TRIM: an Architecture for Transparent IMS-based Mobility

Ivan Vidal[a,*], Antonio de la Oliva[a], Jaime Garcia-Reinoso[a], Ignacio Soto[b]

[a] *Universidad Carlos III de Madrid. Avda. de la Universidad 30*
28911 Leganés - Madrid (Spain)
[b] *Departamento de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid*

Abstract

In recent years, the development and deployment of new wired and wireless access network technologies have made the ubiquitous Internet a reality. Users can access anywhere and anytime to the broad set of value-added Internet services, which are delivered by means of the IP protocol. In this context, 3GPP is currently developing the IP Multimedia Subsystem (IMS), as a key element that allows to evolve from the ubiquitous access to the Internet services towards a next generation network model, by providing a set of essential facilities such as session control, QoS, charging and service integration. Nevertheless, several open issues still need consideration before the future Internet becomes real, such as supporting user mobility in IP networks. Although mobility support in the Internet is receiving much attention, IMS networks present inherent particularities that require further analysis. The solutions proposed so far for IMS do not support mobility transparently to the end-user applications, or address the problem by introducing complex changes to the IMS infrastructure. This paper presents TRIM, an architecture for transparent IMS-based mobility. TRIM supports mobility in IMS networks transparently to the end-user applications, which are unaware of the handover management procedures executed between the mobile node and the network. We have performed several experiments with a TRIM prototype, using a real IMS testbed with 3G and WLAN access networks, validating the proposal for UDP and TCP based applications.

Key words: SIP, IMS, Transparent mobility, Handover management.

*Corresponding author
Email addresses: ividal@it.uc3m.es (Ivan Vidal), aoliva@it.uc3m.es (Antonio de la Oliva), jgr@it.uc3m.es (Jaime Garcia-Reinoso), isoto@dit.upm.es (Ignacio Soto)

1. Introduction

IP is becoming the cornerstone technology for communication networks. In this context, telecommunication operators are pushing the introduction of the IP Multimedia Subsystem (IMS) in their IP networks to provide access and session control. IMS enables in IP networks services traditionally associated with circuit switched networks such as telephony. These services can combine several media and QoS requirements. IMS was developed by 3GPP[1] for UMTS networks, and it is currently working together with ETSI-TISPAN to extend the IMS specification for any type of access technology.

Mobility and the ability to access the Internet everywhere and anytime are characteristics that users have come to expect. Mobility support in IP networks has been studied for some time. The key issue is that, in IP, addresses are used both as identifier and locator. So when a node moves, from the point of view of the applications we need to maintain its IP address (because it is an identifier) but from the point of view of the IP layer we should change the IP address to one topologically correct. The IETF[2] has standardized the Mobile IP (MIP) protocols both for IPv4 [1] and for IPv6 [2] to provide mobility support in IP networks. Mobile IP makes mobility transparent to any communication layer above IP, including the applications and, therefore, a node is able to change the IP subnetwork it is using to access the Internet without losing its active communications. Note that some networks support the movement of nodes without requiring them to change their IP address. In these networks mobility is transparent to the nodes at the IP layer, so layers above IP are not affected by mobility and no additional mobility support is required. For example, this is the case when a MN moves within the same UMTS network. Nevertheless, there are trends that will make mobility among different networks, meaning the change of IP address and consequently the need for a mobility support mechanism that deals with this change, much more common. Examples of these trends are the use of devices with several technologies that connect through the most appropriate one in each situation, the availability of new access technologies that will co-exist to cover different requirements, and operators that integrate several offerings (fixed and mobile access).

[1] http://www.3gpp.org
[2] http://www.ietf.org

Mobile IP is a good solution to provide mobility support but its integration in IMS is far from trivial [3, 4, 5]. This is essentially because MIP hides from the application layer, including the IMS control, the IP address used by a MN in an access network, but IMS requires this address to reserve resources in the access network for the traffic of the services used in the node.

Another alternative is to use the Session Initiation Protocol (SIP) to support mobility [6] in IP networks. In this respect, 3GPP has proposed a set of mechanisms to maintain service continuity in the event of terminal mobility [7]. Using SIP to handle mobility presents the advantage of not requiring additional mechanisms outside the signaling used in IMS. But the traditional SIP mobility support is not transparent to the application layer. This means that applications have to be programmed with mobility in mind which is inconvenient. Section 5 reviews different approaches to provide mobility support in IP networks with IMS, comparing them with this work.

This paper proposes TRIM (described in Section 3), an architecture providing mobility support in IMS-based networks. It is based on SIP signaling and does not require any modifications to the IMS specifications, which is an important advantage over the alternative of using Mobile IP in an IMS network. TRIM makes the change of IP address in a mobile node, required to support the movement, transparent to the applications running in the mobile node and also to the peer nodes. We implemented a prototype of the TRIM architecture and tested the proposal using different access networks (3G and IEEE 802.11) and both UDP and TCP user traffic (see Section 4).

2. Background on IMS

The IP Multimedia Subsystem (IMS) is an architectural framework introduced by 3GPP as part of the standardization process of the UMTS technology. It was designed as a key element to enable the delivery of value-added multimedia services, including those that were traditionally provided through circuit-switched networks, over packet-switched networks. In this respect, the IMS is designed to support facilities related to session control, QoS, charging, interworking with the Internet and circuit-switched networks, service control and creation, integration of heterogeneous access technologies, security and roaming. Figure 1 depicts a simplified overview of the architecture defined by 3GPP for the IMS (see [8] for further details).

In the IMS architecture, session control is based on the Session Initiation Protocol (SIP) and the Session Description Protocol (SDP). Although these

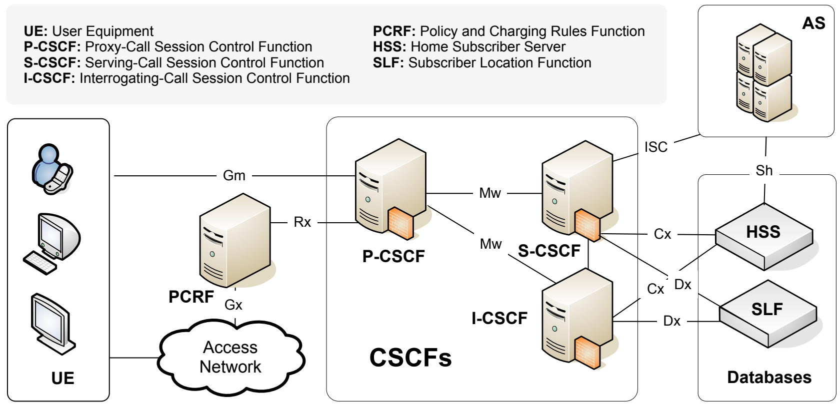


Figure 1: IMS architecture

are Internet protocols standardized by the IETF (see [9] and [10] respectively), session control in the IMS follows a specific profile for SIP and SDP defined by the 3GPP [11]. In the figure, the Call Session Control Functions (CSCFs) are the functional entities responsible of processing the SIP signaling messages: the Proxy-CSCF (P-CSCF) is the first contact point between the UE and the IMS, and processes every SIP message that originates or terminates in the UE; the Interrogating-CSCF (I-CSCF) is the contact point in the user home network for every incoming session addressed to the UE; the Serving-CSCF (S-CSCF) performs session control and registration functionalities. The S-CSCF may also route incoming SIP signaling messages to a set of Application Servers (ASs), which provide services to the end user. In addition, other functional entities are specially relevant in the architectural framework: the user databases, i.e. the Home Subscriber Server (HSS), the Subscriber Location functions (SLF), and the Policy Control and Charging Rules Function (PCRF), which provides policy control decision and flow-based charging control functionalities.

3. The TRIM architecture

This section describes the TRIM architecture. TRIM supports mobility transparently to the end-user applications, which can use UDP or TCP in the user plane, and are unaware of the handover management procedures executed between the MN and the network. Unlike the previous proposals, the described solution does not require changes to the IMS infrastructure

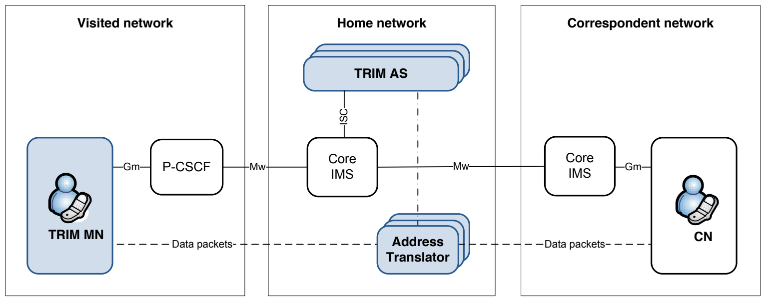


Figure 2: Overview of the TRIM architecture

as defined by 3GPP [8]. In addition, the proposal does not entail any upgrades to the correspondent node (CN, the node communicating with the MN) architecture, being the extensions only necessary in the MN and the home network corresponding to the TRIM user. Figure 2 depicts a general overview of the TRIM architecture, highlighting the relation of its different components with the IMS infrastructure[3].

The main component of the TRIM architecture is a SIP Application Server (AS), namely the TRIM AS. Each user subscribed to the transparent handover service described in this section will be served by a TRIM AS located in its home network. This application server will receive all the SIP signaling messages corresponding to the multimedia sessions originated or terminated in the MN. This way, the TRIM AS stays on the signaling path utilized between the MN and any CN. To guarantee scalability and redundancy, the home network may contain a certain number of TRIM ASs.

On the other hand, the TRIM AS controls a set of Address Translators, which are located in the home network of the mobile user[4]. As the AS stays on the signaling path between the MN and the CN, it ensures that all the media exchanged by them in the user plane passes through one of these translators. The address translator is configured by the TRIM AS to address the media received from the CN to the appropriate location where

[3]In this figure, it is assumed that the MN has acquired connectivity from a visited network and uses a P-CSCF located in that network. However, a mobility scenario is possible where the MN uses a P-CSCF located in the home network of the mobile user. The procedures presented along this section can handle transparent mobility in both scenarios.

[4]Note that the TRIM AS and the address translators can be deployed in different network nodes. In this case, a protocol such as DIAMETER is needed to communicate them (a similar use case is described in [12]).

the MN is willing to receive it. Analogously, it forwards the media received in the opposite direction, i.e. from the MN to the CN. In concrete, the address translator simply changes the IP addresses and ports contained within the data packets according to the configuration provided by the TRIM AS. Therefore, the CN always observes the same remote addressing information for the MN, no matter where the latter is located. Again, for the sake of scalability and redundancy, there may be a certain number of address translators within the home network.

As it will be subsequently explained, the TRIM enabled MN keeps the AS updated with its current addressing information. This information, as well as the mechanisms to update it in the TRIM AS, are transparent to the end-user applications that run on the MN. The AS configures and updates the address translator assigned to the MN with the current addressing information that it utilizes in the user plane. Thus, the MN can always communicate with the CN irrespective of its location.

Hence, the presented solution introduces two new intermediate elements, the TRIM AS and the address translator, as well as the architecture of a TRIM enabled MN. In the user plane, the address translator plays a similar role to the Home Agent (HA) in MIP. Nevertheless, the proposal does not require changes to the IMS infrastructure as defined by 3GPP. Next subsections cover (1) the details of the TRIM architecture at the MN, as well as the signaling procedures that are necessary (2) to establish a multimedia session between the MN and the CN and (3) to handle the transparent handover of the MN in IMS networks.

3.1. Architecture of a TRIM enabled MN

As previously indicated, TRIM guarantees that the addressing information corresponding to the current location of the MN, as well as the procedures to notify this information to the TRIM AS, are transparent to the end user applications that run on the mobile device. To make this possible, we define new functionalities to an IMS enabled terminal (i.e. the MN).

Figure 3(a) illustrates the reference model that has been considered in this article for an IMS enabled terminal. In this model, an IMS SIP User Agent (IMS SIP UA) is the functional entity in the terminal in charge of executing all the SIP signaling procedures with the IMS network (e.g. registration, session establishment and release, etc.). The SIP UA can be triggered when some relevant lower-layer event takes place, for instance when the terminal gets IP connectivity to a new access network and the address of a new P-CSCF has been obtained. These lower-layer triggers can imply certain signaling procedures to be executed (e.g. an IMS registration after obtaining IP connectivity to a new access network). The SIP UA is enabled to open network sockets towards the lower-layers, to support the exchange of SIP signaling messages with the IMS network.

On the other hand, the user terminal may execute a set of applications requiring network connectivity, such as VoIP, online gaming or IPTV. These applications can contact the SIP UA in order to establish multimedia sessions towards remote parties. For instance, in case of a VoIP call, an audio session needs to be established with the callee before any voice data can be sent or received. In this case, the VoIP application running on the terminal would generate a description of the audio session to be established, according to the procedures specified in the 3GPP profile of SDP and would request the SIP UA to establish the session with the callee. In addition, applications are allowed to open network sockets in order to support the data exchange that is necessary for an appropriate operation (e.g. to send and receive audio packets in case of VoIP).

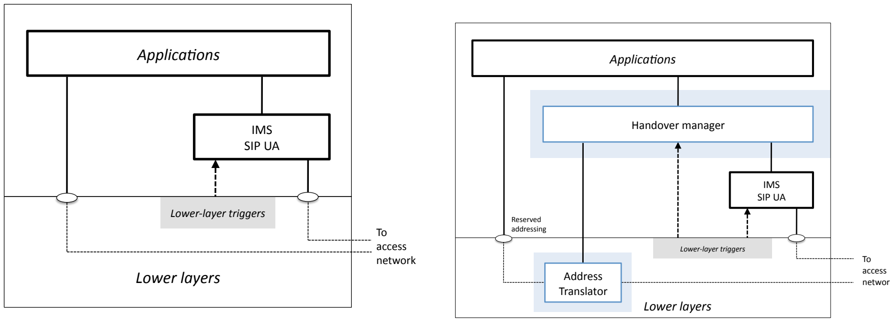
Nevertheless, under this model, mobility is not transparent to applications. In fact, when the MN gains IP connectivity to a new access network:

- Any ongoing multimedia session, established with a CN, needs to be re-established, in order to guarantee an appropriate resource reservation in the new access network and to inform the CN of the new addressing information that must be used in the user plane. In this process, the application must provide the SIP UA with a new SDP description, reflecting the new addressing information for the media components that will be exchanged within the session.

- Any network socket bound to an IP address that belongs to the old network becomes invalid. Hence, applications need to re-configure or open new network sockets.

Figure 3(b) shows the extensions proposed in this paper to an IMS enabled MN to support transparent handover for the end-user applications. In this new architecture, any call to the lower layers to retrieve the MN's IP address receives back a private/loopback IP address. In addition, TRIM introduces two new functional entities into an IMS terminal: a handover manager and an address translator.

The handover manager (HM) is located in the user space of the MN, and acts as an intermediary between the applications and the SIP UA. This entity receives requests from the applications to establish, modify and release multimedia sessions towards any specific destination. In addition, each request can contain an SDP payload. If so, the payload comprises the



(a) Architecture of an IMS terminal (b) Architecture of a TRIM enabled MN

Figure 3: Introducing the TRIM architecture into an IMS terminal

description of the different media components (e.g. audio or video) that the application wants to include in the multimedia session. For each media, among other things, the application indicates the addressing information (IP address and port) where it is willing to receive the component.

The HM modifies the SDP payload received from the application, changing the addressing information associated with each media component. In particular, the manager replaces the internal IP address, which is visible for the applications, by the real IP address that the MN got from the current access network (port numbers are left unchanged). Then, the handover manager forwards the request to the SIP UA, which performs the signaling procedures that are necessary to serve the application request. Additionally, the manager stores the current SDP payload and an identifier of the multimedia session for further use.

Analogously, the SIP UA can receive a request from a remote party to establish, modify or release a multimedia session with any application that runs on the MN. This request is sent to the handover manager. As before, the request can contain an SDP payload. This payload indicates, for each described media component, the addressing information where the remote party wants to receive the component. This addressing information does not need to be changed and, consequently, the request is transparently forwarded to the intended application.

The handover manager will be triggered by the lower-layers when the MN gets IP connectivity to a new access network. If this happens, the HM will retrieve the information of the ongoing multimedia sessions. For each of them, it updates the stored SDP payload, changing the addressing information for each media component. In concrete, the manager replaces the old IP address with the IP address that the MN has obtained from the new access network. Then, the handover manager contacts the SIP UA to modify the multimedia session. As a result, a signaling procedure is executed between the SIP UA and the TRIM AS, and the AS obtains the new SDP description for the multimedia session from the MN. With this information, the TRIM AS can update the address translator configuration, so that subsequent data packets, corresponding to the multimedia session, are properly routed towards the new location of the MN. The details of this signaling procedure are provided in Sect. 3.3. The procedure is executed for each multimedia session established with the MN.

The address translator is located in the kernel space within the MN. This address translator runs locally at the MN and is independent from the set of address translators shown in Fig. 2. As previously indicated, the address translators illustrated in that figure are located in the home network, and are responsible of (1) addressing the data traffic received from the CN to the appropriate location of the MN and (2) providing the CN with stable remote addressing information for the MN. As we explain next, the address translator located at the MN complements this functionality, allowing applications that run locally at the MN to utilize internal addressing information that remains stable independently of the MN location.

In general, an application can open a set of network sockets to exchange traffic within a multimedia session. Each of these sockets is bound to the internal IP address and to a given port. Therefore, traffic sent through these sockets includes the internal address as the source IP address. The address translator behaves like a NAT, changing the source IP address of every packet sent in the uplink direction (from applications to network) by the real IP address assigned to the MN in the current access network.

On the other hand, as previously explained, the remote party in a session receives an SDP payload with the real addressing information corresponding to the MN. Therefore, packets belonging to the session will be addressed from the remote party to the current IP address of the device. The address translator also replaces the destination IP address of every packet received from network to applications by the internal IP address.

This way, applications interact with the HM to establish, modify or release multimedia sessions. The HM proxies requests between applications and the SIP UA, changing the local addressing information in the SDP payloads. This way, the handover manager hides the current addressing information to the applications, that always observe an invariant IP address (the internal IP address), no matter where the MN is located. On the other hand, the HM always provides the SIP UA with appropriate SDP payloads, containing the current addressing information corresponding to the MN. In the user plane, the address translator at the MN changes the addressing information of the data packets, so that local applications always receive data in the sockets they are bound to, while the remote parties receive traffic originated from the real address of the MN.

3.2. Procedures for session establishment

Whenever a local application, running on the MN, needs to exchange media with an application running on a CN, the MN must first establish a multimedia session with the CN. This procedure is triggered by the local application, which generates an SDP payload describing the session and requests the HM to establish a multimedia session with the CN. As we have explained, the HM updates the local addressing information contained in the SDP payload, and proxies the request to the SIP UA, which is in charge of the session setup. The signaling process corresponding to the session establishment is illustrated in Fig. 4. It is assumed, in this figure, that the MN establishes the multimedia session from a visited network where it needs to reserve local resources.

The signaling process is initiated by the SIP UA in the MN, by sending a SIP INVITE request that, according to the IMS routing mechanisms, arrives to the S-CSCF allocated to the user in the home network. This request contains an SDP offer, which is the SDP payload provided by the HM. If the user has contracted the transparent mobility service described in this paper, then the IMS initial filter criteria will indicate that the request should be routed to a TRIM AS, also located in the home network.

The TRIM AS is a SIP Back-to-Back User Agent (B2BUA), as defined in [9]. Assuming this role, it stays on the path of subsequent SIP requests and responses exchanged between the MN and the CN. In addition, the TRIM AS processes the SDP offer received from the MN. In concrete, for each media component included in the offer:

- The IP address and port included in the offer represent the addressing information where the MN is willing to receive the data traffic of the media component. The TRIM AS selects an address translator from the home network (see Fig. 2), and requests from the translator a new binding for the IP address and port. As a result, the address translator returns a new pair of IP address and port, which should be used by the CN to send the data traffic corresponding to the media component[5].

[5]For the sake of scalability, we assume that there may be several address translators

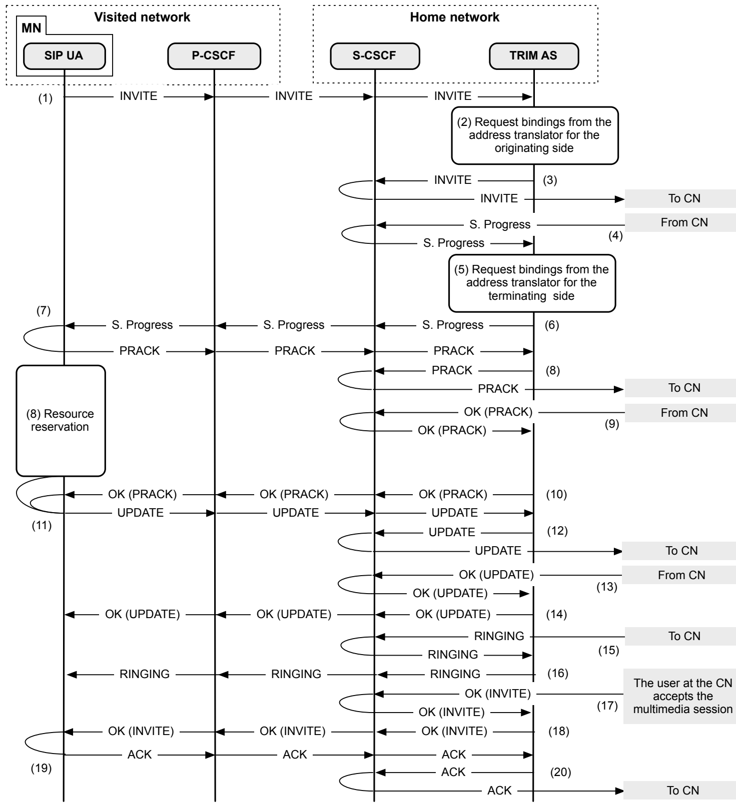


Figure 4: IMS signaling for session establishment

- The TRIM AS replaces the IP address and port specified for the media component with the binding obtained from the address translator.

Afterwards, the TRIM AS generates a new INVITE request, containing the modified SDP offer, and sends this request towards the CN. As this request is related with the INVITE request previously received at the AS, the new request is routed back to the S-CSCF. Finally, after processing the request, the S-CSCF forwards it towards the core IMS of the CN.

Eventually, the CN receives the INVITE request and answers it back with a SIP Session in Progress response. This response contains an SDP answer,

in the home network. In addition, each address translator may be assigned a set of IP addresses. This should enable the TRIM AS to successfully obtain a binding for the IP address and port corresponding to the media component.

describing the media components accepted by the destination. At some point, the Session in Progress response reaches the TRIM AS, that processes the SDP answer. For each media component included in the answer:

- The IP address and port included in the answer indicate the addressing information where the CN is willing to accept the data traffic corresponding to the media component. Again, the TRIM AS requests from the address translator, which had been previously selected, a binding for the IP address and port. As a result, the address translator returns a pair of IP address and port, which should be used by the MN as the destination of the data traffic corresponding to the media component.

- The TRIM AS replaces the IP address and port specified for the media component with the binding obtained from the address translator.

Finally, the TRIM AS generates a new Session in Progress response to the initial INVITE request that was received from the MN. In this response, the AS includes the modified SDP answer. The Session in Progress response is routed back to the MN, passing through the S-CSCF and the P-CSCF. At the MN, the SIP UA notifies the HM about the session disposition, providing it with the SDP answer. As it has been previously indicated, the HM proxies the request to the application that triggered the session setup. Therefore, the application is notified about the addressing information that it should use to send the data traffic within the multimedia session towards the CN.

The session establishment continues according to the regular IMS procedures. When the MN receives the SIP OK response to the INVITE request, the session has been established. At this point, the application is notified through the HM and the data exchange can start in the user plane. Figure 5 illustrates the address translation procedures that take place in the user plane after the session establishment.

The case where the CN initiates the session establishment towards the MN follows a similar procedure. Beyond the direction of the SIP signaling flows (in this scenario the INVITE request is received at the MN), the procedures executed by the TRIM AS remain basically unchanged.

3.3. Handling the transparent handover

When the MN moves to a new network and obtains IP connectivity from the new location, the SIP UA and the handover manager will be triggered by the lower layers. At this point, the SIP UA can register the new contact URI in the S-CSCF, where the MN is now reachable, and de-register the

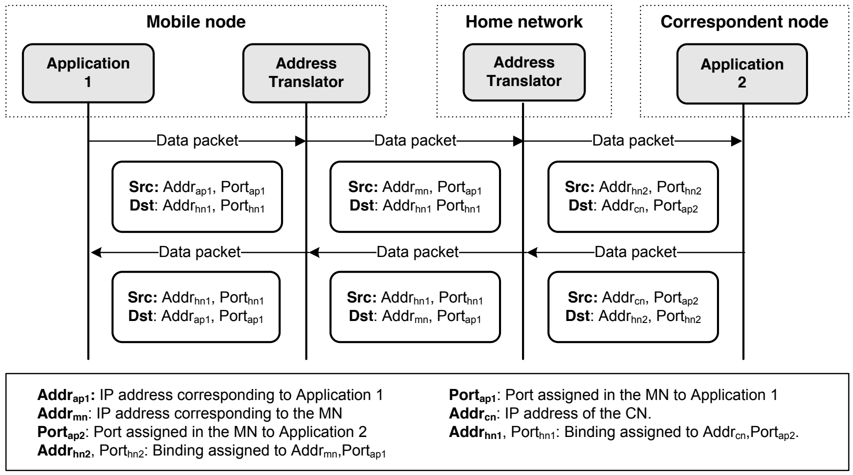


Figure 5: Management of data packets in the user plane

previous URI. On the other hand, the handover manager must ensure that the TRIM AS is informed about the new addressing information that must be used for every multimedia session that involves the MN. In this respect, as it has been previously explained, the manager retrieves the information about the ongoing multimedia sessions and, for each session, it updates the corresponding SDP payload, changing the addressing information for each media component to reflect the new IP address where the mobile node will receive the media. Next, the manager requests from the SIP UA the modification of the ongoing sessions, according to the new SDP parameters. The modification of an existing multimedia session is initiated by the SIP UA, by sending an INVITE request (a re-INVITE if the P-CSCF has not changed or a new INVITE request with a *Replaces* header in case that the P-CSCF has changed). Figure 6 depicts the signaling procedure executed between the SIP UA and the TRIM AS for each ongoing session, under the assumptions that the MN needs to use a P-CSCF and reserve local resources within the new access network.

In this signaling procedure, the TRIM AS receives in the INVITE request the updated information about the multimedia session that is being modified. This information is included in the SDP offer carried by the request. In concrete, the SDP offer specifies, for each media component, the new IP address and port where the MN is willing to receive the data traffic associated with the component. The TRIM AS stores this updated information and answers back the request with a Session in Progress response, that

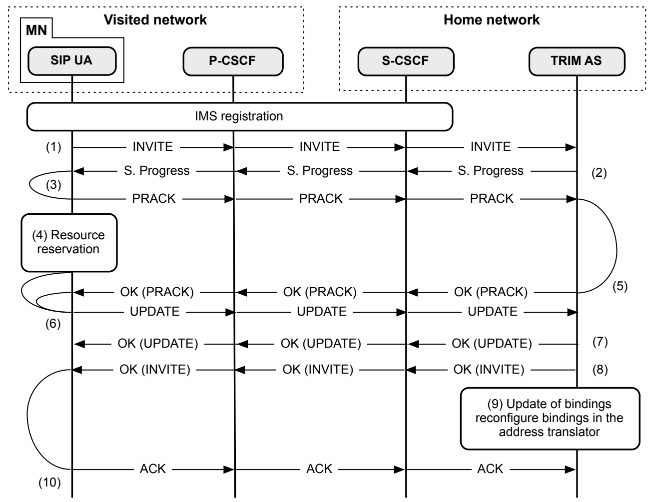


Figure 6: IMS signaling for handover management

includes an SDP answer. This SDP answer does not contain any changes to the addressing information that must be used by the MN to send data traffic (i.e. the MN should continue sending data within the multimedia session towards the address translator). When the SIP UPDATE request arrives to the TRIM AS, meaning that the MN has successfully accomplished the local resource reservation, the TRIM AS accepts the session modification by sending a SIP OK response to the INVITE request, and contacts the address translator in order to update the addressing information corresponding to each media component within the multimedia session. Consequently, the address translator starts forwarding the media belonging to the session to the new location of the mobile node. In addition, when the SIP UA receives the OK response, the HM is notified about the successful session modification, and configures the new IP address assigned to the MN in the local address translator. Additionally, the HM configures the lower layers to use the new interface for the uplink traffic.

The proposal described in this section is valid in a hard handover and in a soft handover scenario. The difference between both cases lies in the instant when local resources are released in the old access network. While in hard handover resources are released before gaining connectivity to the new access network, in soft handover this resources are be retained until data traffic is received through the new access network, being released afterwards.

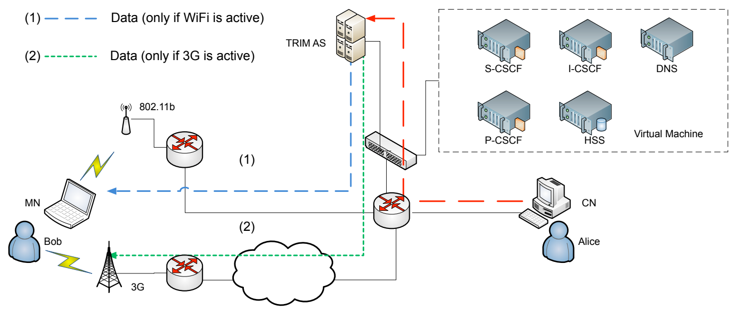


Figure 7: Testbed

4. Evaluation of the proposal

This section presents the results of the validation of the mobility solution described in the previous section. In order to do so, a complete prototype of the solution was implemented and different handovers for TCP and UDP transport protocols were performed. In the following subsections we provide insights of the behavior of the prototype implementation, and provide some considerations about the performance achieved by TRIM.

4.1. Prototype implementation

Figure 7 shows the testbed used to run all experiments. The elements of the TRIM architecture (the MN and the AS) have been implemented using Java and the JAIN-SIP API[6]. All the entities of the core IMS were deployed using the Open IMS Core implementation[7] in a single virtual machine running the Ubuntu 9.10 Linux distribution (all the machines used in this testbed were configured with the same Linux distribution). A DNS server was also configured in that virtual machine so all equipments use the same DNS server in order to resolve the *ims.net* virtual domain. The TRIM AS and one address translator were deployed in a single machine, which was attached to the same local area network of the core IMS (in the following, we will use the term TRIM AS for both entities). In the experiments, we have a user *Alice* registered at the CN, and another user *Bob* registered at the MN. Alice and Bob exchange data using a mobility unaware application. While the CN is always attached to a wired network, the MN has two wireless interfaces: a wireless 802.11g interface and a 3G/GPRS interface.

[6]JAIN SIP Developer Tools, https://jain-sip.dev.java.net/
[7]http://www.openimscore.org/

The TRIM AS is registered in the HSS of the core IMS and the users *Alice* and *Bob* have both an IP Multimedia Private Identity (IMPI) and an IP Multimedia Public Identity (IMPU) configured in the HSS and associated to the TRIM AS.

The data forwarding part of the prototype has been implemented using the NAT functionality included in Linux, by means of the *iptables* interface[8]. Whenever a handover is performed, the corresponding addresses configured in the MN and AS must be changed, requiring dynamic changes to the Linux NAT. As we show in the following sections, this functionality is not implemented in the Linux NAT services. Although we have been able to make a functional prototype of the solution using the standard Linux NAT, it is worth to notice that we would need a specific implementation of the address translator functionality of our proposal to overcome some limitations of the prototype that will be described later.

In this section, we will use the following address nomenclature:

- IP_{CN} is the address used by the CN to connect with the TRIM AS.
- IP_{AS_CN} and IP_{AS_MN} are the addresses used by the TRIM AS to connect with the CN and the MN, respectively.
- IP_{MN_3G} is the address used by the MN in the 3G leg.
- IP_{MN_WLAN} is the address used by the MN in the WLAN leg.
- IP_{MN_LO} is the address used by the MN in the loopback interface. Note that this is the internal address available to the applications.

In the validation, UDP and TCP experiments follow the same procedure. This procedure consists of a set of steps that are described next.

The MN is registered in the IMS core by issuing a REGISTER message containing its IP address. In these particular experiments we always started in the 3G leg so the IP address initially registered in the IMS core is IP_{MN_3G}. Once the MN is registered in the IMS core, it opens a communication with the CN, by issuing an INVITE message which will reach the CN through the AS. In all cases, the application running at the MN is listening at the address IP_{MN_LO} (private address manually assigned to the loopback interface)[9].

[8] See http://www.netfilter.org

[9] Please note that conceptually this is similar to use the loopback address 127.0.0.1, but the Linux kernel does not allow changing the destination IP address of an incoming packet to the loopback address

At this point, the MN and the AS have already configured their NAT modules. In the CN to MN direction, the NAT configuration allows the AS to forward all the packets arriving at the AS with destination the MN to the current MN's location (IP_{MN_3G}). The NAT configuration at the MN allows to address the packets of the session, received at IP_{MN_3G}, to the address IP_{MN_LO}, where the application is listening. In the MN to CN direction, the NAT configuration at the MN allows to change the source IP address (IP_{MN_LO}) of the packets belonging to the session to the address of the active interface (IP_{MN_3G}). The NAT configuration in the AS allows to forward the packets received from the MN to the CN, changing the source IP address from IP_{MN_3G} to IP_{AS_CN}, the destination IP address from IP_{AS_MN} to IP_{CN}, and the transport ports, as illustrated in Fig.5.

After a certain time, the MN decides to perform a handover to the WLAN interface. In order to do so, the MN registers itself again with the core IMS with the new IP address obtained (IP_{MN_WLAN}).

Once the new IP address is registered within the IMS, the MN modifies the multimedia session by issuing a re-INVITE towards the AS, which is able to start sending the packets to the new MN address, by changing its NAT configuration to reflect the IP address in use in the MN (IP_{MN_WLAN}). In addition, the MN configures its NAT according to its new IP address, and also sets up the correct default route through the new interface.

4.2. UDP Setup

In order to validate the solution under the assumption of UDP traffic, we tested the architecture by sending from the CN to the MN a video stream of variable bit rate. The experiment set up and the different steps followed to perform the handover are the ones described in subsection 4.1.

Figure 8 shows a complete cycle of handover, starting in the 3G leg and moving to the WLAN for a few seconds before coming back to the 3G interface. The solid line in Fig. 8 shows the amount of traffic received through the 3G interface while the dotted line corresponds to the traffic received in the WLAN interface. It is important to note that in this experiment we were not saturating the network, since our aim was just to validate the proposal. The handover to WLAN in Fig. 8 occurs between the 66.5 and 67 seconds as can be seen in the close-up of the figure. The close-up also shows how packets are routed through the WLAN interface exactly at the same moment that the 3G interface drops to zero. As conclusion from this graph, in the case of UDP flows, our solution is able to achieve zero packet loss. In next section, where we worked with TCP traffic, we provide insights of

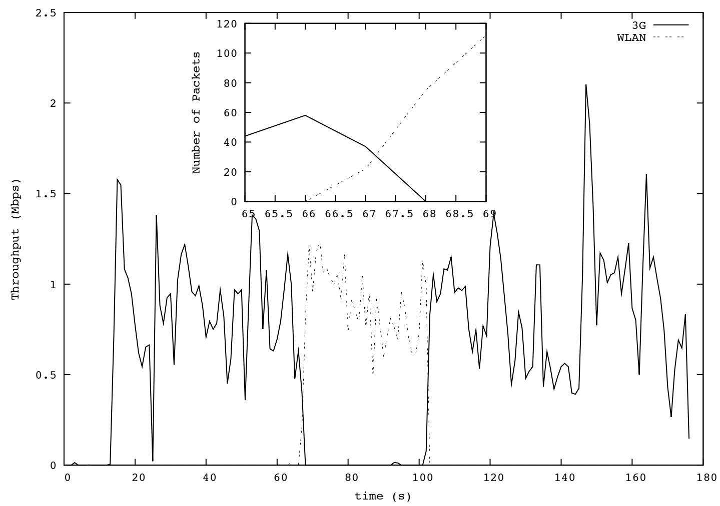


Figure 8: Handover of UDP data flow

how the nature of TCP traffic, the linux routing mechanisms and the NAT implementation complicate the prototype functionality.

4.3. TCP Setup

This section focuses on the validation of the proposed solution when TCP traffic is used. Unlike in the UDP case, in this scenario we only validate the user plane as a proof-of concept, hence the handovers were triggered manually without SIP intervention (the interaction with SIP was already tested in the UDP setup). In order to develop a prototype of the proposed solution we opted to use the NAT functionality of the Linux Kernel, as explained in section 4.1.

The NAT mechanism in Linux relies on the connection tracking (*conntrack*) properties of the kernel. The first time a packet (matching a certain rule in the NAT table) traverse the kernel, an entry in the *conntrack* system is created and no more packets of this certain flow transverse the NAT table. In order to modify the packets with the new IP addresses after a handover, we opted to remove the entry in the *conntrack* table each time a handover is performed. Due to TCP bidirectional nature, a *conntrack* entry matching the complete bidirectional flow is set, so once the *conntrack* entry has been removed, the reception of any TCP segment, data or ACK, prior to the setup of the new entry on the NAT table will register a new entry in the *conntrack* table. This leads to two workarounds in our prototype for

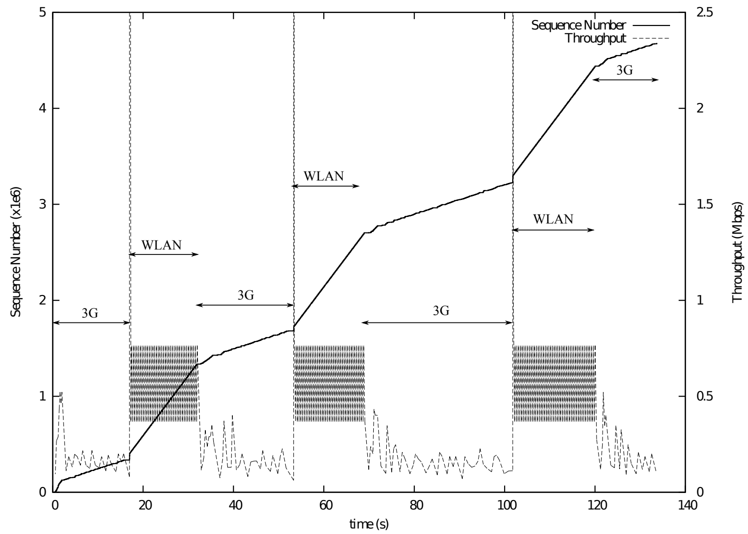


Figure 9: Handover of TCP data flow

TCP traffic: the removal of *conntrack* entries after the handover, and the shutdown of the old interface in the handover procedure, so we have hard handovers in the TCP scenario.

For the validation of the solution with TCP, we used the *iperf*[10] tool to generate the traffic from the CN to the MN. We restricted the traffic generated by the *iperf* tool to 500 Kbps. Figure 9 shows a Sequence number/Throughput vs Time plot with several handovers between 3G and WLAN following the steps mentioned in subsection 4.1. If we focus on the Throughput graph we can see how the bandwidth obtained from the WLAN is clearly higher than the one obtained while using 3G. The throughput in the WLAN is limited by the application rate, while in the 3G it is limited by the network bandwidth[11]. Note that after each handover from 3G to WLAN, there is a throughput peak caused by the sending of the packets buffered by TCP during the 3G period, which now can be delivered at a higher rate through the WLAN. The peak lasts for a short time because even if some packets are lost during the handover, the RTT is very small in the WLAN case (in the order of 2 ms), enabling a very fast increase of

[10]http://iperf.sourceforge.net

[11]The experiments for TCP and UDP were done in different days, and the achieved throughput through the 3G was different due to network conditions.

the congestion window (note also that selective acknowledgements and fast retransmit are used in our version of TCP).

4.4. Considerations about performance

This section covers a set of considerations about the performance achieved by TRIM. In this proposal, an intermediate element (an address translator located in the home network of the MN) is placed in the media path between the MN and the CN. The behavior of the address translator is similar to a home agent in Mobile IP, from the point of view that the data traffic has to go through the home network. Nevertheless, and although this may increase the end-to-end delay between the MN and the CN, it does not prevent real time communications. As an example, conferencing in IMS, as it has been defined by 3GPP, uses an intermediate element (i.e. the Multimedia Resource Function Processor) to receive, combine and redistribute media streams to the conference participants.

Another parameter to characterize the performance achieved by a mobility management solution is the handover delay. To estimate the handover delay achieved by TRIM, we assume that a MN is participating in a multimedia session with a CN via a given access network. In this situation, we define the handover delay, T_{ho}, as the time that elapses from the instant the UE moves to the new network until it starts receiving data through this new network. To estimate T_{ho}, we identify all the delays involved in the handover process:

- T_{attach} is the time necessary to attach the MN to the new network.
- $T_{act-sig}$ is the time necessary to activate a signaling channel in the new network (e.g. a primary PDP context in the case of UMTS). This delay includes the time to configure the MN with a new IP address, and to discover the address of the new P-CSCF (if a P-CSCF in the new network is to be used).
- $T_{register}$ gathers the delay corresponding to the IMS registration of the MN to the S-CSCF.
- T_{sip} gathers the delay corresponding to the INVITE transaction through the new network. This is the time that elapses from the instant the MN generates the INVITE request until it receives the corresponding OK response (after sending this response, the address translator starts forwarding the media belonging to the session to the new location of the MN). T_{sip} also includes the time necessary to execute the resource

reservation procedures in the new network (i.e. the time necessary to activate the transport bearers that are necessary to deliver the media).

Taking this delays into consideration, the handover delay can be estimated as: $T_{ho} = T_{attach} + T_{act-sig} + T_{register} + T_{sip}$.

It is important to highlight that this handover delay is the same that can be achieved with the IMS service continuity procedures defined by 3GPP [7], as these procedures comprise attaching and gaining IP connectivity to the new network, registering and initiating a new multimedia session towards a Service Centralization and Continuity AS[12] (SCC AS).

TRIM can achieve zero packet loss during soft handovers (see Sect.4.2). However, hard handovers may affect performance because, during the handover delay, we cannot continue using the old network while setting up the communication through the new one. It is necessary to develop optimizations to TRIM to achieve zero packet loss in hard handover scenarios. We plan to address this issue as a future work (see conclusions).

5. Related Work

Providing mobility support in IP networks is a well-known problem. At the IP layer, the Mobile IP protocols ([2] for IPv6, and [1] for IPv4) have been developed to provide mobility support. An alternative solution for SIP based services over IP is to provide the mobility by means of SIP signaling [6]. In a SIP based mobility solution, the mobility is not transparent for the application that has to deal with the change of address when one node moves. This has also an impact on the CN which is aware and has to do operations when the MN moves. A SIP based mobility solution has the advantage of not depending on having mobility functionality at the IP layer. This is specially important in IMS-based networks, because the integration of MIP in the IMS framework is complex and requires modifications in IMS.

In the following subsections we provide an overview of related works, differentiating between plain SIP mobility and IMS-based mobility.

[12] We do not consider here the delay of the signaling process executed between the SCC AS and the CN to update the session over the remote access, as this process could run in parallel to the session setup between the MN and the SCC AS. However, note that this signaling is not required in TRIM, as the CN always observes stable remote addressing information for the MN.

5.1. Mobility with plain SIP

In recent years there has been several works studying terminal mobility using SIP such as the seminal work in [13], performance studies like [14], support of mobility across heterogeneous domains [15], and the use of Mobile IPv6 and SIP in heterogeneous networks [16].

The works in [17] and [18] are very related to our proposal. The authors propose to use an intermediate node to hide the mobility of the MN to the CN. The SBC could be a B2BUA (Back-to-back User Agent) or a special SIP proxy so both the MN and the CN exchange signaling through this intermediate entity. All media flows pass through the intermediate node, again to hide mobility to the CN. Nevertheless, in this proposal the mobility is not transparent for the application in the MN, which has to deal with the change of address when moving.

5.2. Mobility in IMS-based networks

Faccin *et al.* [19] is one of the firsts works tackling the problem of integrating MIP in IMS. This work covers several scenarios combining IMS in 3GPP and IMS in 3GPP2 with MIPv4 and MIPv6, but there are several challenges in order to integrate MIP in IMS as will be discussed next.

The addressing considerations when using MIP with SIP are discussed in [19] with an analysis of the usage of the HoA[13] or the CoA[14] at the MN. The HoA should be used both for SIP signaling and for session establishment (data plane) to obtain the advantages of transparent mobility that MIP provides. The CoA would be transparent to the application layer (both for SIP and the application) and MIP would take care, at the IP layer, of the matching between the CoA used to send and receive traffic in the MN and the HoA. Unfortunately, in IMS we cannot simply use the HoA, because resources are reserved in the access network for the traffic corresponding to the SIP service. Because packets coming to the MN have to have the CoA as the destination IP address, and packets coming from the MN have to have the CoA as source address, in both cases with the HoA present in a special extension header (IPv6) or in an inner IP header (when a tunnel is used in IPv4 or in IPv6) the P-CSCF needs to know the CoA as well as the HoA in order to install the proper filters in the network edge, so it

[13] Home Address, permanent address used by the MN and topologically valid in its home network.

[14] Care-of Address, temporal address used by the MN while visiting a network and topologically valid in that network.

is not enough that SIP manages only the HoA. In [19] authors argue that with MIP for IPv6 (MIPv6), the MN could use MIPv6 signaling (a Binding Update) to inform the P-CSCF of the association CoA-HoA. This is true at the IP layer, but we would still have the problem that SIP would be unaware of the CoA. This problem was already studied in [3] for cdma2000 and WLAN access, and in [4, 5] for GPRS/UMTS access. They basically propose to have fully MIP-aware P-CSCFs and access routers introducing an interface between the MIP binding tables and the proper functionalities, but this means introducing changes in the IMS specifications.

Related with the previous item, there is also a problem with the registration process as, when using MIP, the MN has to register the HoA. The P-CSCF will reject the registration because the HoA is not a valid IP address inside its access network domain.

Faccin *et al.* [19] also deal with the return routability (RR) test when using MIPv6. In RR, the CN challenges the MN to verify its reachability using the CoA. In order to do so, the CN sends challenges to the CoA and HoA so the MN receives these two messages and sends a response using both challenges. This test is performed at a certain frequency (a few minutes) so there is an extra overhead in the communication between the MN and the P-CSCF in case the P-CSCF behaves like a CN as proposed by the authors.

In [20], the authors propose supporting mobility for IMS-based IPTV peer-to-peer services by combining MIP with IMS, and using a context transfer mechanism to support efficient handovers. This proposal requires modifications to the IMS infrastructure to support the integration of MIP.

Another approach to handle mobility has been proposed by 3GPP in [7]. This specification describes a service that supports the use of session transfer mechanisms to maintain service continuity in the event of terminal mobility, for the case when this mobility is not transparent to the IMS session. In this solution, session transfer procedures are initiated by the MN and are handled by a Service Centralization and Continuity Application Server (SCC AS), which fits in the signaling path between the MN and the CN. After obtaining IP connectivity in a new network, the MN can preserve the continuity of an active multimedia session by starting the session transfer procedures. Specifically, it registers to the S-CSCF from the new network and it initiates a new multimedia session towards the SCC AS by sending an INVITE request. The MN includes in this request an SDP payload, with its current addressing information in the new network. Upon receiving the INVITE request, the SCC AS matches the new multimedia session with the session to be transferred, and modifies the remote leg of the session providing the CN with the SDP payload indicated by the MN. This way, the new location of the MN is available to the CN and session continuity can be maintained.

Nevertheless, mobility management procedures described in this technical specification are not transparent to applications at the MN and the CN. Specifically, when the MN moves to a new network, its local applications must generate a new SDP description and trigger a new session setup towards the SCC AS. Applications running at the CN must be contacted to update the addressing information corresponding to the MN. In addition, any network socket bound to the previous IP address of the MN becomes invalid. Therefore, applications at the MN and CN may need to re-configure or to open new network sockets. This is particularly problematic in the case of TCP transport, as active TCP connections need to be re-established.

Compared with the related work above, our proposal has the advantage of not requiring a MIP deployment, therefore avoiding the problems of integrating MIP in IMS-based networks, while keeping mobility transparent to the applications. In addition, it does not require any modifications to the IMS infrastructure. In the home network, the proposal introduces a SIP application server (i.e. the TRIM AS) and an address translator. In the UE, the proposal only requires some new functions at the MN, but those can be installed as a simple software upgrade.

6. Conclusion

This paper proposes TRIM, an architecture to provide mobility support in IMS-based networks. TRIM is based on SIP signaling, but unlike other approaches to offer mobility support in IP networks based on SIP, TRIM makes mobility transparent to applications, which do not have to do any operation to support a change of access network by the mobile node. This functionality is similar to the one provided by Mobile IP, but the compatibility of Mobile IP and IMS requires modifications to the IMS specifications, a requirement that TRIM does not have. We have tested a prototype of TRIM in a testbed combining 3G and IEEE 802.11 access technologies, and an IMS core. In the experiments we showed that TRIM correctly supports mobility both for UDP and TCP user traffic.

There are several lines of future research for TRIM. TRIM forwards the user traffic through a network element in the mobile node home network, which is similar to Mobile IP without route optimization. We want to explore an optimization to place this network element within or close to the network being visited by the mobile node. Additionally we want to study the performance of handovers when only one network interface is available in the mobile node and soft handovers are not possible. An approach to improve performance could be to develop optimizations based on buffering in the intermediate network element.

Acknowledgements

This article has been partially granted by the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement 214994 (CARMEN project), and by the Madrid Community through the MEDIANET project (S-2009/TIC-1468).

References

[1] C. Perkins, IP Mobility Support for IPv4, RFC 3344, Internet Engineering Task Force (Aug. 2002).

[2] D. Johnson, C. Perkins, J. Arkko, Mobility Support in IPv6, RFC 3775, Internet Engineering Task Force (Jun. 2004).

[3] T. Chiba, H. Yokota, A. Dutta, D. Chee, H. Schulzrinne, Performance analysis of next generation mobility protocols for IMS/MMD networks, in: Wireless Communications and Mobile Computing Conference, 2008. IWCMC'08. International, IEEE, 2008, pp. 68–73.

[4] T. Renier, K. Larsen, G. Castro, H. Schwefel, Mid-session macro-mobility in IMS-based networks, IEEE Vehicular Technology Magazine 2 (1) (2007) 20–27.

[5] X. Chen, J. Wiljakka, Problem Statements for MIPv6 Interactions with GPRS/UMTS Packet Filtering, Internet draft, Internet Engineering Task Force (2007).

[6] H. Schulzrinne, E. Wedlund, Application-layer mobility using SIP, ACM SIGMOBILE Mobile Computing and Communications Review 4 (3) (2000) 57.

[7] 3GPP, IP Multimedia Subsystem (IMS) service continuity; Stage 2, TS 23.237 v9.6.0 Release 9, 3rd Generation Partnership Project (3GPP) (Sep. 2010).

[8] 3GPP, IP Multimedia Subsystem (IMS); Stage 2, TS 23.228, 3rd Generation Partnership Project (3GPP) (Mar. 2010).

[9] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, SIP: Session Initiation Protocol, RFC 3261, Internet Engineering Task Force (Jun. 2002).

[10] M. Handley, V. Jacobson, C. Perkins, SDP: Session Description Protocol, RFC 4566, Internet Engineering Task Force (Jul. 2006).

[11] 3GPP, Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3, TS 24.229 v9.3.1 Release 9, 3rd Generation Partnership Project (3GPP) (Sep. 2010).

[12] F. Brockners, S. Bhandari, V. Singh, V. Fajardo, Diameter Network Address and Port Translation Control Application, Internet draft, IETF (2010).

[13] E. Wedlund, H. Schulzrinne, Mobility support using SIP, in: Proceedings of the 2nd ACM international workshop on Wireless mobile multimedia, ACM, 1999, pp. 76–82.

[14] N. Nakajima, A. Dutta, S. Das, H. Schulzrinne, Handoff delay analysis and measurement for SIP based mobility in IPv6, in: Communications, 2003. ICC'03. IEEE International Conference on, Vol. 2, IEEE, 2003, pp. 1085–1089.

[15] M. Cardenete-Suriol, J. Mangues-Bafalluy, M. Portoles-Comeras, M. Requena-Esteso, M. Gorricho, VoIP performance in SIP-based vertical handovers between WLAN and GPRS/UMTS networks, in: Communications, 2007. ICC'07. IEEE International Conference on, IEEE, 2007, pp. 1973–1978.

[16] M. Bernaschi, F. Cacace, G. Iannello, M. Vellucci, Mobility management for VoIP on heterogeneous networks: evaluation of adaptive schemes, IEEE Transactions on Mobile Computing (2007) 1035–1047.

[17] S. Salsano, L. Veltri, A. Polidoro, A. Ordine, Architecture and testbed implementation of vertical handovers based on SIP session border controllers, Wireless Personal Communications 43 (3) (2007) 1019–1034.

[18] E. Boysen, H. Kjuus, T. Maseng, Proactive Handover in Heterogeneous Networks Using SIPs, in: Seventh International Conference on Networking, IEEE, 2008, pp. 719–724.

[19] S. Faccin, P. Lalwaney, B. Patil, IP multimedia services: analysis of mobile IP and SIP interactions in 3G networks, Communications Magazine, IEEE 42 (1) (2005) 113–120.

[20] I. Vidal, J. Garcia-Reinoso, A. de sla Oliva, A. Bikfalvi, I. Soto, Supporting mobility in an IMS-based P2P IPTV service: a Proactive Context Transfer mechanism, Computer Communications 33 (14) (2010) 1736–1751.