	(+0.25dB)	(+0.50dB)
+ 5.00	0.0081	0.005	0.003
+ 5.75	0.0021	0.0014	0.00078
+ 6.50	0.00044	0.00025	0.00014
+ 7.25	0.000081	0.000044	0.000022
+ 8.00	0.00001	0.0000048	0.0000021
+ 8.75	0.00000081	0.00000037	0.00000016
+ 9.50	0.000000061	0.000000016	0.0000000056

Table 4: cck 11 BER

As it can be seen for selected modulation S/N ratio which is bigger than 10 dB defines channels BER as almost zero, thus in simulation scenario we set threshold $T_{S/N}$ to 10 dB.

3.2 Simulation results

In the simulation we focused on performance measures which can affect user experience. These are: (i) handover time and (ii) packet loss during handover.

As we defined that MN1 constantly measures S/N ratio of WLAN the new network detection time can be eliminated from calculation. The biggest part to overall handover time is contributed by handover execution i.e. new SIP session establishment, which is defined as a time interval from decision for handover to the time when new session is established via new network. When SIP messages for new session are exchanged, both MNs start to send RTP traffic via new network. To get overall handover time this time interval needs to be added to new SIP session establishment time. Handover time is than calculated according to (2):

$$T_{handover\_time} = T_{new\_session} + T_{first\_RTP\_packet} \qquad (2)$$

The $T_{new\_session}$ is defined as time needed that all SIP messages are exchanged and $T_{first\_RTP\_packet}$ as a time needed that the first RTP packets of the new stream traverse the network. Time interval $T_{first\_RTP\_packet}$ represents silence in the conversation. In case that MN1 would not be capable of sending SIP INVITE messages and RTP traffic at the same time on different interfaces the silence in the conversation would be equal to $T_{handover\_time}$.

When new SIP session is established and both MNs start to send RTP traffic via interfaces to the new network, there are still some packets that traverse the old network. Those packets are lost in the conversation and will be discarded. The proportion of such packets represents the packet loss during handover and is defined as in (3).

$$packet\_loss = 1 - \left(\frac{received\_packets}{sent\_packets} \right) \qquad (3)$$

The number of packets is measured on the MN1 (received packets) and MN2 (sent packets) in time period from SIP INVITE_HANDOVER message to the time when first RTP packets arrive. To distinguish which packets were sent via new network, we mark those packets with flag defined in IP packet.

3.2.1 Scenario A

In Fig. 10 throughput on all interfaces of MN1 is presented for scenario A.

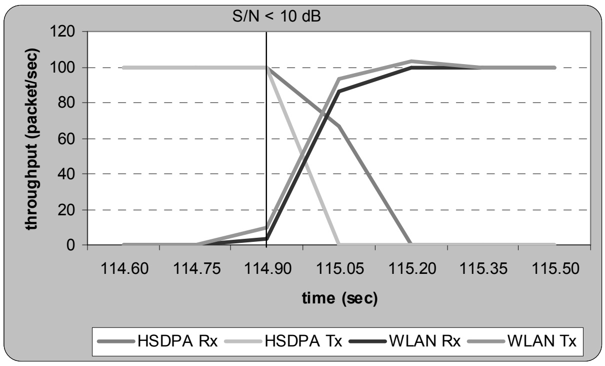

Fig. 10: Throughput on all interfaces of MN1 (scenario A)

As it can be seen $T_{S/N}$ is exceeded at 114.9 sec of the simulation. After that the throughput on HSDPA receiver and transmitter start to decrease, while the throughputs on WLAN interfaces start to increase. It can be also seen that the handover is executed almost immediately after $T_{S/N}$ is exceeded.

Table 5 shows the handover characteristics measurements for scenario A.

Parameter	Value
$T_{new\_session}$	51 ms
$T_{first\_RTP\_packet}$	15 ms
$T_{handover\_time}$	66 ms
Packet loss	33 %

Table 5: Handover characteristics measurements for scenario A

From the results we can see that handover time to WLAN network is not critical. This is especially due to very fast connection in WLAN network, which enabled quick SIP session setup of 51 ms and first RTP packets transmission of 15 ms, which represents silence in the conversation.

On the other hand it can be seen that packet loss during handover is 33 % which can be critical. Such a proportion of discarded packets happened due to bigger delay in HSDPA network and large number of VoIP packets in HSDPA network.

3.2.2 Scenario B

In Fig. 11 throughput on all interfaces of MN1 is presented for scenario B.

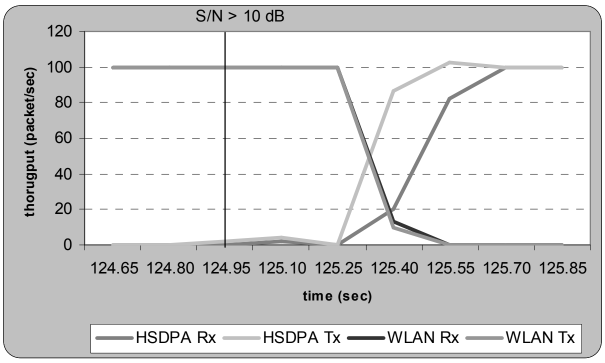

Fig. 11: Throughput on all interfaces of MN1 (scenario B)

The S/N ratio of WLAN network felt below $T_{S/N}$ at 125.0 sec of the simulation. This started handover process.

It can be clearly seen that the SIP INIVTE_HANDOVER message was sent via HSDPA Tx and Rx of MN1, while for the RTP stream still old network has been used (i.e. packet traversing via WLAN Tx and Rx interface of MN1).

Because of bigger delays in HSDPA network the SIP message exchange was longer than in scenario A. After the SIP acknowledge MNs started to sent RTP packets via new network. When this happened throughputs on HSDPA Rx and Tx interfaces started to decrease, while throughputs on WLAN Rx and Tx interfaces started to increase.

Table 6 shows the handover characteristics measurements for scenario B.

Parameter	Value
$T_{new\_session}$	411 ms
$T_{first\_RTP\_packet}$	105 ms
$T_{handover\_time}$	516 ms
Packet loss	5 %

Table 6: Handover characteristics measurements for scenario B

From the results we can see that total handover time is more than half a second. This is especially due to bigger delay in HSDPA network. Bigger delay also longer also the silence interval to 105 ms.

The second parameter is packet loss which is not critical in this scenario. Because delay of the WLAN network is very low there is very small number of packets that traverse WLAN network, which sets packet loss to 5%.

3.2.3 Discussion

The results showed that overall handover time is highly dependent on network's end-to-end delay, which is used for sending handover requests. In scenario A handover was done from network with bigger delay to the network with smaller delay. We showed that the handover time is not critical and can be performed also for real-time applications. On the other hand in scenario B when we performed a handover from smaller delay to the network with bigger delay the handover time can affect user experience. In simulation we assumed that MN1 is capable of sending handover signalization messages and RTP stream at the same time. If MN1 would not have such capability the interval with silence in conversation will be equal to handover session setup, which is 516 ms in case of scenario B. Such conversation disruption is noticeable by the user.

The second parameter that we measured was packet loss during handover, which is also dependent on new network's end-to-end delay. In scenario A large proportion of packets was discarded. Such a proportion of lost packets can affect QoS of the service. Packet loss is especially critical for VoIP application as packets are small and frequent, which means that a lot of packets traverse the network at the same time in case the delay of the network is big. In scenario B the packet loss was not critical.

With results we showed that performing handover still has some challenges. When handover is done from slower network, like HSDPA to faster network, like WLAN, packet loss can be critical and when performing handover form slower network to faster network conversation can be disrupted due to longer interval of silence.

In simulation we assumed that MN1 is in the open space and that there is line of sight between MN1 and AP1. In real WLAN network users can often experience big signal oscillation, which means that handover could be executed several time one following another, which could have big effect to the user experience.

4 Conclusion

In this paper we focused on mobility management using SIP protocol. Due to its independence of access technologies SIP can be the right candidate for making handovers in real operator environment. Although in this paper we were focusing more on simulation model development. The simulation results show that today latencies in mobile wireless networks are too high to offer seamless handover for real-time applications. In the simulation we also measured packet loss during handover, which also can be very critical.

Pre-4G networks which will be available in the future will offer much bigger bandwidth and will improve characteristics of mobile wireless networks.

In our future work we will further enhance handover mechanism in order to improve handover delay and packet loss.

References:

[1] British Telecom, www.bt.com, 2007.

[2] C. Perkins, Ed., IP Mobility Support for IPv4, IETF RFC 3344, August 2002.

[3] K. El Malki, Low Latency Handoffs in Mobile IPv4, IETF INTERNET-DRAFT, 3 October 2005.

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

[5] Seok J. Koh, Mee Jeong Lee, Maximilian Riegel, Mary Li Ma, Michael Tuexen, Mobile SCTP for Transport Layer Mobility, IETF Internet Draft, June 2004.

[6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, SIP: Session Initiation Protocol, IETF RFC3261, June 2002.

[7] R. Rajavelsamy, S. Anand, Osok Song, Sungho Choi, A novel scheme for mobility management in heterogeneous wireless networks, Wireless Personal Communications, Vol. 43, No 3, 2007, pp. 997-1018.

[8] Shantidev Mohanty, Ian F. Akyildiz, Performance Analysis of Handoff Techniques Based on Mobile IP, TCP-Migrate, and SIP, IEEE Tranactions on mobile computing, Vol. 6, No. 7, 2007, pp. 731-747.

[9] Group of Experts on IP Telephony / ITU-D, The Essential Report on IP Telephony, ITU 2003.

[10] OPNET Technologies, www.opnet.com, 2007.

[11] P.M.L., Chan, R. A. Wyatt-Millington, A. Svigelj, R. E. Sheriff, Y. F. Hu, P. Conforto, C. Tocci, Performance analysis of mobility procedures in a hybrid space terrestrial IP environment. Comput. networks, Vol. 39, 2002, pp. 21-41.