HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Electrical and Communications Engineering
Telecommunications Software and Multimedia Laboratory

Anna Larmo

Usability of Instant Messaging over WCDMA

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Engineering

Espoo, 26.1.2005

Supervisor Professor Petri Vuorimaa
Instructor M.Sc. Stefan Wager

TEKNILLINEN KORKEAKOULU

Diplomityön tiivistelmä

Tekijä:	Anna Larmo
Työn nimi:	Pikaviestinnän käytettävyys WCDMA:n yli
Päivämäärä:	26.1.2005 **Sivumäärä:** 60
Osasto:	Sähkö- ja tietoliikennetekniikan osasto
Professuuri:	T-111 Vuorovaikutteinen digitaalinen media
Työn valvoja:	Petri Vuorimaa, Professori
Työn ohjaaja:	Stefan Wager, DI

Diplomityössä tutkitaan pikaviestinnän käytettävyyttä kolmannen sukupolven radioverkon, laajakaistaisen CDMA:n (WCDMA), yli käytettävyysarviointien avulla. Pikaviestintä, eli lähes reaaliaikainen tekstipohjainen keskustelu, on suosittu viestintätapa Internetissä. Jotta pikaviestejä voi vaihtaa, on tärkeää tietää mikä on vastapuolen läsnäolotila, eli halukkuus ja mahdollisuus viestiä.

Aluksi työssä esitellään WCDMA verkon rakenne. Esittelyssä keskitytään erityisesti radiorajapinnan ominaisuuksiin ja hallintaan. Tämän jälkeen esitellään työssä käytetty emulaattori ja siinä käytetyt kaksi erilaista WCDMA kokoonpanoa.

Lyhyt johdanto pikaviestintään ja tilapalveluihin esittelee suosituimmat pikaviestiohjelmat lyhyesti ja kertoo standardisointi rintamalla tehdystä työstä avoimen standardin aikaansaamiseksi. Suuren kannatuksen saanut Session Initiation Protocol (SIP) on yksi tärkeimmistä pikaviestinnän mahdollistavista avoimista ratkaisuista.

Pikaviestimien toimintakykyä testataan ja analysoidaan WCDMA:n yli. Tässä osiossa vertaillaan suositun pikaviestiohjelman ja avoimeen SIP määritelmään perustuvan ohjelman suoriutumista yksinkertaisten käyttötapausten avulla.

Käytettävyysarviointi suoritetaan lopuksi kahdeksan käyttäjän kanssa, jotka ovat jo käyttäneet pikaviestintää aiemmin. He testaavat pareittain pikaviestinnän toimivuutta kolmen eri verkkokokoonpanon yli. Esikyselyn, käytettävyysarvioinnin ja jälkikyselyn tulokset esitellään.

Avainsanat: pikaviestintä, läsnäolo, WCDMA, käytettävyys

HELSINKI UNIVERSITY OF TECHNOLOGY

Abstract of the Master's Thesis

Author:	Anna Larmo
Name of the Thesis:	Usability of Instant Messaging over WCDMA
Date:	26.1.2005 **Number of pages:** 60
Department:	Department of Electrical and Communications Engineering
Professorship:	T-111 Interactive Digital Media
Supervisor:	Petri Vuorimaa, Professor
Instructor:	Stefan Wager, M.Sc. (Tech.)

In this Master's thesis, the usability of instant messaging (IM) over the 3rd generation radio network WCDMA is studied with usability evaluations conducted with real users. IM is currently used by hundreds of millions of people as a communication method in the fixed Internet.

First, the network architecture of WCDMA is presented with focus on the radio interface control features. This is followed by the introduction of the set-ups used later in the usability evaluations.

IM and presence services are explained, and the most popular applications are introduced as well as the work done in the standardizing organizations. Session Initiation Protocol (SIP) is one of the main efforts to produce a joint standard for IM and presence.

A technical performance analysis for IM over WCDMA is done. In this analysis, a popular IM application is compared to an application based on the open SIP specification.

The usability evaluations are then made with users, who are already familiar with IM. The users evaluate IM over three different network set-ups. The preliminary survey, evaluation and after test questionnaire results are presented.

Keywords: instant messaging, presence, WCDMA, usability

Acknowledgements

This thesis was made in Nomadiclab, the research unit of Oy L M Ericsson Ab, Finland.

I would like to thank my previous manager Janne Peisa for giving me this opportunity to write the thesis on this interesting subject, and my section manager Johan Torsner for supporting me to finish it. I, also, owe my greatest thanks to my instructor Stefan Wager who guided me through this process, and thought me a great deal about making research.

I thank Professor Petri Vuorimaa for supervising my thesis.

Finally, I want to thank my parents and Ville for their love and support during this work.

Kirkkonummi, 26th January, 2005.

Anna Larmo

Table of Contents

TABLE OF FIGURES

Abbreviations

3GPP	3rd Generation Partnership Project (WCDMA standardization)
3GPP2	3rd Generation Partnership Project 2 (Cdma2000 standardization)
AN	Access Network
AOL	American On Line
ARQ	Automatic Resend Query
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CN	Core Network
DCH	Dedicated Channel
DECT	Digital Enhanced Cordless Telephone
DL	Downlink
EDGE	Enhanced Data GSM Environment
ETSI	European Telecommunications Standards Institute
FACH	Forward Access Channel
FDD	Frequency Division Duplex
GSM	Global System for Mobile Communications
GPRS	General Packet Radio Service
HTTP	HyperText Transfer Protocol
ICQ	"I seek You"
IETF	Internet Engineering Task Force
IM	Instant Messaging
IMPS	Instant Messaging and Presence Service
IMS	IP Multimedia Subsystem
IMT-2000	International Mobile Telecommunications, 3rd generation networks are referred as IMT-2000 within ITU.
IP	Internet Protocol
ITU	International Telecommunications Union
MSS	Mobile Satellite Systems
NAT	Network Address Translator
OMA	Open Mobile Alliance
OSI	Open Systems Interconnect
PCS	Personal Communication Systems, 2nd generation cellular systems mainly in Americas.
PSTN	Public Switched Telephone Network

QoS	Quality of Service
RACH	Random Access Channel
RFC	Request For Comments
RLC	Radio Link Control
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RTT	Round Trip Time
SCIP	Simple Conference Invitation Protocol
SDU	Service Data Unit
SIMPLE	SIP for Instant Messaging and Presence Leveraging Extensions
SIP	Session Initiation Protocol
SIPv1	Session Invitation Protocol
SMS	Short Message Service
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TIA	Telecommunication Industry Association
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunication Services
UTRA	UMTS Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
WCDMA	Wideband Code Division Multiple Access
WWW	World Wide Web

1. Introduction

Instant messaging (IM) is an increasingly popular communication method. There are numerous applications available on the Internet that make it possible to communicate with other people all around the world in almost real-time. Ordinarily, IM is used through a client application on a computer with an Internet connection. Some IM applications enable sending and receiving of instant messages to and from mobile phones, which makes it a mobile service as well.

IM is usually text-based communication between two users, although it is possible to have more people messaging in a group as well. Thus, it can be seen as a kind of private chat room. IM is however not chat; users can choose the people they want to communicate with from a private buddy list they have made themselves. The users outside this list can be denied of contacting.

Usually, the client application shows, which of the user's contacts are online, so that the user knows with whom it is possible to communicate, even before starting an IM session. This so-called presence information can also contain user specified data about his mood and location, besides the availability information.

IM and presence services are currently extremely popular in the Internet world, and are already used to a small extent in the second generation (2G) cellular networks via mobile data services, such as General Packet Radio Service (GPRS). Thus, they are important application candidates for third generation (3G) mobile networks. It is important to see how well the 3G network is suited for these services.

Different networks will cause different amounts of delay to the data transmissions, which will influence the user's experience of the Quality of Service (QoS). In Europe, the main technique that has been chosen for 3G network operation is Wideband Code Division Multiple Access (WCDMA).

In this thesis, we have concentrated on the WCDMA delays and their impact on the IM service. First, we have made a technical comparison of the delay values for a limited set of use cases. Then, these use cases were tested with real users with different network configurations. The usability of the IM applications used in the test sessions is out of the scope of this study.

The user group that was evaluating the usability aspects was chosen among young adults already using IM services, because IM is particularly popular as a communications channel among these age groups. It's to be expected that most likely the first users of IM over WCDMA will be the people who already use it now over the fixed Internet.

1.1. Content of the Thesis

In Chapter 2, the 3G network architecture and the WCDMA technology is shortly introduced. The emulator used in this study and the used measurement set-ups are also illustrated and explained. Chapter 3 introduces the concepts of IM and presence services. We look in to the background of theses services, introduce some currently used applications, and finally, talk about standardization work being done in this front. A technical performance analysis is made in Chapter 4. Two different solutions for IM and presence services are tested over two WCDMA set-ups introduced in Chapter 2. The usability evaluations and their results are described and explained in Chapter 5. First, the requirements are specified, then the test organization and progression is described and finally conclusions are drawn from the analysis made on the results. Chapter 6 concludes this thesis.

2. WCDMA Overview

WCDMA is a 3G radio access technology. It is a part of the 3G mobile communications systems, often referred to as Universal Mobile Telecommunications Services (UMTS). 3G converges the telecom, computer and media industries to form a type of multi-service network of networks.

In the first deployment of WCDMA, Release 99 (Rel-99), the bit rates for data transfer are up to 384 kbps in downlink (DL) direction. Later on, in release 5, it will be possible to have rates up to 12 Mbps in DL. 3G will make it possible to use the Internet more efficiently on the move and will also enable, e.g., video calls.

2.1. Background

The evolution of services from the first generation (1G) mobile communication systems to the current 3G systems is depicted in Figure 2-1.

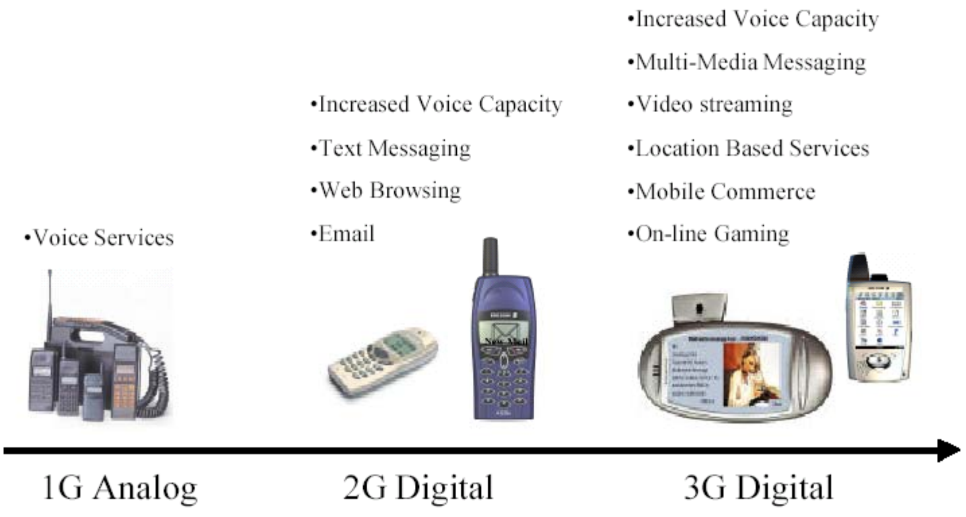

Figure 2-1. Application and devices introduced along with the technology evolution (source: Ericsson[1]).

The development of the 3G mobile systems started already in 1992, when the World Administrative Radio Conference (WARC) of the International

[1] http://www.ericsson.com/cdmasystems/pdf/products/CPPArticle.PDF (31.12.2004)

Telecommunications Union (ITU) identified the frequencies available for future use. These 3G systems are called International Mobile Telephony 2000 (IMT-2000) within the ITU. There are three different 3G systems defined within the IMT-2000 family, although the original goal was to produce a single global air interface for all systems. The three systems are explained briefly in Figure 2-2.

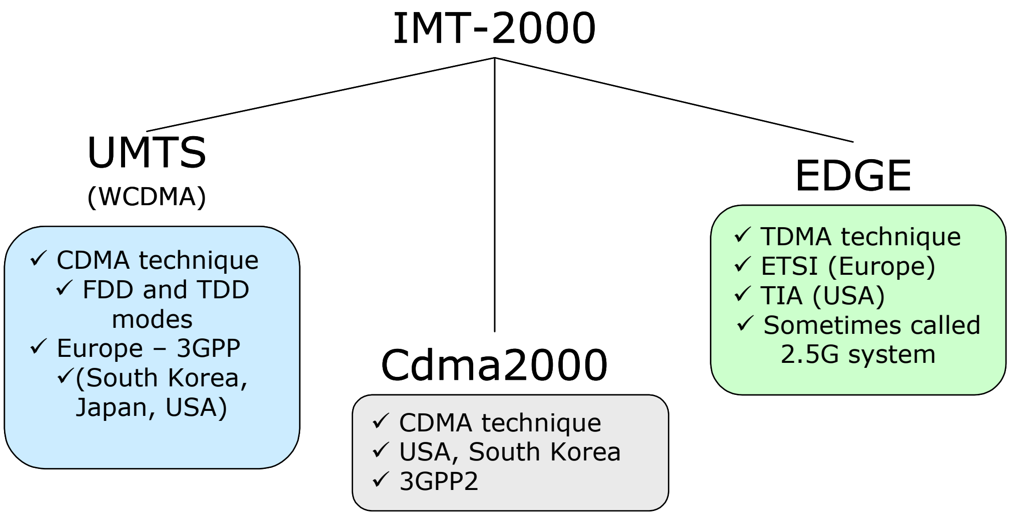

Figure 2-2. 3G systems around the world.

The most widely spread of the air interfaces is WCDMA, which is used, e.g., in Europe and most of Asia. WCDMA has two modes of operation: Frequency Division Duplex (FDD) and Time Division Duplex (TDD), but so far only the FDD mode has been deployed. The 3rd Generation Partnership Project (3GPP) standardizes WCDMA worldwide, and the European Telecommunications Standards Institute (ETSI) in Europe. Within 3GPP, WCDMA is called the Universal Terrestrial Radio Access (UTRA).

In the North America, the frequencies for WCDMA were not available when the decision to move to 3G was made. Thus, they have their own air interface called CDMA2000. CDMA2000 will be developed from the currently used CDMAone (also called IS-95). 3rd Generation Partnership Project 2 (3GPP2) standardizes the CDMA2000-technique.

The Enhanced Data GSM Environment (EDGE) is based on the Global System for Mobile Communications (GSM) standard. It brings faster data transfer rates to GSM. Just like GSM, EDGE is based on Time Division Multiple Access (TDMA) technique. EDGE is sometimes called a 2.5G technique, because its deployment does not require building of new networks.

2.2. Spectrum Allocation

The spectrum allocation for 3G systems was originally made at the 1992 WARC. The spectrum allocation for Europe, China, Japan, Korea and North America is shown in Figure 2-3.

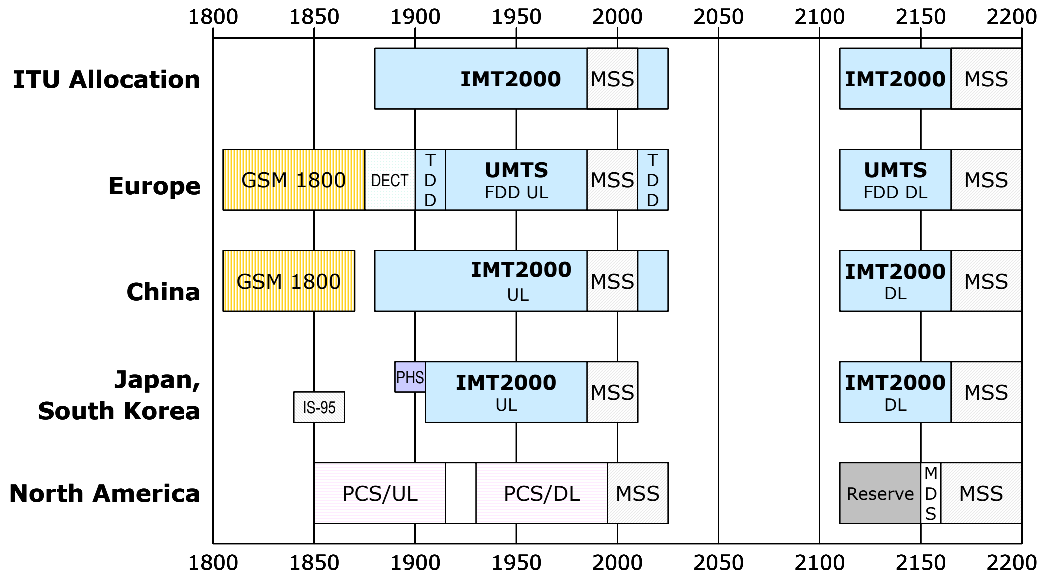

Figure 2-3. WARC '92 spectrum allocation.

WCDMA FDD requires separate 5 MHz frequency bands for uplink (UL) and DL, and a guard band in between them. WCDMA TDD requires only one 5 MHz frequency band, because UL and DL operation is separated with a guard time between the timeslots. In 1998, ETSI chose the WCDMA paired FDD bands at 1920-1980 MHz (UL) and 2110-2170 MHz (DL) for the UMTS public, wide-are service. The unpaired TDD bands at 1900-1920 MHz and 2010-2025 MHz were chosen for private indoor services [Hol 00].

Mobile Satellite Services (MSS) were allocated spectrum at the 1992 WARC, because it was seen as an area with growing needs. The MSS will, supposedly, constitute the satellite component of IMT-2000 [Man 00].

2.3. 3G Network Architecture

Based on the 3GPP reference network model, UMTS consists of four main parts listed below and shown in Figure 2-4.

- User Equipment (UE)
- Access Network (AN)

- Core Network (CN)
- External Network (e.g., the Internet, Public Switched Telephone Network (PSTN))

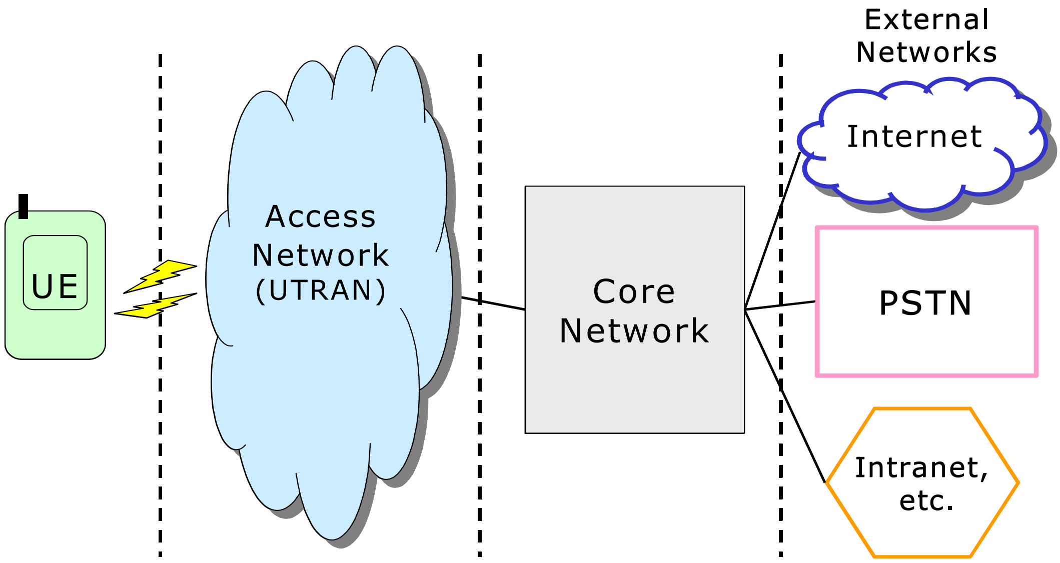

Figure 2-4. The 3GPP network reference model simplified.

The UE is the users device that enables the user to connect to the WCDMA network. AN is usually called WCDMA Radio Access Network (RAN). In 3GPP and ETSI, RAN is also called with the acronym UTRAN (UMTS Terrestrial Radio Access Network). CN is the part that joins the UMTS system with other networks like PSTN, the Internet, etc. CN is normally common with GSM. During 3G startup the two systems are run side by side.

2.3.1. Radio Access Network Nodes

The WCDMA RAN is shown in Figure 2-5. It consists of Node Bs (name for Radio Base Stations in 3GPP), which provide the radio resources and convert the data flow between the Iub and Uu interfaces, and Radio Network Controllers (RNC), which control the Node Bs and the radio resources. The RNC provides all WCDMA RAN services to the CN. One RNC can control several Node Bs.

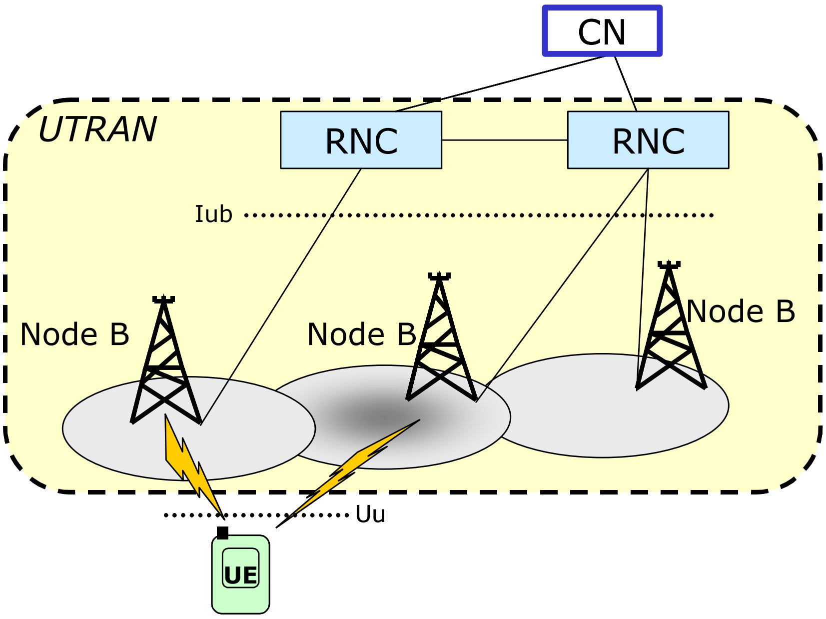

Figure 2-5. The WCDMA RAN (UTRAN) with the relevant interfaces.

2.4. Radio Resource Control Protocol

The signalling between UE and UTRAN is mainly Radio Resource Control (RRC) messages. For example, the mobility of the UE in the connected mode is controlled by RRC signalling. This includes, e.g., measurement, cell updates and handovers. RRC signalling also does paging of idle mode UEs and broadcasting of system information.

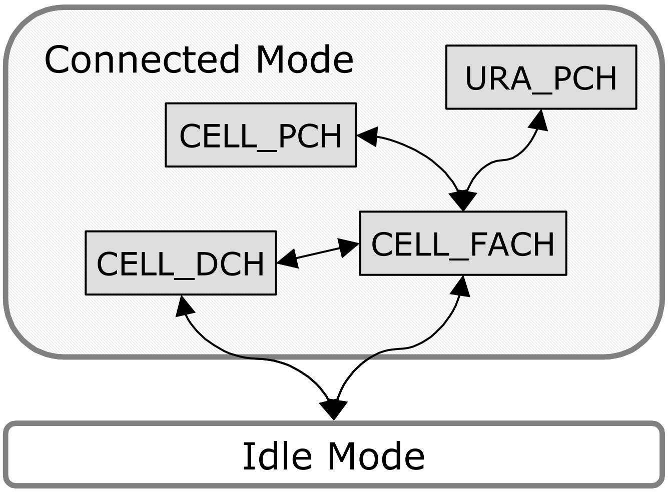

Figure 2-6. UE modes and RRC states [Hol 00].

An UE has two basic modes, shown in Figure 2-6: idle mode and connected mode. The RRC has four possible states in the connected mode: CELL_DCH, CELL_FACH, CELL_PCH and URA_PCH. In the state names, CELL means the current cell the UE is located in. URA is an acronym for UTRAN Registration Area. The state names are explained in the next section. The ways to switch between these states were shown with arrows in Figure 2-6.

In the idle mode, the UE is switched on, and it is able to receive system information and broadcast messages. There is no RRC connection at this point between the UE and UTRAN and UTRAN does not have any information about the idle mode UEs. The UTRAN can, however, contact all idle mode UEs through, e.g., paging.

When the UE is in connected mode, it has an RRC connection open to the RNC. The RNC can locate the UE inside a cell or UTRAN Registration Area (URA). In some RRC states, the UE can transmit and receive data.

2.4.1. FDD Channel Structure

There are two kinds of transport channels in WCDMA FDD: common channels and dedicated channels. They are listed in Table 2-1 [Dah 98].

Table 2-1. WCDMA transport channels.

Common Channels		***RRC State***
BCCH	Broadcast Control CHannel (DL)	CELL_FACH
FACH	Forward Access CHannel (DL)	CELL_FACH
PCH	Paging CHannel (DL)	URA / CELL_PCH
RACH	Random Access CHannel (UL)	CELL_FACH
Dedicated Channels		
DCH	Dedicated CHannel (DL & UL)	CELL_DCH

In the CELL_DCH state, the UE has been allocated dedicated physical channels. The Dedicated Channel (DCH) is mapped on to two physical channels: Dedicated Physical Data Channel (DPDCH) and Dedicated Physical Control Channel (DPCCH). The user data is sent on the DPDCH, which can provide the user different transfer rates depending on the user's needs, distance from the Node B and level of mobility. The DCH is characterized by fast power control, the possibility of fast rate change (every 10 ms) and inherent addressing of UEs [Hol 00].

In the CELL_FACH state, the UE is allocated no dedicated channel, but it can use the Forward Access Channel (FACH) in downlink and the Random Access Channel in uplink (RACH) instead, which are shared between users. A low bit rate data transfer is possible in both directions. FACH is used to carry control information to the UE when the system knows its location cell. RACH is used to carry control information from the UE to the Node B. RACH is characterized by a collision risk.

The CELL_PCH state is a state where the UE is still known on a cell level, but it can only be reached via the Paging Channel (PCH). The PCH is always transmitted over the entire cell.

URA_PCH is quite similar to CELL_PCH, except that this time the UE does not have to be inside a specific cell, but inside a specific URA. The paging is this time transmitted over the entire URA.

2.4.2. Radio Resource Management

Radio Resource Management (RRM) controls the radio resources. It is located in RNC in Figure 2-5. One important task that RRM does is the channel switching function. This means controlling which UE has which transport channel. The possible channels were shown in Table 2-1. RRM uses RRC protocol messages to manage the UEs.

2.5. WCDMA Link Emulator

A WCDMA link emulator tool was used as the radio network part in order to perform the measurements and usability evaluations. The emulator modeled an unloaded, single cell radio network. It computed IP packet delays over a UMTS WCDMA bearer using Radio Link Control (RLC) in acknowledged mode. The module used Service Data Unit (SDU) delay statistics acquired through simulations, to calculate the delay and throughput through the emulated channel.

2.5.1. The Queuing model and the RNC Buffer

The transmission delay over any packet switched network is dependent on the traffic load. Assigning delays from delay distributions of recorded traffic is therefore not sufficient since it does not consider the simultaneous traffic load. The approach in this emulator is to combine simulated SDU transmission times with a queuing model. The delay simulations were performed for isolated SDUs with no queuing in the system, which meant that they did not include any queuing delay. The queuing delay was captured by the queuing model, which is described here.

The transmission of an SDU can be divided in two phases, the initial transmission phase and the retransmission phase (cf. Figure 2-7). In the figure, T_A is the arrival time of the SDU, T_S is the time when the initial transmission phase of the SDU has ended, and T_E is the release time of the SDU.

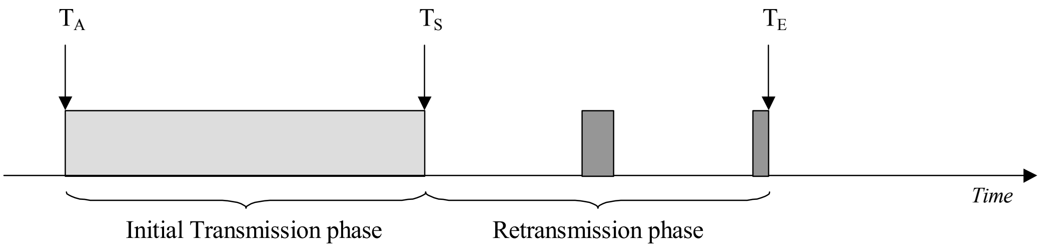

Figure 2-7. Transmission phases of a SDU.

Characteristic of any selective repeat Automatic Resend Query (ARQ) protocol is that the transmission of the next SDU can begin immediately after the initial transmission phase of the previous SDU has ended (T_S), although the previous SDU might need to be retransmitted before successfully received. Since retransmissions have higher priority than initial transmissions, retransmissions will not be affected by later SDUs, except in cases where retransmissions fill an entire Transmission Time Interval (TTI). Later SDUs, however, will be affected by retransmissions of previous SDUs, which will be experienced as a lower available bit rate, assuming a constant link rate.

To get a more correct behavior to the emulator, the module included some RNC functionalities. Figure 2-8 shows how the RNC buffer worked in conjunction with the queuing model described earlier.

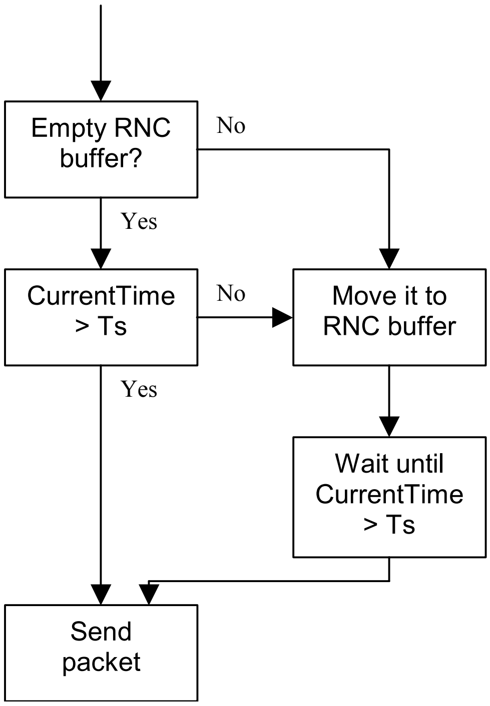

Figure 2-8. A model of the conjunction between the RNC buffer and the queuing model.

2.5.2. Measurement set-ups

The emulator was run in interactive mode, meaning that channel switches were performed when needed due to measured traffic or buffer size. The channel switch scenario and the execution delay times for channel switches are shown for each set-up in the following sections. In the notation, the UL and DL values are separated with a slash, and the UL value is shown first (UL/DL).

The next two sections introduce the two set-ups that will be used in this study. For each set-up, the relevant UE states are shown. Also, transmission bit rate is shown. The states are connected with arrows, showing what kinds of transitions are possible. The trigger times and buffer sizes for transitions are shown next to the arrows they are connected to.

2.5.3. Set-up A: Performance Optimized

The first set-up was configured so that it offered optimized performance for the application. The RRC states, used transport channels and delay and threshold values are shown in Figure 2-9.

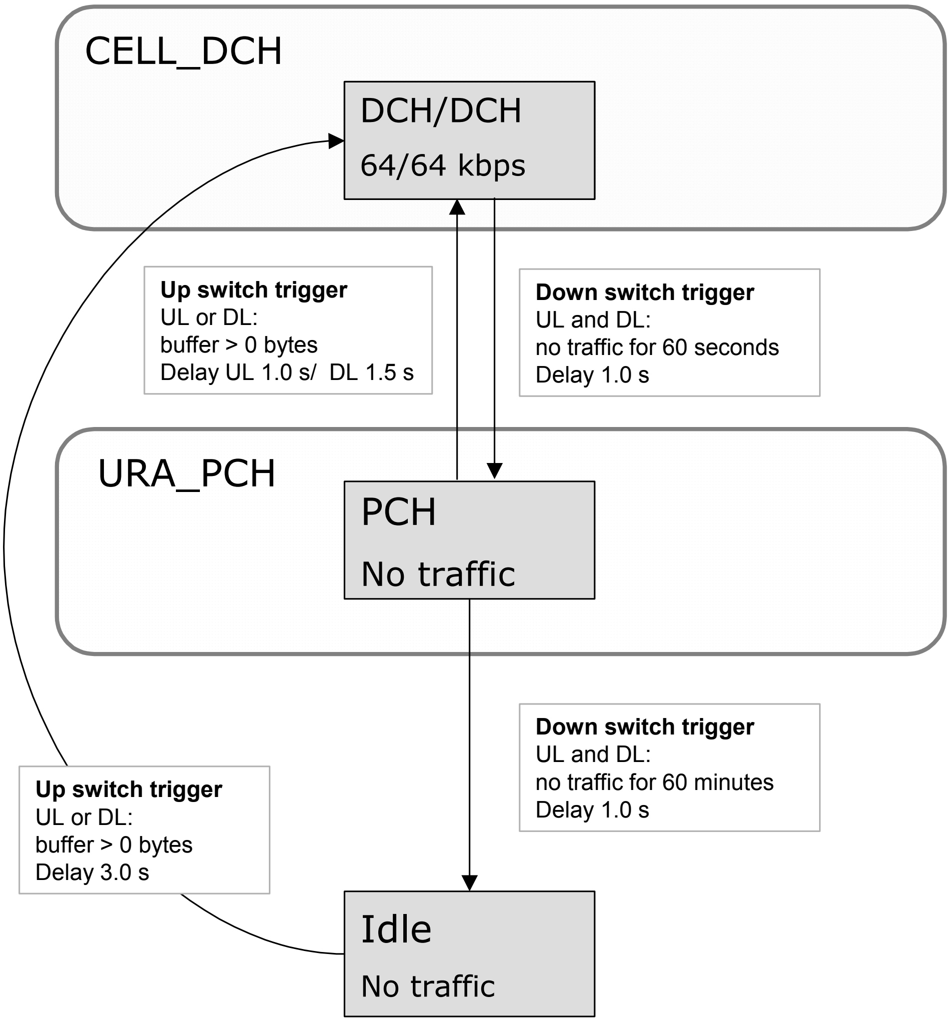

Figure 2-9. Set-up A, performance optimized. The modes, RRC states and transport channels with transition delays.

The CELL_FACH state, which is between the CELL_DCH and URA_PCH states, was used as a transition state, in other words the UE just passes through that state to URA_PCH and no data is exchanged on that channel. The CELL_FACH state is, therefore, not shown in the figure. URA_PCH was used, because the UE could be there for long periods of time without sending any messages. Also, the receiving and decoding of control messages, which would be required on CELL_FACH, could be left out. The most important thing, making this set-up optimized for the application performance, is that the UE can stay for 1 minute in CELL_DCH state even if it is not sending anything. This means that no time-consuming channel switch is needed, if data is sent 1-59 seconds after the last activity.

The up switch from URA_PCH to CELL_DCH (via CELL_FACH) takes longer in DL than UL. This happens because, if the first packet comes in DL, the RNC must first locate the UE by paging it.

The idle mode is used, if the UE is silent for a very long period of time, set to one hour in this configuration. This means that, e.g., at night when the UE is not used, it will switch down to idle mode to save power and network resources.

From the idle mode the UE is again switched up to CELL_DCH state in the connected mode, if needed. The state transition takes a long time (three seconds here), because the UE has no RRC connection when it is in idle mode. When the UE wants to switch to the connected mode, the connection must be first built.

2.5.4. Set-up B: Network Capacity Optimized

The second set-up is optimized from the network capacity point of view. The states and transitions are shown in Figure 2-10.

The starting point for the UE in this configuration, just as in the last one, is the idle mode. When the UE starts either sending or receiving data, it is switched up to the CELL_DCH state. Because this set-up is optimized from the network capacity point of view, the UE is switched down from CELL_DCH if it is not sending or receiving anything for 1 second. There are only a limited number of DCH channels in CELL_DCH state, whereas almost unlimited number of users can be put to CELL_FACH.

From the CELL_DCH state the UE is switched down to CELL_FACH, where it has a shared traffic channel with other users in the cell. In CELL_FACH, the traffic channel can transmit a maximum of 32 kbit/s for all users together. This is not very fast, but if the UE happens to be the only one transmitting in the CELL_FACH state, it gets the whole bandwidth to itself.

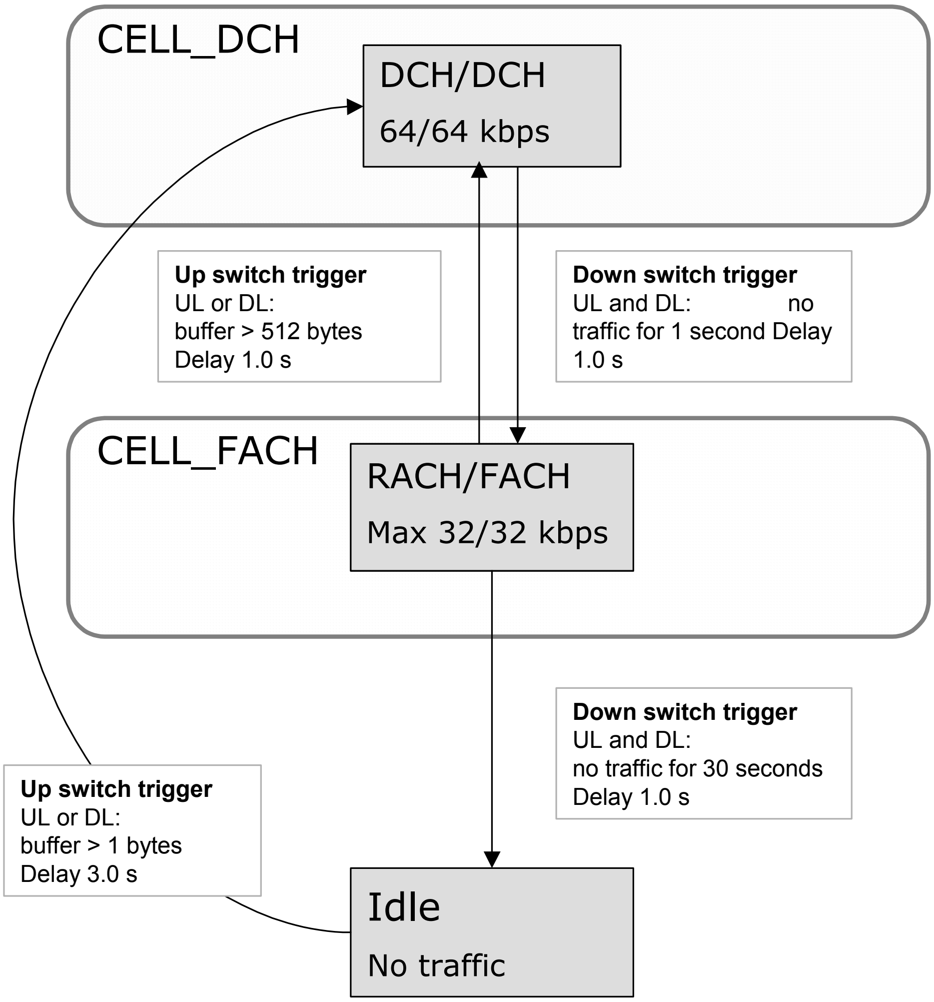

Figure 2-10. Set-up B, network capacity optimized. The RRC states, transport channels and the transitions between states in the emulator.

The UE can stay in CELL_FACH state for 30 seconds if it is not receiving or sending anything. After this time, it is switched down to idle mode to save power, as the UE must receive and decode all messages received on CELL_FACH.

3. Instant Messaging and Presence Services

IM and presence services have gained enormous popularity in the last few years. IM is almost real-time communication between two or more people. The most significant thing separating these services from other network based communication services such as Short Messaging Service (SMS) or email is the presence function. Usually, in an IM application the user has a pre-defined list of buddies. Next to each buddy is a symbol that tells the user whether the particular buddy is online or not, and whether the buddy is available for IM.

An IM application called Windows Messenger is shown in Figure 3-1. The user has three contacts in his buddy-list. One of the contacts is offline, and two are online, but only one is available for IM. An IM dialog is showed in Figure 3-2. Also, files, audio and video can be shared in this particular application.

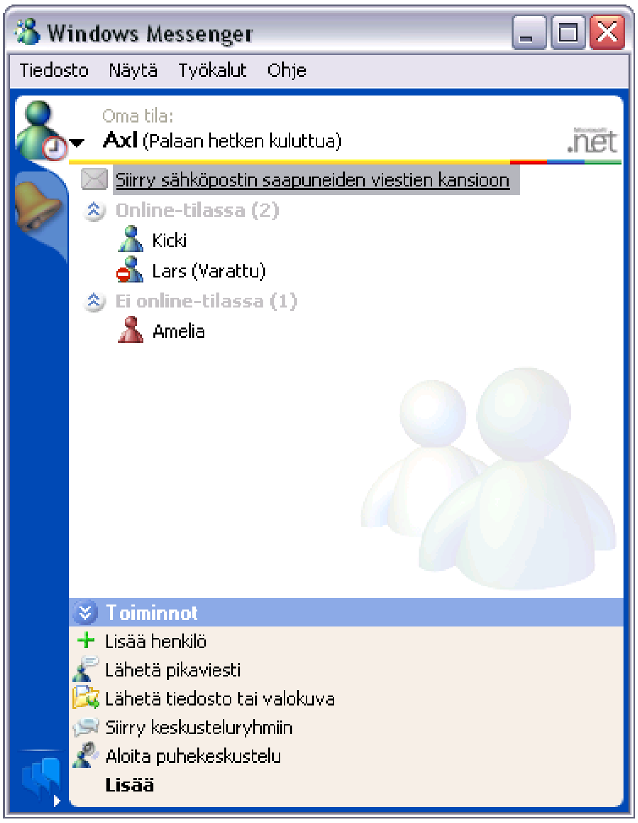

Figure 3-1. Windows Messenger application (Finnish user interface).

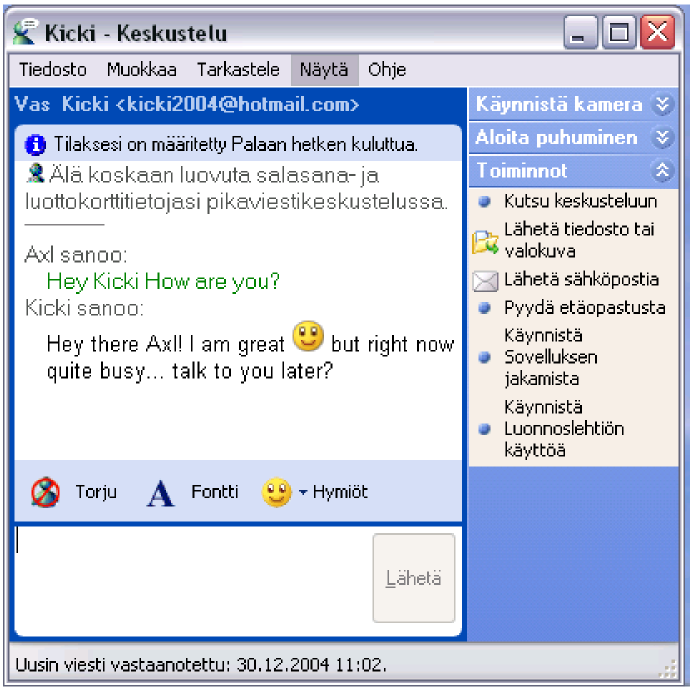

Figure 3-2. An IM session (Finnish user interface).

Many applications give the user possibilities for categorizing his contacts in to several groups, using colorful emoticons (also called ‘smileys’), like the one used in the dialog shown in Figure 3-2, and different presence options like “Away from the computer,” “Be right back,” “Out to lunch,” etc. The newest trend in presence messages is showing the users mood in addition to the normal online/offline presence information.

3.1. Background

IM was first invented in 1996, by a small Israeli company called Mirablis, when they introduced the first IM application called ICQ (“I Seek You”). In May 1997 ICQ had already 850 000 registrations, and it still was the only product on the market in it’s category. Later, in 1998 AOL bought ICQ, but it still exists as an independent application.

Today IM is still gaining popularity all around the world. The amount of users of these services is not precisely known. In 2003, however, BBC communicated some numbers announced by the companies themselves. According to BBC AOL’s products, AOL Instant Messenger and ICQ, had about 335 million users world wide, MSN Messenger about 75 million users world wide and Yahoo! Messenger about 19 million users in the US only [Rub 03].

The role of IM and presence services has also been noticed in standardizing organizations such as the Internet Engineering Task Force (IETF). The IETF is at the moment standardizing protocols and extensions also for IM and presence services. One example of such a protocol is the Session Initiation Protocol (SIP) with Instant Messaging and Presence Leveraging Extensions (SIMPLE). In the mobile world, 3GPP is working on SIP interoperability and the Open Mobile Alliance (OMA) is making it's own specification for IM and presence services.

3.2. Currently Used Applications

The IM market is currently dominated by three big companies, which offer their applications for free use. These are Microsoft with the Windows and MSN Messengers, Yahoo with it's own Yahoo messenger and American Online (AOL) with it's AIM and ICQ software.

The commercial applications use their own proprietary protocols. This means that applications from different vendors cannot communicate with each other. In practice, if a person is using the MSN Messenger he cannot communicate with a friend who is using the Yahoo! Messenger. In the enterprise world, the situation is getting better with new deals of interoperability [Hal 04], but it will still take time before the common user has access to all his/her contacts through a single application.

The next sections describe shortly what is known about the four most popular applications in the IM world.

3.2.1. AOL Instant Messenger

The AOL Instant Messenger (AIM) uses mainly a protocol called OSCAR (Open System for Communication in Real-time). This binary protocol is proprietary, although the name might suggest otherwise. In 1998, the AOL released a new string based protocol called TOC under the GNU General Public License to be used parallel with OSCAR. AOL does not use TOC itself; actually, TOC acts as a proxy for third party clients wanting to connect to AOL's AIM network.

AIM is funded with advertisements, which are shown on a small window in the application. This kind of software will be called ad-ware in the remainder of this chapter. and can be downloaded from the Internet for free. The current version is 5.5.

3.2.2. ICQ

ICQ uses the same OSCAR protocol as AIM. It used to have its own protocol, but AOL abandoned that protocol in favor of OSCAR when they bought ICQ in 1998. The ICQ version of OSCAR has only minor differences to the AIM version. Also, the ICQ version of OSCAR is proprietary.

ICQ is ad-ware and can be downloaded from the Internet for free. The current version number is 4.

3.2.3. MSN Messenger and Windows Messenger

The MSN Messenger's protocol is commonly referred to as .NET Messenger Service. August 18th in 1999, Microsoft announced that they are going to publish the MSN Messenger Service Protocol (then called MSNP) to industry. Microsoft did actually submit the protocol as an Internet draft to the IETF, but it has now expired. MSN Messenger is ad-ware and can be downloaded from the Internet for free. The current newest version is 6.2.

Windows Messenger, which is very similar to MSN Messenger, comes with newer versions of the Windows operating system. Windows Messenger can communicate with the MSN Messenger using the same .NET Messenger Service. Windows Messenger can also communicate using the IETF SIP protocol (explained in Section 3.3.1). The newest version of Windows Messenger is 5.0 and it does not show advertisements.

Although the protocol used by these messengers is proprietary, many people have made efforts to reverse-engineer it. Also, because is it text based, it is easy to conclude what each message actually delivers. For example, an instant message is delivered in plain text and can be read with a protocol analyzer tool. Presence information is sent with the help of codes, but their meaning can also be reasoned. For example, the code BRB means "Be Right Back" and NLN means "Online."

3.2.4. Yahoo! Messenger

Yahoo! Messenger uses it's own proprietary YMSG protocol. The current version is YMSG12.

Yahoo! Messenger is Ad ware and can be downloaded from the Internet. The current version is 6.0 for the Windows client and 2.5.3 for the Mac client.

3.3. Open Specifications

Because IM is very popular, there are joint efforts to produce an open standard for IM. However, so far there are no popular clients based only on an open standard.

The IETF started to produce a standard by setting up a working group called Instant Messaging and Presence Protocol (IMPP). This working group has concluded its work and produced a group of documents called Request For Comments (RFC), which now act as recommendations for more specific protocol definitions.

The IETF has, also, concluded work on two more specific standards for IM and presence. Namely these are PRIM (short for Presence and Instant Messaging Protocol) and XMPP (short for Extensible Messaging and Presence Protocol). The XMPP working group's work ended in October 2004 when they published four RFCs specifying the XMPP Protocol, also known as Jabber.

The third IETF working group working on IM and presence is SIMPLE. SIMPLE working group has not yet concluded its work. It and SIP are further discussed in Section 3.3.1.

In the mobile world, the 3GPP is working in co-operation with IETF on the SIP and SIMPLE specifications. The main reason 3GPP is involved in the IETF SIP and SIMPLE work, is the IP Multimedia Subsystem (IMS) that will be a core part of the 3G networks in the later releases of the UMTS standard [Cam 04].

There is also an existing standard for IM and presence in the mobile networks, which is produced by the Open Mobile Alliance (OMA). The OMA standard is specified to work in 2G networks through WAP and SMS as well. The OMA standard is based on work made in the former Wireless Village (WV) Initiative, which has now consolidated into OMA. In many places the OMA standard is still called WV. The WV can be used both, as a way to transport the messages and as a base for applications. For example, the company Yamigo[2] has developed an independent WV service, which is compatible with the most common IM software used and can be set up to any mobile phone that has the WV functionality.

[2] http://www.yamigo.com (31.12.2004)

In this study we will concentrate on the SIP and SIMPLE standards.

3.3.1. Session Initiation Protocol

Session Initiation Protocol (SIP) is currently an IETF proposed standard for setting up and ending sessions between two or more participants. SIP is often used for Voice over IP sessions, but it can be used for other kinds of sessions just as well. SIP is only involved in the signalling part of a session and that is why it needs to interoperate with several other protocols as well. SIP carries the Session Description Protocol (SDP), which defines the media content of a session. SDP is defined in IETF RFC 2327.

3.3.1.1. SIP Background

On February 22nd, 1996, Eve Schooler submitted the first version of the SIP, then called the Session Invitation Protocol (SIPv1), as an Internet draft to the IETF. On the same day Henning Schulzrinne submitted a draft describing the Simple Conference Invitation Protocol (SCIP) to the IETF as well. In December 1996, the IETF decided to merge the two protocols and the new protocol that resulted was the Session Initiation Protocol (SIP). SIP was initially developed in the Multiparty Multimedia Session Control (MMUSIC) working group. Since 1999, the work on SIP has been done in a separate SIP working group [Cam 02].

The 3GPP has interest in the IETF work on SIP, because of the development of the IP Multimedia Subsystem (IMS). The 3G network, as mentioned earlier, is intended to merge the Internet and the cellular worlds. The network already makes packet switched data transfer possible, but the IMS is still needed in order to maintain the QoS requirements, make charging possible and help in the integration of different services [Cam 04].

The 3GPP and IETF collaboration is documented in RFC 3113. Also, the 3GPP2 has it’s own specification for IMS but it is out of the scope of this study.

3.3.1.2. SIP Protocol Design

SIP was first specified in RFC 2543, in 1999, which was later in 2002 replaced by a clarified version RFC 3261. The RFC 3261 defines the SIP protocol in detail. The core SIP specification concentrates on methods for session initiation and ending, and does not commit to what kind of sessions it carries.

3.3.1.3. Messaging & Presence Procedures

In SIP, the procedures consist of handshakes combined from different messages. More specifically, the messaging and presence procedures belong to an extension of SIP called SIMPLE. Session initiation and ending procedures belong to SIP.

RFC 3265 describes the "Session Initiation Protocol (SIP)-Specific Event Notification," which is an extension to SIP providing a framework for event notification. This asynchronous notification of events is particularly interesting in the case of services for which cooperation between end-nodes is required, such as buddy lists (user presence events).

IM requires still one more extension to SIP, namely "Session Initiation Protocol (SIP) Extension for Instant Messaging," defined in RFC 3428. This RFC proposes the MESSAGE method, which allows the transfer of instant messages.

As mentioned earlier, 3GPP is involved in SIP and SIMPLE work as well. 3GPP has it's own system of standardizing and the naming and numbering is different from IETF. Because of IMS work, 3GPP is interested in presence services. Technical Specification (TS) 23.141 describes the architecture and function of a presence service in a 3G network. TS 22.141 specifies the first stage of the presence service. Technical Report (TR) 24.841 shows the protocol details, functional models and information flows related to a SIP based presence service.

4. Performance Analysis

In this chapter, we investigate the technical performance of Windows Messenger and a SIP-based messenger over the two WCDMA configurations explained in Chapter 2.

4.1. Measurement Set-up

The laboratory set-up for the measurements is shown in Figure 4-1.

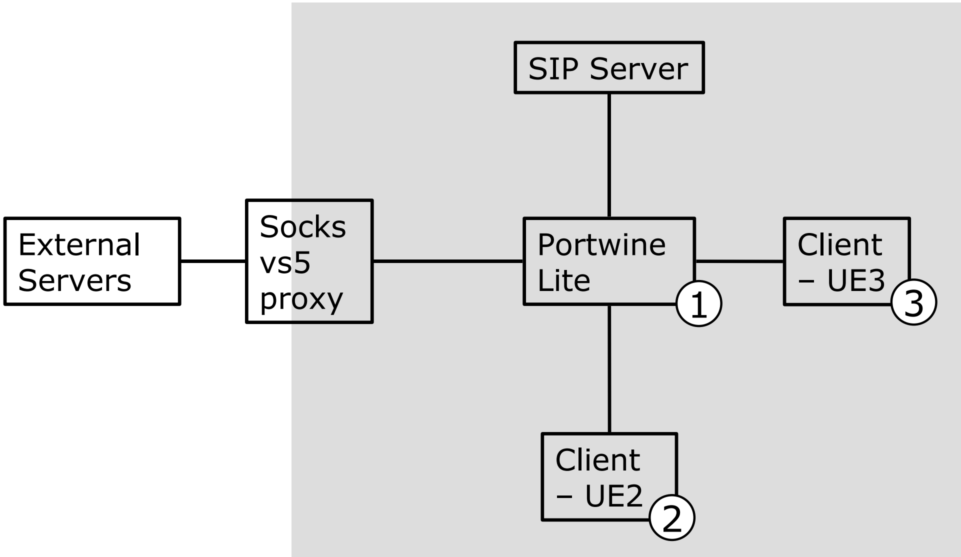

Figure 4-1. The lab set-up for the measurements. The grey area denotes the local network and the white circles indicate the three measurement points.

The laboratory set-up consisted of five laboratory computers and some outside servers. The client applications were run on Windows XP operating system, the SIP server on Windows NT, and the proxy server and the emulator – explained in Section 2.5 – on Red Hat Linux. Packet traffic was caught and analyzed with the Ethereal Network Analyzer Protocol tool[3] in all three measurement points.

In both set-ups, all IP level traffic between the clients and servers was routed through the emulator, which then added the appropriate WCDMA delay to the packets.

---

[3] http://www.ethereal.com (31.12.2004)

4.1.1. Windows Messenger

The Windows Messenger client application was in contact with several public servers in the Internet while active, which caused a problem with the emulator configuration, since all servers had to be entered by hand. To solve this problem, a Socks version 5 proxy server, defined in RFC 1928, was used to route all Windows Messenger generated traffic through the emulator. The proxy used its own IP address, instead of the remote address, when communicating with either the client application or the external server. This meant that all traffic flowing through the emulator was between the proxy servers IP address and the clients IP addresses.

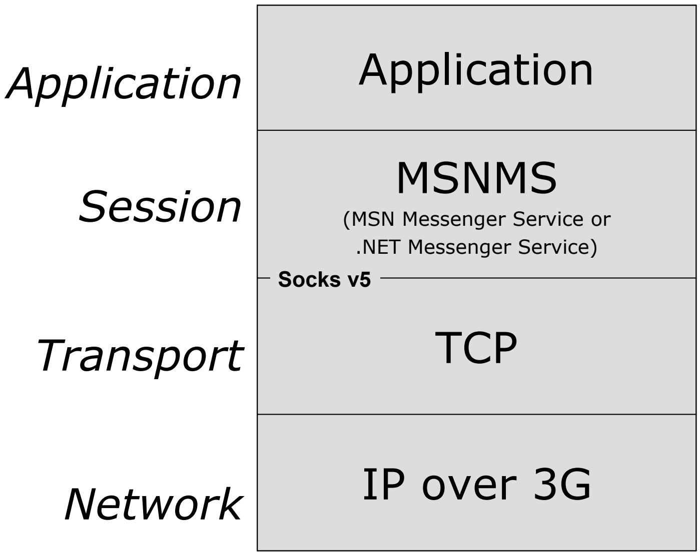

Figure 4-2. Windows Messenger protocol stack.

Windows Messenger's protocol stack, compared with the Open Systems Interconnect (OSI) levels, is depicted in Figure 4-2. As can be seen in the figure, Windows Messenger uses Transmission Control Protocol (TCP) in the transport layer and its own MSN Messenger Service (MSNMS) – also called .NET Messenger Service – protocol in the session layer. The Socks v5 protocol used in the measurement setup is located somewhere between the session and transport layers.

The use of the Socks proxy generated some extra traffic to the procedure, because the proxy connection had to be established first. Because the Socks proxy is two-way, meaning that certified services could contact the client from the outside, no periodical keep-alive functions were needed between the messenger clients and their servers. All Windows Messenger measurements were automated with the AutoIT3[4] scripting language.

4.1.2. SIP Based Messenger

The SIP set-up was entirely located at the laboratory network. The used configuration consisted of a SIP server, two basic SIP clients, and the emulator.

There are several protocols on different OSI levels with which the session layer protocol SIP could be used. Figure 4-3 shows the ones that are relevant to our case.

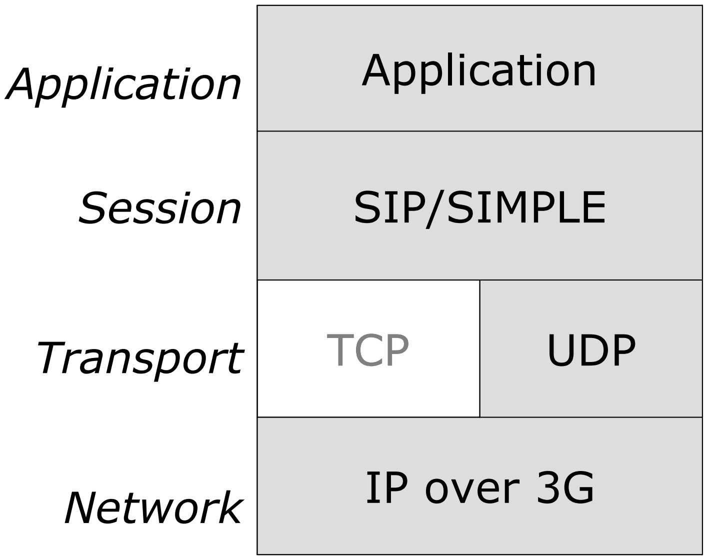

Figure 4-3. SIP protocol stack

The SIP client used the User Datagram Protocol (UDP) on the transport layer, as can be seen in the figure. The SIP protocol was used on the session level to establish, manage, and end sessions and SIMPLE protocol was used for IM and presence services. The SIP client incorporated a keep-alive function for updating the user profile to the server. The timer was configurable and was set to one hour for minimal interference with the measurement results. Also, another timer called retransmission timer (RTT) was changed from its original value 500 ms to 4000 ms for the same reason. All SIP measurements were automated with AutoIT3 scripts.

---

[4] http://www.hiddensoft.com/autoit3/ (31.12.2004)

In the capacity optimized set-up, the UE's initial state was always the idle mode. In the performance-optimized set-up, the UE was always initially in the URA_PCH state.

4.1.3. Use Cases

Two use cases were measured with both client applications, over both WCDMA set-ups. The initial state for the UE was idle mode, when the capacity optimized set-up was used, and URA_PCH state when the performance-optimized set-up was used.

The use cases were:

1. **Login**: User A logs in to the service. The use case was measured from the first packet the client PC sent, to the last packet received/sent before the client application showed to the user it was online. User sees he is online when the presence state is online, and the application window is opened. This measurement interval does not include client processing before the first measured packet or after the last measured packet. E.g., the time from when the user presses the "Login" button to when the first packet leaves the PC is not included. The measured interval does not include buddy list presence updating.

2. **IM session**: two users interact via instant messages. First, User A opens the conversation with one message, and then User B replies to that message. The use case was measured from the first packet sent after User A had opened the conversation window, to the packet in which User B gets the message content. The second interval was measured from the packet in which User B sends the response message content, to when User A gets the message content. The time required to read the message and write a response was approximated with 15 seconds. As in the previous use case, client application processing between the user action and the actual first or last packet is not included.

4.2. *Windows Messenger Measurements*

4.2.1. Use Case 1: Login

The first use case was the login procedure. The signalling in the use case is shown in Figure 4-4. The measured interval is shown with a purple curly bracket. All messages in the figure are part of the MSN Messenger Service protocol, unless they are marked with Socks or TCP.

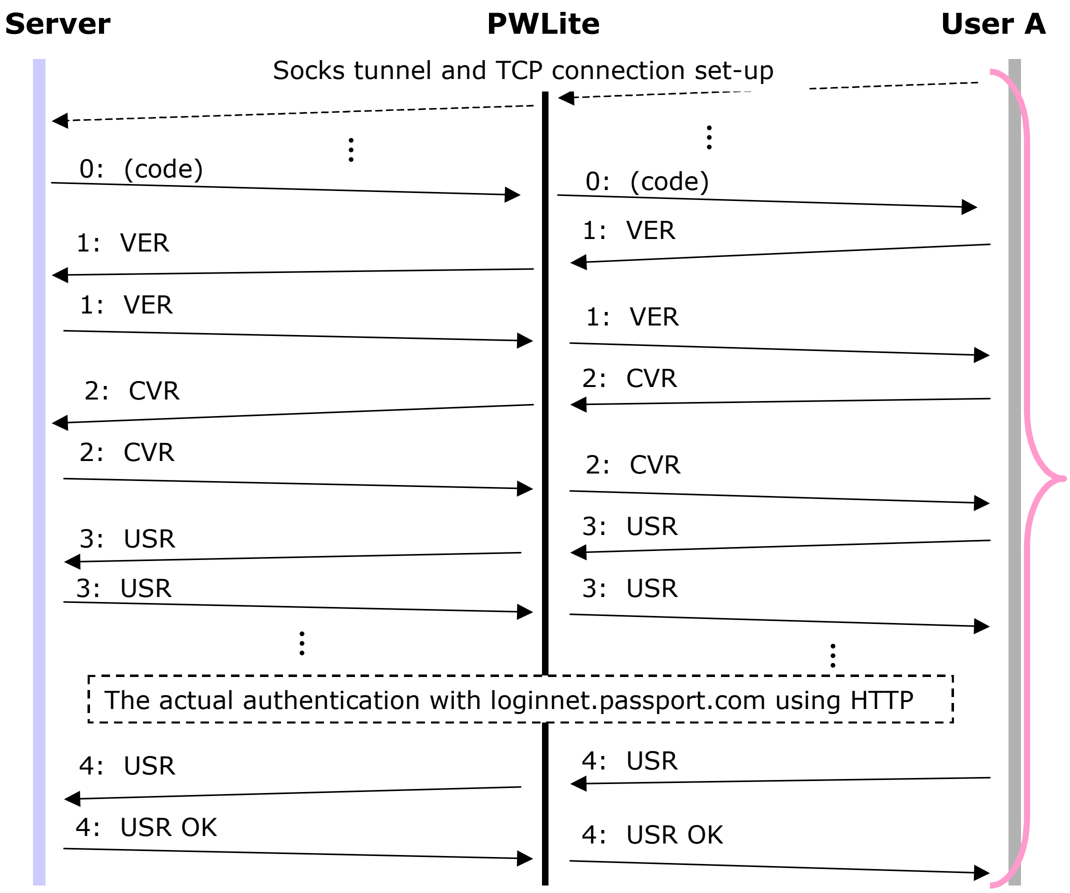

Figure 4-4. Signalling in Windows Messenger during login.

The IP packet sizes and a more detailed description of the messages, sent in the Windows Messenger login procedure, are shown in Table 4-1. Note, however that Figure 4-4 and Table 4-1 do not show all messages sent during the depicted procedure. First the client negotiates a TCP connection with the Socks version 5 - proxy server. Then it opens a second TCP connection to use for the remote server. Between MSNMS USR messages 3 and 4, the client makes the actual authentication with the MS Passport at `loginnet.passport.com` through HTTP. All together, about 45 messages were sent each time between the socks proxy and the client as part of the measured login sequence.

Table 4-1. Details of messages sent during Windows Messenger login.

Direction	Message	Size (bytes)	Description
< (UL)	TCP [SYN]	60	TCP connection opening
> (DL)	0: MSNMS (code)	60	A code sequence sent to start the MSNMS connection
<	1: MSNMS [VER]	70	Specifies supported versions of the MSNMS protocol
>	1: MSNMS [VER]	70	Server denotes that version is ok
<	2: MSNMS [CVR]	140	Sends version information about the client and OS to the server
>	2: MSNMS [CVR]	200	Server sends the download address of newest version
<	3: MSNMS [USR]	90	Initial authentication. Username is sent with this message.
>	3: MSNMS [USR]	210	Subsequent USR response. Contains authentication string.
At this point the client does the actual authentication with the authentication server			
<	4: MSNMS [USR]	370	Final USR command, sent after authentication.
>	4: MSNMS [USR]	100	Server responds with USR OK.
Now the client is successfully logged in. The remainder of the messages are for intial presence.			

The application indicated to the user that it was online after MSNMS message “4 USR OK” arrived to the client by then opening the application window and showing the not yet updated contact list. After the client had logged in, it updated the presence information by synchronizing itself first with the presence server and then comparing its contact list to the list at the server. The initial presence updating was not included in these measurements.

The message names and sizes in Table 4-1 were taken from the Ethereal log acquired from the measurement and the descriptions were concluded with the help of the MSN Messenger reverse-engineered protocol explanation[5].

The results are presented as Cumulative Distribution Functions (CDF). A CDF is defined as follows: CDF tells the probability distribution for a random variable X. The CDF for every real number x is given by (Equation 4-1), where the right-hand side represents the probability that the variable X takes on a value less than or equal to x. The CDF function is always denoted with a capital F.

(Equation 4-1) $$F(x) = P(X \leq x)$$

CDF is used here, because it shows delay distribution of the observations. It also shows what kind of delays one should expect with these use cases and these applications.

[5] http://www.hypothetic.org/docs/msn/index.php (31.12.2004)

The graph for use case 1, for both WCDMA set-ups, is depicted in Figure 4-5.

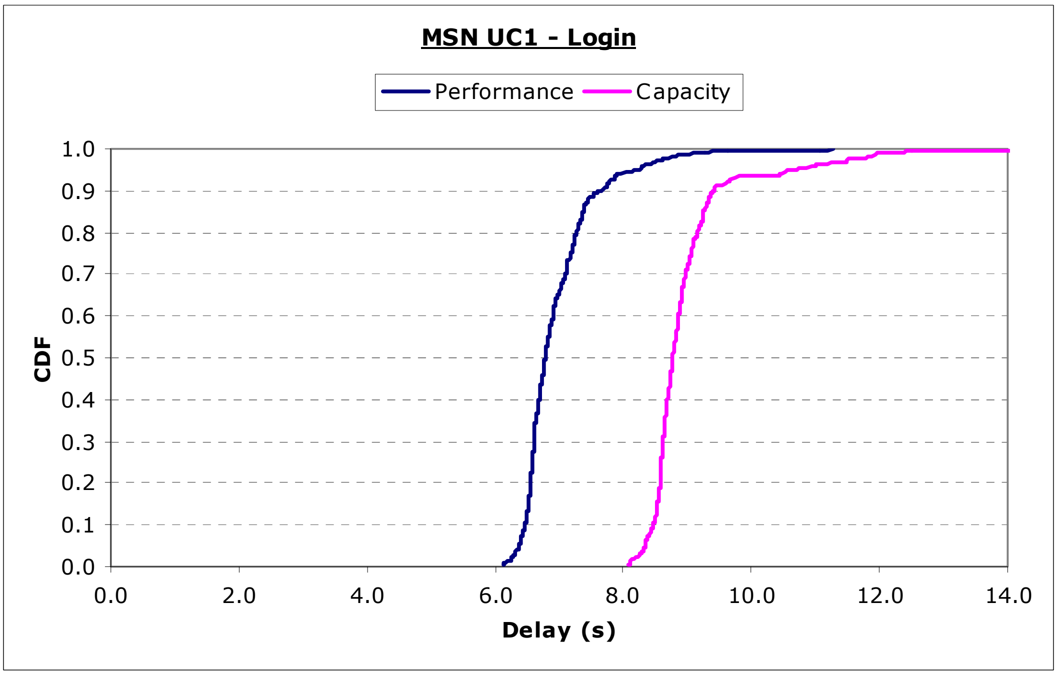

Figure 4-5. Results for Windows Messenger use case 1.

In the performance optimized set-up, the up switch from URA_PCH to the CELL_DCH state takes around 1 second. The up switch from idle state to CELL_DCH takes about 3 seconds in the capacity optimized set-up. The two-second difference between the channel switching delays can be seen clearly in the graph; the curve for the capacity optimized set-ups results starts 2 seconds later than the curve for the performance optimized set-up. The part affecting the user perceived total delay the most was the authentication, which was also included in the measured interval, as could be seen from Figure 4-4. The time needed for the authentication varied from a little over 2 seconds to 11 seconds or longer. A more careful look at the delays revealed that server processing often caused the extra delay in authentication. However, as can be seen from the graphs tails, the users experience a longer delay in only about 10% of the cases. The processing was slow either when opening the HTTP connection for authentication, or between the USR 4 query and response after authentication. During login a total amount of about 7200-7600 Bytes was sent over the radio interface. If this were to be sent over the radio interface in one packet, it would take about 0.1 seconds to transmit the data without retransmissions, assuming a 64 kbps channel. Sending this amount of data takes in this setting so long, because it is chopped up to small packets sent over the radio interface on at a time. If sending one packet takes about 100 ms, sending 45 packets will take about 4500 ms. To this value we still need to add the retransmissions and the channel switches. All in all, this adds up to about 6 seconds. The biggest load was caused by the authentication, during which packets of sizes up to 1500 Bytes were sent across the network.

4.2.2. Use Case 2: Instant Messaging

The second use case was instant messaging. In this use case two users, A and B, sent instant messages to each other. In practice, the use case was simplified so that first User A sent an instant message and then User B replied to it.

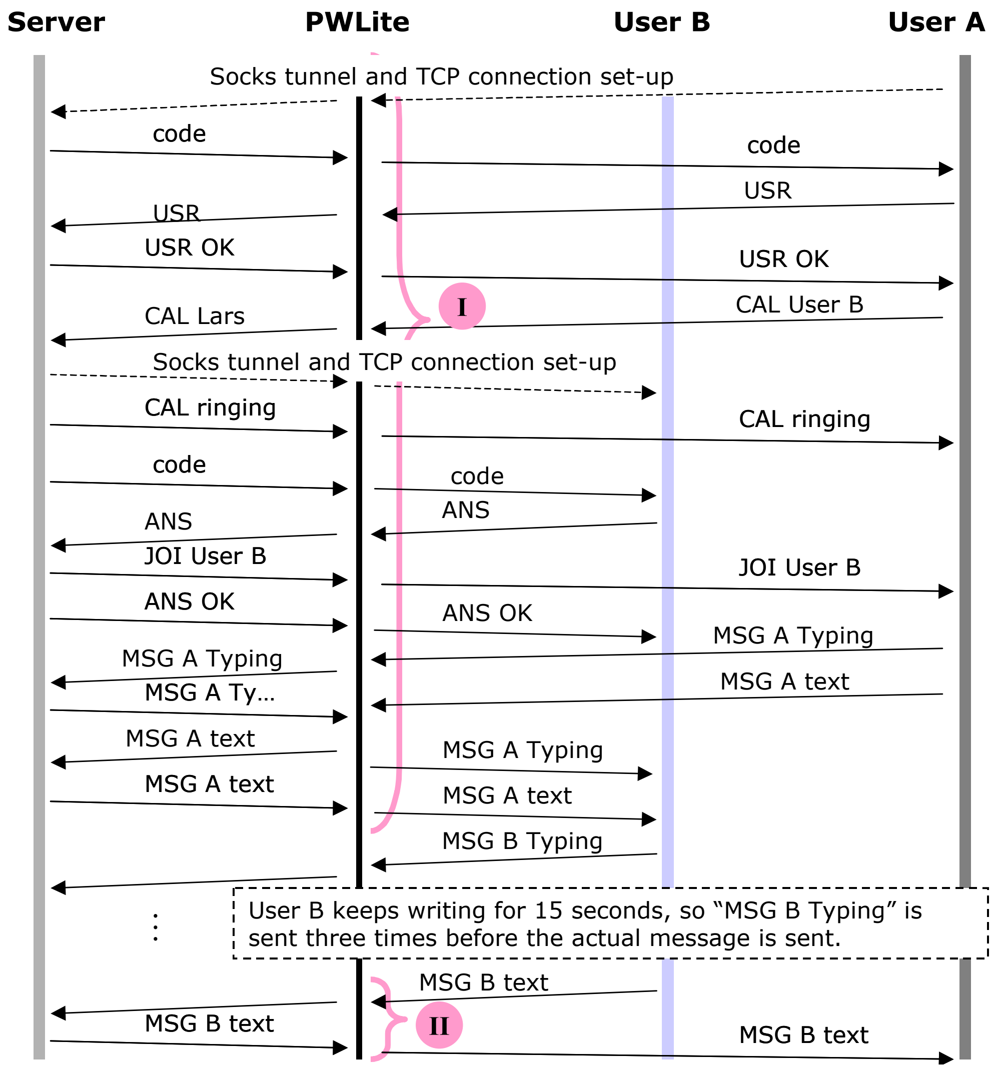

Figure 4-6. Signalling in Windows Messenger in an IM session.

The signalling in this use case is shown in Figure 4-6, and the packet sizes and descriptions in Table 4-2. One should note that the TCP connection opening and the Socks connection opening are not shown in detail, although, they were always made. All together, 45 messages were sent through the radio interface during the measured interval I, depicted in Figure 4-6 with a curly bracket.

During measured interval II only a few TCP [ACK] messages were sent besides the messages shown in Figure 4-6.

Table 4-2. Packets and their descriptions in Windows Messenger use case 2.

Direction	UE	Message	Size (bytes)	Description
< (UL)	3	TCP [SYN]	60	User A opens a TCP connection to server from UE3
> (DL)	3	MSNMS (code)	60	A code sequence sent to start the MSNMS connection
<	3	MSNMS [USR]	110	User A sends his user ID
>	3	MSNMS [USR OK]	100	Server responds if ID ok
<	3	MSNMS [CAL User B]	90	User A sends a call request for the user User B
>	2	Socks (Unknown)	130	Server contacs User B in UE2 through Socks
>	3	MSNMS [CAL Ringing]	80	Server denotes that the call to User B is ringing
<	2	TCP [SYN]	60	User B opens TCP connection to server
>	2	MSNMS (code)	60	A code sequence sent to start the MSNMS connection
<	2	MSNMS [ANS ID]	120	User B answers with his user ID
>	2	MSNMS [ANS OK]	110	Server replies if ID ok
>	3	MSNMS [JOI User B]	90	Server tells User A that User B has joined the conversation
At this point the IM session has been opened.				
<	3	MSNMS [MSG A Typing]	170	User A starts to type a message
>	2	MSNMS [MSG A Typing]	190	Server forwars the info to User B
<	3	MSNMS [MSG A txt]	230	The actual message is sent
>	2	MSNMS [MSG A txt]	250	Server forwards the message to User B.
<	2	MSNMS [MSG B Typing]	170	User B starts to type a message
>	3	MSNMS [MSG B Typing]	190	Server forwards the info to User A
Because it takes User B 15 seconds to answer to User A, the "MSG B Typing" message is sent several times.				
<	2	MSNMS [MSG B text]	310	The actual message is sent
>	3	MSNMS [MSG B text]	340	Server forwards the message to User A

Results for both WCDMA set-ups can be seen in Figure 4-7. The marking in the picture is as follow: the abbreviation MSG means a message and the letters A and B denote if the message was the first message (A) or the response to it (B). The word “Perf” refers to the performance optimized set-up and the word “Capa” to the capacity optimized set-up.

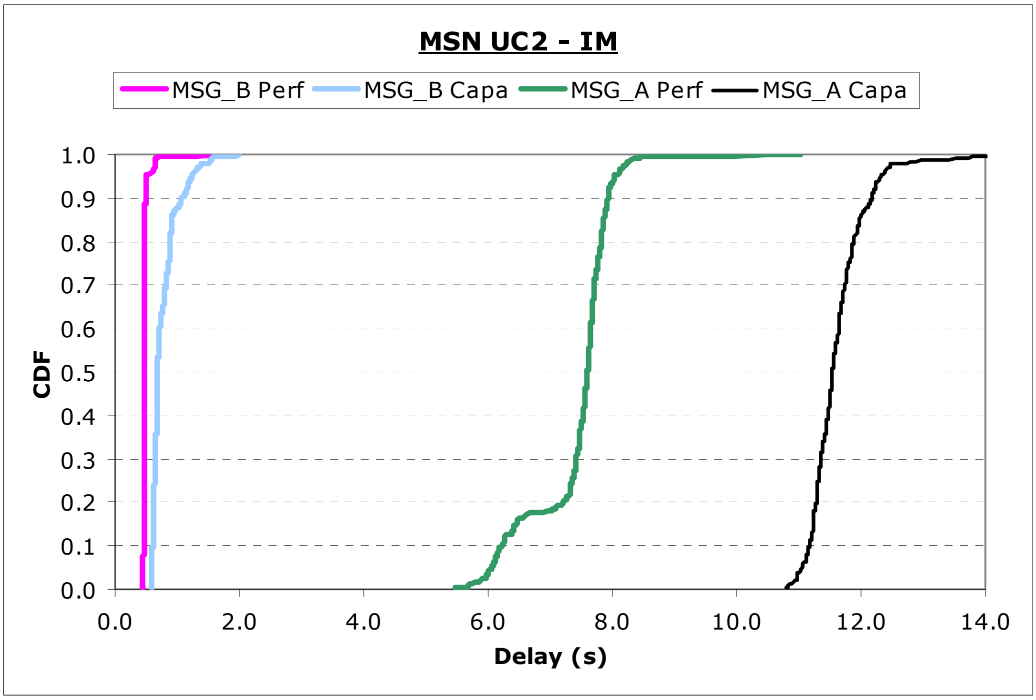

Figure 4-7. Results for MSN use case 2.

As can be seen from Figure 4-7, opening the conversation takes fairly long, whereas sending a message after the session is established is quick. The difference between the two set-ups can be clearly seen from the delays in the first messages. The difference between the channel switch delays in these two set-ups was two seconds, and because, in this use case, traffic is going to two different UEs, we have two radio interfaces to cross. This means that the difference between the set-up delays is four seconds.

The second message was sent after the IM session had already been set-up. The sending of the second message took less than a second in both set-ups. Because the writing delay was approximated with 15 seconds, the UE stayed all the time in the CELL_DCH state in the performance-optimized set-up. In the capacity optimized set-up, the UE was switched down to CELL_FACH state after one second of silence. The packets, in which the message was sent, were, however, so small that they did not cause an up switch from CELL_FACH. Instead the packets were sent on the common channel. Because the CELL_FACH is slower than CELL_DCH the “MSG_B Capa” had longer delays than “MSG_B Perf.”

4.3. SIP Based Client

4.3.1. Use Case 1: Login

The first use case was the login procedure. In this implementation, the authentication in the login procedure was done using the HTTP Authentication defined in RFC 2617. The signalling in the use case is shown in Figure 4-8, and the measured sequence is marked with a curly bracket. After login, the client always sent subscribe-messages to the server to get the presence information of its contacts and to update its own initial presence. This, however, was not included in these measurements.

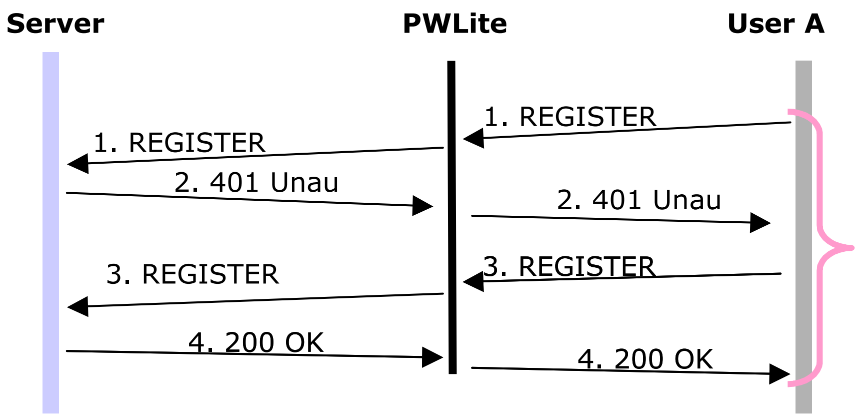

Figure 4-8. The measured signalling and IP packet sizes used in SIP login.

The IP packet sizes and a more detailed description of the messages are shown in Table 4-3.

Table 4-3. Description of messages in SIP use case 1.

Direction	Message	Size (bytes)	Description
< (UL)	[SIP] REGISTER	520	Client send a register request to the server
> (DL)	[SIP] 401 Unauthorized	540	Server responds with 401 Unauthorized containing a nonce
<	[SIP] REGISTER	780	Client replies with a register message again, this time containing the nonce.
>	[SIP] 200 OK	630	Server replies with a 200 OK
Now the client is successfully logged in. The remainder of the messages are for intial presence and updating the contact list.			

After the measurements it was noticed that a user modifiable parameter RTT, introduced earlier in Section 4.1.2, had significant effect on the login procedure. This parameter was apparently used to control the retransmissions of SIP messages. Somehow, however, it also affected the login procedure. In the login procedure, a message named REGISTER was first sent from the UE to server and later sent a gain as a response to the server's "401 Unauthorized" message. RFC 2617 defines that the same query must be sent again as a response to the "401 Unauthorized" message. Although, the two REGISTER messages were different sized and had different content, the time between them was unfortunately controlled by this parameter. By default, the timer was set to 500 ms, but this value was changed to 4000 ms with the intent to avoid unnecessary retransmission causing extra load during channel switches.

The results for SIP use case 1, for both WCDMA set-ups, are shown in Figure 4-9.

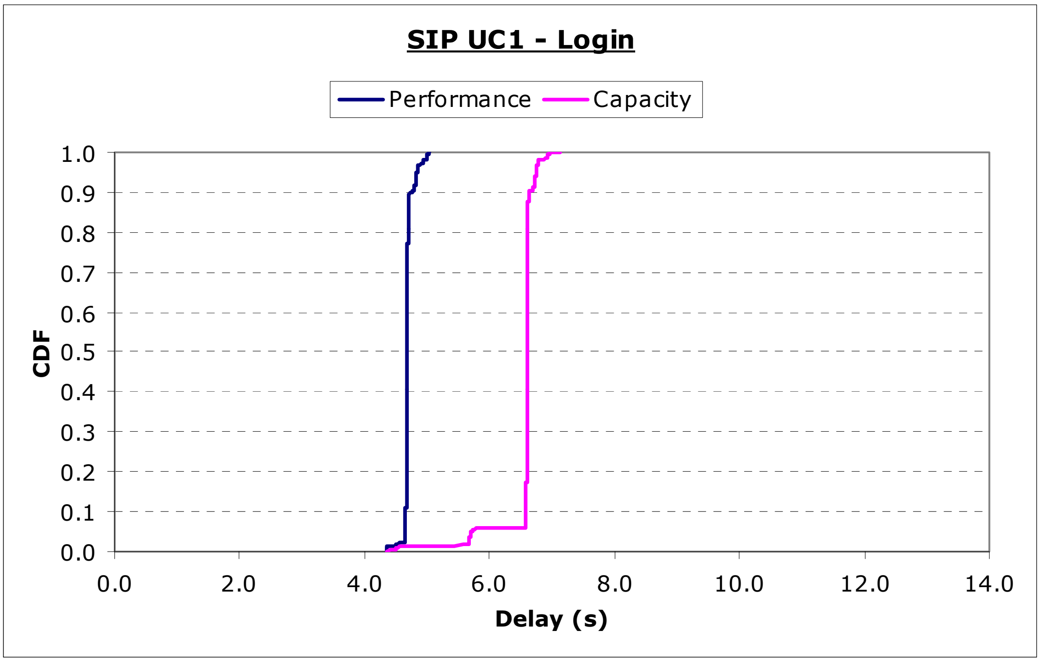

Figure 4-9. Results for SIP use case 1.

The effect of the retransmission timer discussed earlier in this section, can be clearly seen from the resulting CDF curves. The minimum delays for both set-ups are around four seconds, which is not surprising remembering that the timer was set to 4000 ms. The transmission delays of couple of hundreds milliseconds, which comes from sending the messages in the second handshake (3. REGISTER and 4. 200 OK), are added to the four seconds. The performance, however, still varies a great deal between the two set-ups. In the capacity optimized set-up, the up switch from the idle state up to the CELL_DCH state lasts 3.0 seconds. After that the "401 Unauthorized" response is sent right away. This leaves still one second from the 4.0 seconds the retransmission timer was set to, which is silent time in the network. If the silent time lasts more than 1000 ms in the network, a down switch to CELL_FACH state is started. The down switch then lasts 1.0 second, during which the second “REGISTER” message comes to the buffer. As could be seen in Table 4-3, the second “REGISTER” message is more than 512 bytes in size, and so exceeds the limit causing an immediate up switch back to the CELL_DCH state. This channel switching results in at least 2.0 second extra delay to the login procedure, and happens in most of the cases measured in the capacity optimized set-up, as can be seen in the graph. If the transmission timer had not been set to 4000 ms, the delay graphs would look very different. In the capacity optimized set-up, the login would last about 3 seconds and, in the performance optimized set-up, it would last about 1 second.

The curve obtained from the results for the performance optimized set-up looks simple. The small steps on the top of the curve are caused by the RLC layer retransmissions.

4.3.2. Use Case 2: Instant Messaging

The second use case was IM. In this scenario, two clients interacted by sending instant messages: one client first sent a message to which the other then replied. A period of 15 seconds was used to approximate the time needed by the responding user to read the message and write a response to it. The SIP signalling used in the client is shown in Figure 4-10.

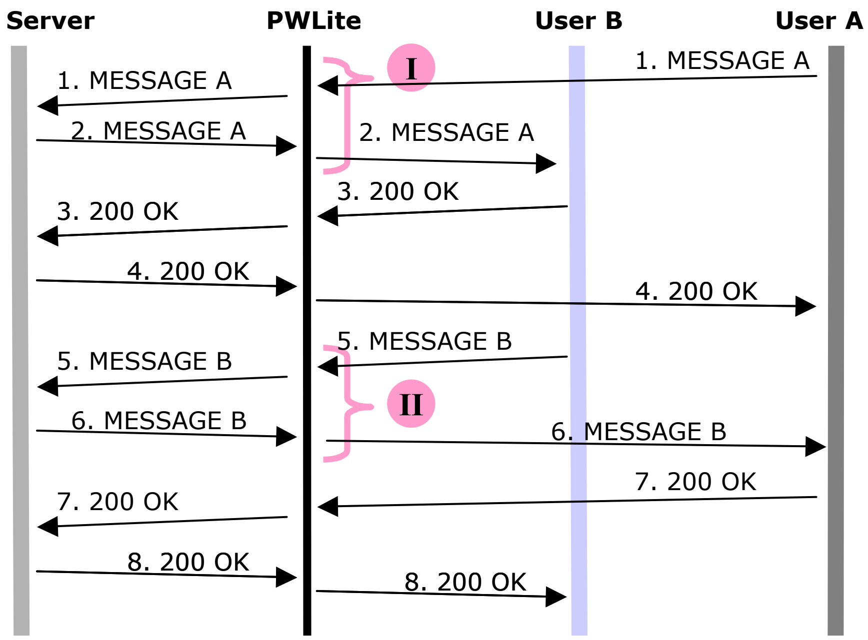

Figure 4-10. Signalling in a SIP-based IM session.

The SIP IM works much more straightforward in this implementation than IM in Windows Messenger. Once the SIP session has been created, the users can interact without having to create a separate session. One must keep in mind, however, that this client application is not in commercial use and therefore it cannot be verified that it would function in this way in real use as well.

The message sizes and their descriptions are shown in Table 4-4. As can be seen from the table, the original sent message has grown considerably in size when it leaves the server. This is because the server for some reason attaches information to it, e.g., about the route the message travels.

Table 4-4. Description of messages in SIP use case 2.

Direction	UE	Message	Size (bytes)	Description
< (UL)	3	Message A	750	User A sends the message to the server.
> (DL)	2	Message A	1210	Server forwards the message to User B and adds to it some information about the route.
<	2	200 OK	570	User B replies that it got the message.
>	3	200 OK	520	Server forwards the reply to the sender, User A.
<	2	Message B	840	User B sends the reply message.
>	3	Message B	1300	Server forwards the message to User A.
<	3	200 OK	560	User A replies that it got the message.
>	2	200 OK	520	Server forwards the reply to the sender, User B.

The results for SIP use case 2, for both WCDMA set-ups, are depicted in Figure 4-11. In this application, the difference between the two set-ups can be clearly seen from the CDF curves in the figure.

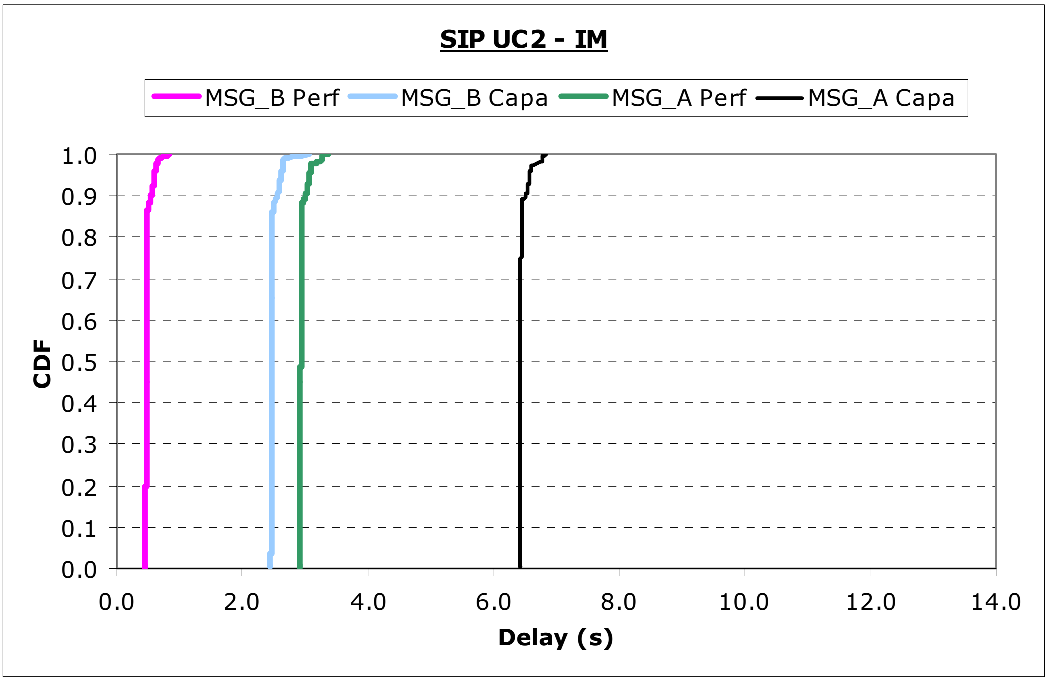

Figure 4-11. Results for SIP use case 2.

The initial state for both UEs in the capacity optimized set-up was the idle state and URA_PCH in the performance optimized set-up. This can be seen from the first messages: in the capacity optimized set-up it takes the first message over six seconds to get through, while in the performance optimized set-up it takes only around three seconds. The three seconds can be explained mostly with the up-switch times from the URA_PCH, which were 1.0 second in uplink and 1.5 seconds in downlink direction. When the first message was sent, there was one up switch first in uplink and then one in downlink, which adds up to total of at least 2.5 seconds delay, to which we then need to add the transmission delays. The small steps on top of each curve were caused by RLC retransmissions.

The delay of the second message also has a big difference depending on the used WCDMA set-up. In the performance-optimized set-up, the UEs stay all the time on the CELL_DCH, because the writing time is only 15 seconds. This explains the short delays experienced by the second message in the performance set-up. In the capacity optimized set-up, the UEs are both switched down to the CELL_FACH state before the second message is sent. Because the message is greater than 512 bytes, both UEs need to be switched back up to the CELL_DCH state, so that the message can be sent.

4.4. Summary

In this chapter, we have studied the technical performance of Windows Messenger used behind a Socks v5 proxy server and a SIP-based messenger solution run on two clients with emulated WCDMA links, with different characteristics.

The Windows Messenger behaved similarly over both WCDMA solutions. Although it took longer to log in and start an IM session in the capacity optimized set-up, the actual messaging was almost as fast as with the performance-optimized WCDMA, because the IP packets were so small that they required no up switch from the common channel. The difference in the transmission times for small packets stayed at couple of hundreds milliseconds, which is unnoticeable to the human user in this kind of application.

For comparison, also a SIP-based client was tested with the same use cases and WCDMA set-ups. The used simple SIP-based client produced far less traffic when performing the use cases, and especially, it performed better in the login procedure and session handling. On the other hand, in an IM session the SIP client performed worse. The packets the SIP client sent were bigger in size than the packets sent by the Windows Messenger. This meant that the SIP packets always needed an up switch from the CELL_FACH state, while the Windows Messenger packets were small enough to be sent in the CELL_FACH. It is good to note that this SIP solution is not in commercial use, so it is hard to estimate, if it were to work the same way with millions of users.

In conclusion, it seems that the delay caused by the WCDMA network is, first of all, in all cases dependable greatly on the networks switching choices, but also, on the number of handshakes on application level. Although the WCDMA delays sometimes might cause an extra delay of even six seconds to a procedure, the application itself might also cause random length delays that are even longer with, e.g., complicated authentication systems. It is important to remember that the user perceived delay might be even longer than the measured delays here, because here we did not include that application processing to the measurements.

5. Usability Evaluations

In this chapter, we first define the requirements and explain the motive for the usability tests. Guidelines for the evaluation are from [Fau 00] and [Nie 93]. The remainder of the chapter explains the organization of the tests and reports the obtained results.

5.1. Usability Requirement Specification

5.1.1. User Group

The users of IM and presence services vary a great deal in age, educational background, gender and current profession. However, the most significant user group of these services are teenagers (aged 13-19 years) and young adults (over 19 years).

Young adults are mostly students in higher educational institutes such as universities or polytechnics or are already in working life. Most of the young adults have their own family or at least their own home. In other words, they take care of their own finances and economy. We decided not to place a strict upper age limit to this group, but we expect these people to be around their thirties at the most.

Teenagers can be seen as young people still in middle school or high school. Most likely they live with their parents, although there are always exceptions. Teenage users form a very interesting group, because they show interest in new technology and ways of communication, and often adopt new communication styles easier than other people. Teenagers' use of IM and communication in general has been studied previously by, e.g., Grinter and Palen in [Gri 02] and by Grinter and Eldridge in [Gri 01]. However, teenagers are financially dependable on their parents. Thus, they probably are not in the first wave of 3G service users.

The user group that was chosen to participate in this usability evaluation was young adults aged around 19-30 years. The group was then narrowed a bit more; only people who had already used IM were chosen. It was thought to be important that the test users were familiar with IM, since the main point was to find out whether using IM over WCDMA differs a great deal from using it over the Internet. All users with some experience on IM were qualified, so there were both novice and expert users in the group. People in this group are easily reachable for this kind of evaluation, and are old enough to decide themselves whether to participate or not.

We found some statistic about the young adults. For example, from all American IM users 31% are 18-27 years old, and 59% are under 40 years old [Shi 04]. In Sweden, 39% of 18-24 year old population use IM or chat. From people aged 25-44 years, 15% in the low-income class and 12% in the high-income class use chat or IM [Nil 04]. Young people are also the major group using mobile text messaging, SMS. In Finland, almost all young people using mobile phones also use SMS. On average about 40% of all people aged 10-30 years send SMS weekly [Sne 04]. Although SMS is not interactive by nature, as IM is, one could conclude that people who most likely will adopt mobile IM are the same people who are used to messaging through the Internet and through a mobile device.

5.1.2. Task Analysis

By using IM, the users seem to content two basic needs: acquire or share knowledge and on the other hand just socialize. Basically, IM is just one communication channel among others, but it is preferred in some situations as the primary channel. Among young people, the motive to use IM in free time seems to be the price. IM is free in the sense that one pays for the Internet connection, but not directly for the particular service. In the workplace, IM is probably the least disturbing way to contact someone. One can see the other person's presence info before contacting, so one knows if the person is there and available. On the other hand, an instant message is usually answered fairly quickly compared to an email, which might be answered the next day or later. In our survey, one person actually defined that IM was used at work for "quick questions."

One respondent stated that the plus sides of IM were that it is cheap, easy and works well. Another one said the key thing is that it is faster than email and SMS. Some users found the fact that the popular IM systems are not interoperable a disadvantage. A few users told that they, especially, use IM to keep in touch with friends and family abroad. Several users stated that the best thing about IM is that you can be with your friends, although you are not physically present. One respondent put it all in one sentence: "Best thing is that it's cheap and quick. Worst that nobody is ever online…"

Usually, IM is not used as a standalone activity. Shiu and Lenhart found out in [Shi 04] that IM-ers are multi-taskers; they surf on the Internet, watch TV, talk on the phone, or work offline on their computer, almost every time they are involved in an IM session. From this, one could conclude that IM is rarely a standalone activity, and so the IM session never gets the user's full attention.

We are concentrating on IM use over a mobile network. Let's picture a scenario where the user is in a train. Some people are sleeping, as is the person sitting next to our user. Our user is reading a magazine as a pastime activity and is also online in an IM application through a mobile device in case some friend would come online. The scenario is presented in Figure 5-1.

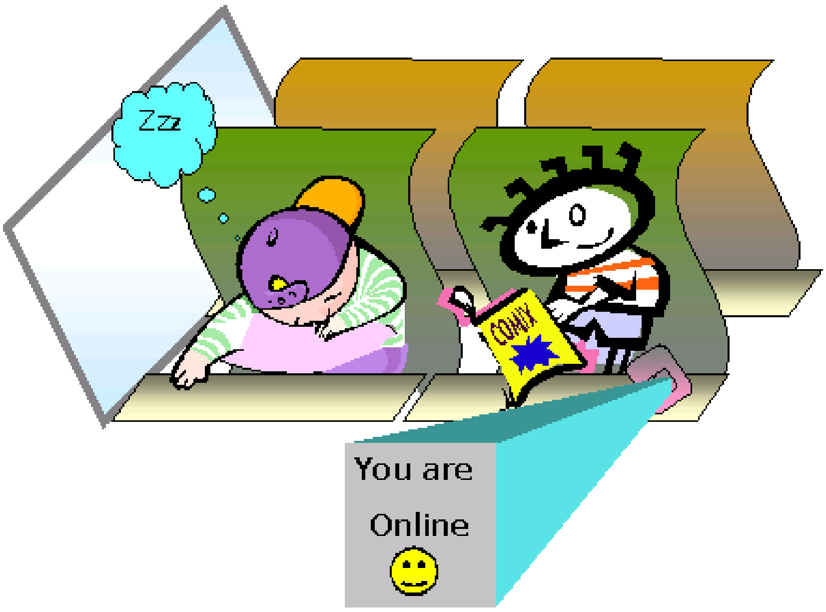

Figure 5-1. A use case scenario for mobile instant messaging.

The most important current tasks are to see if a buddy is available and if so, start an IM session with the buddy. A flow chart of the process to open an IM session is shown in Figure 5-2. The motivation for the whole process is the need to communicate. This could origin from several sources: maybe the subject needs to ask something from a buddy, or maybe the subject is just bored and wants to socialize. In any case, after the need to communicate has been identified, the subject must choose a communication channel through which he will contact the buddy. This could be, e.g., telephone, email, or IM. If the subject chooses IM he must then log in to the service. Once the subject is logged in he can see the contact list and which of the contacts are available. If the particular buddy, who he wants to talk to, is available, an IM session with this buddy is opened.

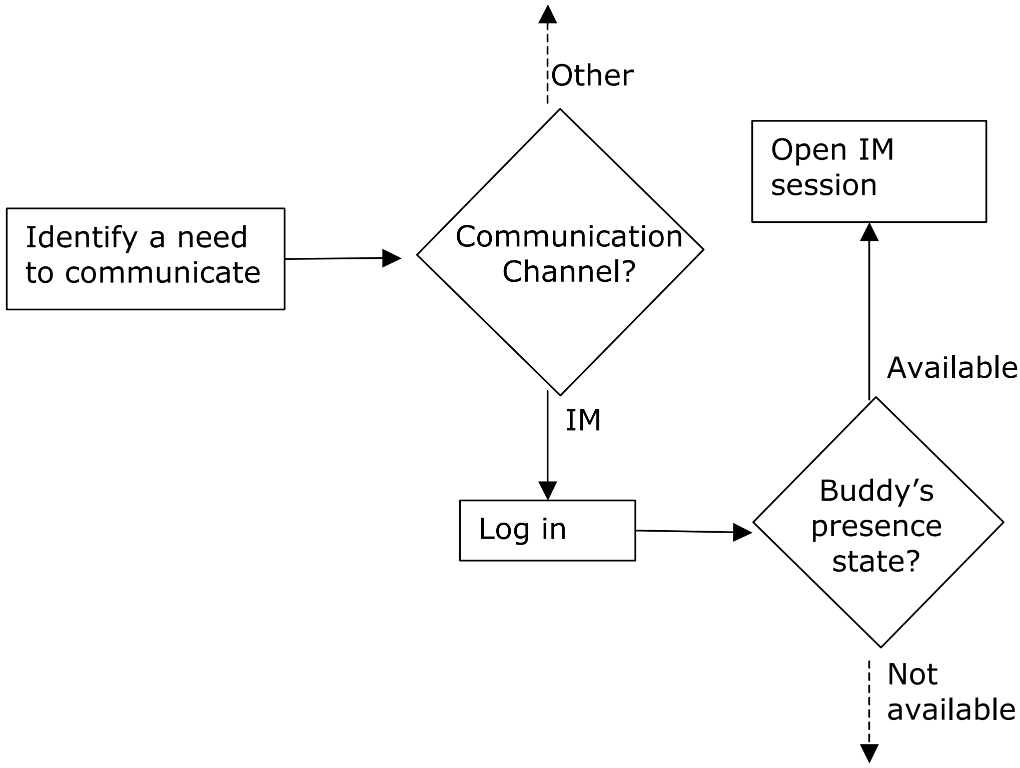

Figure 5-2. A flow chart of the IM session opening process and decisions.

5.1.3. Use Environment

The physical environment for mobile IM use varies greatly. The conventional way to use IM is to sit at a computer either at home, classroom or workplace. Mobile IM can be used anywhere, but probably it will be used when attention can be focused on the mobile device, e.g., in a train, bus, plane or similar spot. The mobile device itself can be a phone, laptop, PDA or some other device capable to run an IM application in some network.

The social environment could be anything but most likely users log into IM, and have their presence set to available, when they have time and are not involved in a face-to-face conversation. The fact that IM-ers are multi-taskers is important here; IM is often used with distractions and everyone seems to be aware of that.

5.1.4. Requirements

The user requirement on the IM over WCDMA service is that the communication should be almost, or preferably as fluent as over the fixed Internet. The presence information needs to be updated as quickly as possible, to avoid communication attempts made after the counterpart has already said he is unavailable. Using IM over WCDMA should not be noticeably different from using IM over the fixed Internet.

Requirements for the service:

1. The user cannot tell from the delays, is he using IM over 3G or fixed Internet.

2. Use cases should be carried out the same way in each set-up. Underlying network should not have impact on use cases.

We also define requirements for the usability test tasks. With these requirements it is easier to judge the test validity.

Requirements for the test tasks:

1. Conversation should be as natural as possible. Users should feel they are using IM as they normally are.

2. Tasks accomplished over different set-ups should be as similar as possible, while still being interesting enough. This requirement is important, because this ensures that the results can be compared afterwards.

5.2. Usability Evaluations

5.2.1. Test Users

The test users were chosen from a larger group of people who were asked to fill out a preliminary survey, which can be found in Appendix A. The complete result sheet acquired from this survey is in Appendix D. The main characteristics looked for were: age over 19 years and that the user had used IM before. The goal was to get both novice and expert users of IM to the test group. All together 12 people replied to the survey out of whom, four user-pairs were chosen to participate in the evaluation. The user pairs were chosen so that they knew each other in advance, because this was thought as an important factor in making the IM conversations as natural as possible.

The results show that 50% of the respondents used IM daily and 92% of the users at least a couple of times per month. One respondent had never used IM. This person's answers are not included in the percentage calculations in the remainder of this section. The majority of the users (82%) used IM because it was an easy way to communicate. 73% stated that they used IM because it was a fast way to communicate, and 64% said that one reason to use IM was that it was a cheap way to communicate. Only 18% said they used IM because all their friends use it. When teen IM usage has been studied, peer pressure has been identified as a major factor in making IM so popular [Gri 02]. It seems, however, that with young adults peer pressure is no more a major issue motivating the use.

73% of the users reported they used IM to chat with friends, while 36% reported they used IM to solve problems at work. Only one person did both activities. It seems that the IM users in this user group can roughly be divided into free-time users and work users. This means that there is, also, more than one category of tasks to accomplish with IM.

Interestingly enough, the use environment for IM varied more than the purpose: 82% of the users reported they used IM at home and 45% said they used it at work. 18% of the users stated that they used IM almost anywhere: train, school, and so on. 36% of the users said they only used IM when they had spare time, whereas 82% said they used IM when they had something to ask or tell to a person they knew was using IM. IM is often used several hours at a time. Our user group backs this up: 55% of the users said they were online while working at the computer and 18% stated they were always online while the computer was running. The IM applications commonly have sound notifications. Thus, incoming messages can be noticed although one would not be directly at the computer all the time. Only 27% said they just sign in to see if someone else was online.

All test users were living in Finland – a country where SMS has proven to be extremely popular among people of all ages but especially young people. All users also stated that they use SMS, although only 27% said they use it daily, 45% weekly, and 27% a couple of times per month. When asked if they were to use IM from their mobiles phones, 82% said that they might use it and only 18% said they would not use it. One person said he did not understand the difference between mobile IM and SMS.

After the test, more information about the subjects was gathered with another questionnaire, which can be found in Appendix C. Two of the eight users were female and six were male. Two of the test users defined themselves as novice, four as intermediate, and two as expert users in using IM. The experience factor was evaluated with two questions: the first was how many times the user had used

IM in general, and the second asked the user to define the level of expertise he felt he was from four options: beginner, novice, intermediate and expert.

5.2.2. Test Organization

The software application used in the test was Windows Messenger. It was expected to be the most popular application among the test candidate population. This assumption turned out to be wrong after all. The most popular application among the test candidates was ICQ (64% used it), while the MSN/Windows Messenger was used by 45% of them. A question about the user's familiarity with the used application revealed that three users used it for the first time at the test. Two users said they had tried the application before, but were not using it regularly and three users told they had used the particular messenger several times before.

The used usability test methods were pair testing and a pair interview afterwards. The interview could be better described as a group discussion. The interview questions are listed below.

- What did you think of this evaluation? Were any of the tasks hard? Is there something we could do better with the tasks? Better scenarios maybe, or something else?

- Did you have any technical problems during the test? Did you have problems with the software or with the hardware?

- Did you find differences in the three different set-ups? Do the different tasks make it hard to compare the set-ups?

- What do you think of IM in general? What is good about it? What is bad about it?

- Would you use IM from you phone? What kind of problems do you think it has? What kind of benefits?

- Do you have any additional comments about anything regarding this evaluation session?

After each test session, IM between a 2G mobile phone and a laptop was demonstrated to the subjects and they were invited to try it out themselves as well.

A Sony Ericsson P800 phone with the Agile Messenger[6] was used as the mobile end in this demonstration.

The most important data gathered from the test were the subjects' opinions about the performance of the Windows Messenger in all set-ups. The opinions were collected with the questionnaire filled in afterwards and the interview session. As was suggested in [Sea 00], the subjects were told about the delay differences between set-ups only after data about their satisfaction to the IM system had been gathered. Data about the actual delays experienced during the test was collected with the Ethereal Protocol Analyzer.

The evaluation setting was organized as follows: we had two rooms prepared for the evaluation. User A was always sitting in one room and user B in the other room, from where the emulator was also operated. The rooms were located so that the users could not hear a conversation had in normal voice in the other room, and they could not see the emulator while performing the test tasks. A diagram of the set-up is shown in Figure 5-3.

[6] http://www.agilemobile.com/ (3.1.2005)

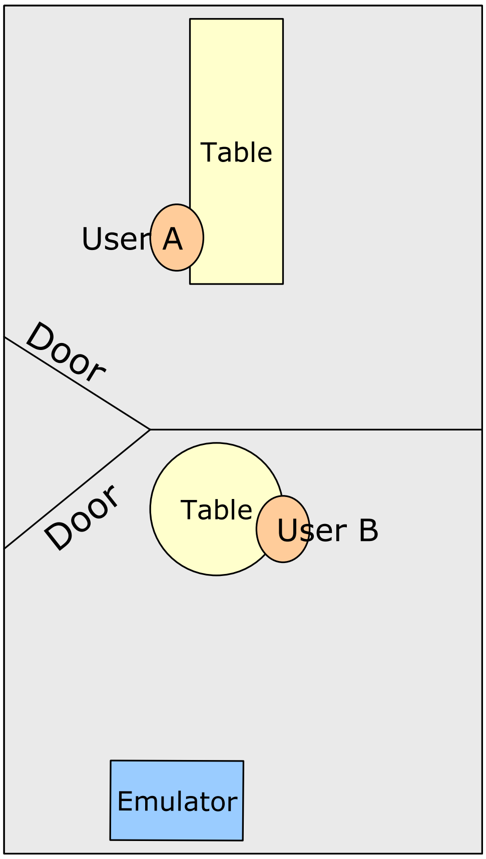

Figure 5-3. The usability evaluation set-up.

Both users were provided with a standard laptop, running Windows XP, which was connected to the emulator. They each also had an instructor guiding them through the evaluation session. The instructor in the emulator room was also managing the emulator. Each use case was given to the users in written form on a piece of paper.

In the beginning of the test session, the users were asked to sit down and they were introduced to the instructors. The test progression was explained to both users. They were told that they would have three tasks to do in three different set-ups, which differ from each other in delay characteristics. After the short introduction, the users were taken into separate rooms. In the test rooms, the users were shown the basic functionality of the used client application, if they were not familiar with it before. In the beginning of each test task (use case), the scenario was explained and given to the subject on a piece of paper. The same instructor stayed with the subject throughout the test session.

The order of the three set-ups was different with each pair, so that the effects of learning could be minimized. Learning here means getting used to the hardware and software used in the test as the session progresses. E.g., the touchpad used instead of a mouse in a laptop often takes time to get accustomed to. Also, some users were new to the used software application.

5.2.3. Use Case Scenarios

There were three separate use cases to do, which included at least logging in to the application and IM. The users got to use each of the three set-ups – the two WCDMA set-ups introduced in Section 2.5 and fixed Internet – with one use case. The presence function was tested in some tasks.

The background scenario for all the use cases was the same. The pairs were told about Axl / Kicki and Lars / Amelia (the names were chosen according to the test users gender each time) who were good friends and met each other regularly on their free time.

The actual tasks that were made during the test are shown in Appendix B. The main purpose for the tasks was to make the IM situation as natural as possible. In other words, the outcome of the task was not important, but the fact that the task forced the users to communicate in almost real time was. When the use case task was accomplished, e.g., Axl and Lars had set a lunch date, the subjects were told they can freely experiment with the application and the set-up for couple of minutes before moving on with the test.

5.3. Usability Evaluation Results

After the pilot round, minor changes were made to the test task guidelines. Namely, some extra information was added, so the guidelines would be more understandable. One laptop used in the pilot test round, which was considerably smaller in size compared to the other, was changed. The pilot round user had difficulties with the small keypad and the trackpoint[7] device. After the change both users had almost identical laptops, which were both equipped with a trackpoint device and a touchpad.

All four test sessions went practically as planned. In a couple of sessions, it took a bit longer for the test subjects to relax than in others, and in some sessions the users first had difficulties to get use to the hardware and software they were doing

[7] Trackpoint is a cursor-pointing device in the keyboard, which looks like a small button.

the test with. The test tasks were sometimes found to work a little different than what was thought when planned. Especially, the initiation of conversation did not always work as planned. At times, the users had to wait for the conversation to start because they logged in at different times. The subjects had fairly long conversations each time they were doing the tasks, so in the end the tasks can be considered successful.

The instructors were sitting next to the subjects, while they were performing the test tasks. Instructors were watching the screen as well. At times, it felt that some subjects might have found this disturbing. Sometimes light discussion was needed to make the test situation more relaxed for the subjects.

In the after test questionnaire, the users were asked to rate the set-ups from worst (slowest) to best (fastest), if they could. These results are displayed separately in Table 5-1. The rest of the results acquired from the questionnaire are shown in Table 5-2. Please, note that the subject could mark several answers to each question, which means that the added percentage for one question sometime exceeds 100%.

Table 5-1. Set-up rating after test, ranking 1=best 3=worst.

Set-up rating	**Pair 1**			**Pair 2**			**Pair 3**			**Pair 4**		
A	Performance			Internet	3	3	Performance			Capacity	1	1
B	Internet			Performance	1	2	Capacity			Internet		2
C	Capacity			Capacity	2	1	Internet			Performance		3
No difference		X	X					X	X		X	

In Table 5-1, the set-ups A, B and C denote the three test tasks, which were always made in alphabetical order. The order of the three set-ups is shown separately for each pair. The subjects were asked to rate the set-ups with numbers 1-3, where 1 was the best set-up and 3 the worst. If they saw no difference, they were asked to cross the "No difference" –box. Internet was the quickest set-up, meaning that it had the least extra delay, the second quickest was the performance optimized WCDMA, and the slowest was the capacity optimized WCDMA.

For example, for pair 1 it took 2.59 seconds for User 1 to login in the fixed Internet set-up. In the capacity optimized set-up it took 8.87 seconds to accomplish the same login procedure. Login in the fixed Internet set-up was 6.29 seconds faster.

The IM session is interesting as well. It took 23.80 seconds in the fixed Internet set-up before the IM session was open and User 2 got the opening message. From this, the session opening itself took 1.11 seconds, and sending the message took 0.0001 second. The time needed for User 1 to write the message in and send it was about 22 seconds. For User 2 it took over 1 minute (about 78 seconds) to read the message from User 1 and send an answer to it. In the capacity optimized set-up, it took 58.57 seconds to open the session and deliver the first message. This time the session opening lasted 5.58 seconds, over 4 seconds longer than in the fixed Internet. The time needed for User 1 to write the first message and send it was about 52 seconds, transmission of the message via the server took 0.7099 seconds. Thus, the transmission took 0.7098 seconds longer than over the fixed Internet. This time it took User 2 about 36 seconds to read the message from user 1 and reply to it, twice as fast as in the fixed Internet set-up. This time the answer was very short compared to the previous answer.

The complete time needed to open the session and exchange two messages was about 102 seconds in the fixed Internet set-up and 95 seconds in the capacity optimized set-up. It is easy to see that the writing delay is the main time consumer in the IM session. Writing differs about 30 seconds, because the user was writing a longer message as an opening in the capacity optimized WCDMA set-up than in the fixed Internet set-up.

Some evidence of learning could be seen from the answers pair 2 gave to the listing. User 2 has rated the set-ups from slowest to fastest in the order they were made. On the other hand, there is no correlation between the ratings of different user pairs, except the ones who said they saw no difference. This gives some evidence on the learning not having effect on the results.

From the results in Table 5-1, it can be concluded that the extra delays do not affect the user perceived delay. Two pairs said straight away they could not rate the set-ups in any order. Two pairs tried to rate them, but with poor success. Pair 2 rated the Internet set-up to be the worst, although it was the fastest. The subjects in pair 2 also had different opinions on which of the WCDMA set-ups were faster. In pair 4, both subjects thought the capacity optimized set-up, thought to be the worst, was actually the best set-up. Their opinions for the second and third place were again different; one subject said he could not rate the two WCDMA set-ups, and the other had an opinion to the right direction about the fastness.

When the subjects were filling out the evaluation form after the test, they were in the same room, and one could easily see what the other one was writing. This could explain why inside one user pair both either tried to rate the set-ups or just crossed the “no difference” box. This could, also, partly explain the correlation of the results inside pairs 2 and 4.

Table 5-2. Results from the after test questionnaire.

How familiar are you with instant messaging?	**% of subjects**
I have used it more than 100 times	62.5 %
I have used it more than 10 times	37.5 %
I have used it less than 10 times	0.0 %
I had never used it before	0.0 %
What level of expertise would describe your instant messaging skills best?	
Beginner	0.0 %
Novice	50.0 %
Intermediate	37.5 %
Expert	25.0 %
How familiar MSN was?	
Had used several times before	25.0 %
Had used, but no expert	12.5 %
Had tried it a couple of times before	25.0 %
This was the first time	37.5 %
Did you find the use case scenarios realistic?	
Yes	100.0 %
No	0.0 %
Was the use of messenger natural in this test?	
Yes, usually the conversation flows like this	100.0 %
No, usually the dialog is quicker	12.5 %
No, for some other reason	12.5 %
Gender	
Female	25.0 %
Male	75.0 %
What do you do during the week?	
Study	25.0 %
Work full-time	87.5 %
Work part-time	12.5 %
Something else	0.0 %
Your latest education	
Elementary	0.0 %
High school	0.0 %
Polytechnic	0.0 %
University	100.0 %

Table 5-2 shows that all our test subjects were, in fact, familiar with IM already before the test; actually, most of them rated themselves to the intermediate or expert level in IM use. However, 37.5% of them were using the particular application for the first time, which could have some impact on the results. All subjects stated that their use of messenger was natural in the test and that the IM conversation flow is usually similar to what they experienced in the test session. Some users had added extra comments to the questions stating that their use of messenger was natural, but usually they had more lively conversations. This is an important notion, since, e.g., Ackerman and Palen noted that in a lively conversation in chat or a chat like instant, the messages fly by so fast that some messages are left unanswered [Ack 96]. The IM conversations, also, easily meander off topic. This would suggest that in a lively conversation the IMers do not have time to wait and read every message, which on the other hand would mean that they probably do not pay attention to the transmission delay of a single message either.

In the interview that followed the test, several subjects stated that they usually knew the people they were messaging to and also knew their IM habits. For example, one subject told that he has a regular IM buddy who is always doing several things while he is in an IM session. He told that because he knows this buddy's behavior, he never waits for an answer from him right away. Other important notions were that the subjects said they know in advance it will take time to write an answer. Therefore, they are prepared to wait for a moment for the answer to begin with.

The messaging behavior of the subjects seemed to vary depending on the level of expertise. It was observed during the test that the expert users were writing several short messages to say one sentence, whereas the less experienced users were writing whole sentences at a time. Bradner, Kellogg and Erickson have discovered that many users prefer short messages in IM. One subject in their research actually stated that IM messages should not be longer than one line [Bra 99]. The expert users might also start new topics during pauses in the conversation, when they were waiting on an answer from the opponent. Sending short messages shortens the so-called writing delay (time required to write the message) dramatically and therefore makes the conversation more active. It was also observed during the test sessions that several users complained that their pair was a slow writer. This happened often when there was a longer pause in the IM discussion.

In the interview, the subjects were in the end asked about their views and expectations about mobile IM, partly based on the demo made with the GSM/GPRS mobile phone. Many users showed interest in mobile IM, but almost all also said they find it frustrating and slow to type longer messages on such a small device. One pair agreed they would probably not use mobile IM, because just making a phone call is so much easier.

5.4. Summary

IM over WCDMA does not differ significantly from IM used directly over fixed Internet. One user reported he might have noticed lengthening in the login procedure, but was not sure about this. As we saw in the analysis part, the transmission delay could have been noticed easier during the login than during the IM session. In the IM session the main part of the pauses come from writing and reading the messages. The majority of the users did not see significant differences between the three set-ups. This includes both novice and expert users.

The IM applications display the users own messages in the conversation instantly, although they have not been sent to the receiver yet. This makes the user feel, that the other user on the other end, is just typing his answer slowly and so causing the delay. The user does not always realize that the conversation he is seeing on his end is not the same his counterpart is seeing. This clever design used in the applications, results in that the user cannot separate the network delays from the writing delays.

Nielsen states that three well-known limits for response time are

- 0.1 second to make the user feel the response was instant,
- 1.0 second to make the user feel uninterrupted,
- 10 seconds to keep the users attention [Nie 93].

According to these limits, our results from the technical analysis would suggest that people would not feel uninterrupted and sometimes it would be hard to keep the users' attention. But then again, these limits concern the human-computer interaction. In our case, we should more likely talk about human-human or human-computer-computer-human interaction.

The test tasks might have been forcing the users to wait for answers, but still some of them started to use the application as they were used to; sending messages and not caring if the counterpart had answered or not. A clear distinction in this behavior could be seen between novice and expert users of IM. The expert users already had an idea of how the communication flows in an IM conversation and acted accordingly, where as the novice users often acted as in normal conversation and waited for the reply before sending a new message.

Mobility brings new challenges to IM use. During the interviews it became obvious that the greatest obstacle seems to be the user interface that is offered in mobile devices; the hardware is too small for fluent IM. Many users stated that sometimes they feel that even typing SMS is too painful on such a small device.

6. Conclusions

IM is a popular way of communicating between friends. To use IM, one needs a terminal, e.g., a PC or a mobile phone, and an IM network. Messages are exchanged between users at different terminals in almost real-time.

New, faster mobile networks are opening all around the world. They enable the use of Internet-based services from a mobile terminal, such as a mobile phone. Most IM services reside in the Internet, to which users have to connect to, in order to use their services.

In Europe, the technique chosen for the 3G mobile networks is WCDMA. WCDMA will eventually enable bit rates up to about 14 Mbps for mobile data transfer. Radio resources are, however, limited, and different strategies are used to share them between users in a cell.

We have tested IM over two different WCDMA set-ups. We have compared two different IM solutions over these set-ups: the Windows Messenger and a SIP-based application. Tests were made with two basic use cases: login and IM. The technical performance analysis showed that there is significant difference between the proprietary Windows Messenger's and the SIP-based client's functions. The main differences between the two IM solutions were that the SIP client functioned locally and the server was a private server, while the Windows Server was used over the fixed Internet just as in normal use, and it additionally functioned through a Socks v5 proxy server.

The SIP client needed only a couple of IP packets to accomplish the chosen use cases, whereas the Windows Messenger sent tenfold IP packets to accomplish the same use cases. The SIP client seemed to surpass the Windows Messenger, when the delay statistics in login and IM session opening were compared. The Windows Messenger, however, outdid the SIP client when it came to messages sent during the IM session. It can be argued, which characteristic is more important to the users.

After the technical analysis was concluded, we made usability tests for both WCDMA set-ups with the Windows Messenger. The objective of the user tests was to reveal, if the users noticed the change of underlying network from the lengthening of the delays. Four user pairs tested IM over the two WCDMA configurations and over fixed Internet, with the help of use case tasks.

We saw in Chapter 5 that IMers are often multi-taskers. They are doing other things while IMing. Maybe Nielsen's guidelines for appropriate response times [Nie 93], originally cited from Miller [Mil 68], were in the right direction. The IM users, however, do not care about the long delays; they do other things while they are waiting for their buddy to answer. IMers, also, seem to adjust their communication style and pace different according to the buddy on the other end.

In our test setting, the users had almost no distractions; they had no other activities besides accomplishing the tasks they were asked to do via IM. Still our results imply that the users could not identify, which set-up was technically the slowest or the fastest. The alternation in only server processing time could cause same difference in delay time, in normal use, as the change of underlying network did in our evaluation.

In conclusion, IM use over WCDMA does not hinder the service significantly. As the interview answers suggest, greater challenges probably arise with the choice of terminal device, and getting the users accustomed with the new user interface.

7. Future Work

In the future, it would be interesting to conclude a similar study using smaller mobile terminals, e.g., 3G mobile phones, instead of laptops. Another field to study would be mobile users reaction to delays in this kind of interaction. The existing guidelines are based on stationary users using a computer. The situation where the user is mobile, and the terminal is handheld is quite different.

It should be noted that the SIP-based client used in this thesis is not in commercial use at the moment. Thus, it is hard to say if its function would be the same with millions of users. A comparison between these results and results from a commercially launched SIP client would be interesting.

References

Validity of all Internet sources checked 5th January 2005.

[Ack 96] Ackerman, M.S. and Palen, L.A. "The Zephyr Help Instance: Promoting Ongoing Activity in a CSCW System." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (13-18.4.1996, Vancouver, Canada).* ACM Press, USA, 1996, p. 268-275.

[Bra 99] Bradner, E., Kellogg, W.A. and Erickson, T. "The Adoption and Use of BABBLE: A Field Study of Chat in the Workplace." *Proceedings of 6th European Conference on Computer-Supported Cooperative Work (12-16.9.1999, Copenhagen, Denmark).* Kluwer Academic Publishers, New York, 2002, p. 139-158.

[Cam 02] Camarillo, G. *SIP Demystified*. McGraw-Hill, USA, 2002.

[Cam 04] Camarillo, G. and Garcia-Martin, M. A. *The 3G IP Multimedia Subsystem*. John Wiley & Sons Ltd., West Sussex, UK, 2004.

[Dah 98] Dahlman, E., Beming, P., Knutsson, J., Ovesjö, F., Persson, M., and Roobol, C. "WCDMA – The Radio Interface for Future Mobile Multimedia Communications." *IEEE Transactions in Vehicular Technology*, rev A, 1998, Volume 47, p. 1105-1118.

[Fau 00] Faulkner, X. *Usability Engineering.* Macmillan Press Ltd., UK, 2000.

[Gri 01] Grinter, R. E. and Eldridge, M. "y do tngrs luv 2 txt msg?" *Proceedings of 7th European Conference on Computer-Supported Cooperative Work (16-20.9.2001, Bonn, Germany).* Kluwer Academic Publishers, Netherlands, 2002, p. 219-238.

[Gri 02] Grinter, R. E. and Palen, L. "Instant Messaging in Teen Life." *Proceedings of ACM 2002 Conference on Supported Cooperative Work (16-20.11.2002, Louisiana, USA).* ACM Press, USA, 2002, p. 21-30.

[Hal 04] Hall, M. Article: "Instant Messaging's 'Big Three' Agree to Interoperate." *Instant Messaging Planet*, July 15th, 2004. (http://www.instantmessagingplanet.com/enterprise/article.php/3381531)

[Hin 03] Hindocha, N. "Threats to Instant Messaging," Symantec Security Response, White Paper. *Symantec Corporation*, USA, 2003.

[Hol 00] Holma, H. and Toskala, A. *WCDMA for UMTS*. John Wiley & Sons Ltd., West Sussex, UK, 2000.

[Lin 04] Lintula, P. Suoraviestintää ja läsnäoloa Sip:llä. Master's Thesis, *Tampere University*, Tampere, 2004. (http://www.cs.uta.fi/research/theses/masters/Lintula_Petri.pdf)

[Man 00] Manohar, R. "Mobile Satellite Services: Spectrum Requirements of Mobile-Satellite Services." *World Radiocommunication Conference 2000.* Istanbul, Turkey, 8.5.-2.6.2000. (http://www.itu.int/newsarchive/wrc2000/presskit/)

[Mil 68] Miller, R. B. "Response Time in Man-Computer Conversational Transactions." *Proceedings of AFIPS Spring Joint Computer Conference.* Volume 33, p. 267-277.

[Nie 93] Nielsen, J. *Usability Engineering*. Academic Press Inc, USA, 1993.

[Nil 04] Nilsson, C. and Svensson, J." Informationssamhällets framväxt, En trendrapport." *World Internet Institute*, Sweden, September 2004.

[Rub 03] Rubens, P. Article: "Will instant messaging be the new texting?" *BBC News*, June 30th 2003. (http://news.bbc.co.uk/2/hi/technology/3031796.stm)

[Sea 00] Sears, A. and Jacko, J. A. "Understanding the Relation Between Network Quality of Service and the Usability of Distributed Multimedia Documents." *Human-Computer Interaction*, *EBSCO*, 2000, Volume 15, p. 43-68.

[Shi 04] Shiu, E. and Lenhart A. "How Americans use instant messaging." *Pew Internet & American Life Project*, USA, September 2004.

[Sne 04] Snellman, K. "Mobiilipalvelumarkkinat Suomessa 2003." *Ministry of Transport and Communications*, Finland, April 2004.

Standards and Specifications

Enabler Release Definition for IMPS. Open Mobile Alliance.
Version 1.2 – 22 May 2004.
http://www.openmobilealliance.org/release_program/docs/IMPS/v1.2-20040522/OMA-ERELD-IMPS-V1_2-20040522-C.pdf

RFC 1928. SOCKS Protocol Version 5. M. Leech, et al. IETF, March 1996. Internet Standards Track.
http://www.ietf.org/rfc/rfc1928.txt

RFC 2327. SDP: Session Description Protocol. M. Handley and V. Jacobson. IETF, April 1998. Standards Track.
http://www.ietf.org/rfc/rfc2327.txt

RFC 2543. SIP: Session Initiation Protocol. M. Handley, et al. IETF, March 1999. Standards Track. Now obsolete.

RFC 2617. HTTP Authentication: Basic and Digest Access Authentication. J. Franks, et al. IETF, June 1999. Standards Track.
http://www.ietf.org/rfc/rfc2617.txt

RFC 2779. Instant Messaging / Presence Protocol Requirements. M. Day, et al. IETF, February 2000. Informational.
http://www.ietf.org/rfc/rfc2779.txt

RFC 3113. 3GPP-IETF Standardization Collaboration. K. Rosenbrock, et al. IETF, June 2001. Informational.
http://www.ietf.org/rfc/rfc3113.txt

RFC 3261. SIP: Session Initiation Protocol. Core protocol specification; obsoletes RFC 2543. J. Rosenberg, et al. IETF, June 2002. Standards Track.
http://www.ietf.org/rfc/rfc3261.txt

RFC 3265. Session Initiation Protocol (SIP)-Specific Event Notification. A. B. Roach. IETF, June 2002. Standards Track.
http://www.ietf.org/rfc/rfc3265.txt

RFC 3428. Session Initiation Protocol (SIP) Extension for Instant Messaging. B. Campbell, et al. IETF, December 2002. Standards Track.
http://www.ietf.org/rfc/rfc3428.txt

TS 22.141 V6.2.0. Presence Service; Stage 1. 3GPP, March 2003. Technical Specification.
http://www.3gpp.org/ftp/Specs/html-info/22141.htm

TS 23.141 V6.7.0. Presence Service; Architectural and functional description. 3GPP, September 2004. Technical Specification.
http://www.3gpp.org/ftp/Specs/html-info/23141.htm

TR 24.841 V6.0.0. Presence service based on Session Initiation Protocol (SIP); Functional models, information flows and protocol details. 3GPP, June 2004. Technical Report.
http://www.3gpp.org/ftp/Specs/html-info/24841.htm

WWW-Pages

http://www.3gpp.org/
3rd Generation Partnership Project (3GPP)

http://www.ietf.org/
Internet Engineering Task Force (IETF)

http://www.openmobilealliance.org
Open Mobile Alliance (OMA)

http://www.openmobilealliance.org/tech/affiliates/wv/wvindex.html
OMA: Archived documents from the Wireless Village Initiative (WV).

http://www.cs.columbia.edu/sip/
Session Initiation Protocol (SIP)

http://www.hiddensoft.com/autoit3/
AutoIt v3, An open source scripting language for automating the Windows GUI.

http://epp.eurostat.cec.eu.int/portal/page?_pageid=1090,1&_dad=portal&_schema=PORTAL
Eurostat, Statistical Office of the European Communities

http://www.tilastokeskus.fi
Statistics Finland

http://www.ericsson.com
Ericsson Ltd.

Applications

Ethereal Network Analyzer Protocol. Originally by Gerald Combs, distributed under GNU General Public Licence.
http://www.ethereal.com

MSN Messenger. Microsoft Corporation
http://messenger.msn.com

Yahoo! Messenger. Yahoo! Inc.
http://messenger.yahoo.com

AOL Instant Messenger. American Online, Inc.
http://www.aim.com

ICQ. ICQ, Inc.
http://www.icq.com

Agile Messenger. Agilemobile.com Co. Ltd.
http://www.agilemobile.com

Appendix A

Instant Messaging Preliminary Survey

Instant Messaging Survey

This is a preliminary survey for a usability evaluation for instant messaging. All information you submit in this form will be handled confidentially. It will take about 5 minutes for you to complete this survey. Not everyone who answers this survey is asked to participate in the actual evaluation. To most questions, you can choose one or more suitable answers. You may fill in more details in the text boxes if needed. Please answer all relevant questions.

Please send you answer before Saturday 20th of November. Thank you for your participation in this survey!

Name: ______________________________

E-mail: _____________________________

Age category:
◉ 18 - 25 ○ 26 - 35 ○ more than 35

The actual survey starts here.

How often do you use instant messaging (such as MSN Messenger, ICQ, or similar)

☐ Daily
☐ Weekly
☐ Couple of times a month
☐ Couple of times a year
☐ Never (if you answer this, you can go straight to the last question)

Which client application(s) do you use/have you used?

☐ AOL Instant Messenger (AIM)
☐ ICQ
☐ MSN Messenger / Windows Messenger (they are almost the same)
☐ Yahoo! Messenger
☐ Some other, what? ____________________________

Which one do you prefer, why?

____

Why do you use instant messaging?

☐ It is a fun way to communicate
☐ It is a cheap way to communicate
☐ It is a fast way to communicate
☐ It is an easy way to communicate
☐ I use it, because all my friends use it
☐ Some other reason, what? ____________________________

For what do you use instant messaging?

☐ to chat with friends
☐ to solve problems at work
☐ to play games
☐ to other stuff, what? ____________________________

Where do you use instant messaging?

☐ At work
☐ In school
☐ At home
☐ Some other place, where? ____________________________

When do you use instant messaging?

☐ When I have to ask/tell something to someone who uses instant messaging
☐ When I have spare time during the day
☐ Some other time, when? ____________________________

How long are you usually "online"?

☐ I just sign in to check if someone is there
☐ I stay online while I am doing stuff on the computer
☐ I stay online almost all the time

If you have a mobile phone, how often do you send text messages (SMS)?

☐ Daily
☐ Weekly
☐ Couple of times a month
☐ Less than couple of times a month
☐ Never

If you could use instant messaging from your mobile phone, would you use it?

☐ Yes, definitely
☐ I might, but depending on the situation
☐ No, probably never

☐ *I don't really understand how it differs from SMS*

Additional comments on instant messaging? What is best about it? What is worst? Feel free to tell about your opinions!

The evaluation is done with a pair. If you already know a pair, let me know! Your pair has to fill out this form as well.

If you don't care who is your messaging partner in the evaluation session, leave this blank.

Comments about this questionnaire?

Submit Reset

Thank you for participating in this survey!

Appendix B

Usability Test Tasks

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up A (evaluator 1)

You want to ask one of your friends, Kicki, to go to see a movie on the next day. You know that Kicki is today at an important seminar on a ship, and don't know if he can answer the phone right now. The tickets need to be reserved right away if you want to get them. So you are already in the tram on your way to the place to buy them from.

- You decide to see first if Kicki is online in messenger (it is Monday of week 1)
- log in to the Windows Messenger with the given username and password
  - E-mail (username): LarsMagnus2004@hotmail.com
  - Password: abc983K
- When he is online, you can start the conversation by double-clicking his name on the contact list
- When you agree on the time and the movie you are done and may end the discussion.

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up A (evaluator 2)

- log in to the Windows Messenger with the given username and password
  - E-mail (username): kicki2004@hotmail.com
  - Password: Hp90twK
- Change your presence state so that your friends see you are busy at the moment
  - Ask your instructor for advice if you don't know how to do this
- Imagine you are in a seminar on a ship. You have lectures for the most of the day. During the lecture you can't talk on the phone or anything, but some of the lectures are so boring that you wish you would have something else to do.
- It is Monday of week 1
- This is why you have logged on to messenger. But the current lecture is very interesting, so you decided to change your state to busy so that you can listen.
- Now the interesting part is over and some uninteresting mambo jambo starts, you decide to be available online again to be able to chat with friends.

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up B (evaluator 1)

- You are in the bus travelling from Joensuu to Helsinki on Friday of week 1. The bus is full, and the person sitting next to you is sleeping.
- The bus is arriving in Helsinki at 22:00 and you want to ask one of your friends to join you for a beer then, but you don't want to be rude and wake up the person next to you by talking on the phone
- See if some of your friends are online and available for a conversation
    - Log in to the Windows Messenger with the given username and password
    - E-mail (username): LarsMagnus2004@hotmail.com
    - Password: abc983K

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up B (evaluator 2)

- log in to the Windows Messenger with the given username and password
    - E-mail (username): kicki2004@hotmail.com
    - Password: Hp90twK
- You are in the public library reading room, just passing time by reading magazines while waiting for you Karate class to start in the same building. The class before you started 20 minutes late, so yours will start late as well.
- It is Friday of week 1
- You have also logged on to messenger to see if some friends would be online

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up C (evaluator 1)

- You are in Russia on a business trip. You are leaving tomorrow to go home, but you have already been on the trip for a while.
- It is Monday, of week 2.
- You decide to log on to Messenger to see if you could arrange to go to lunch with some of your friends later this week
- Log in to the Windows Messenger with the given username and password
  - E-mail (username): LarsMagnus2004@hotmail.com
  - Password: abc983K
- You see if anyone is online

Background scenario: Kicki, Amelia, Axl and Lars are good friends who meet each other occasionally for free-time activities like movies, etc.

Set-up C (evaluator 2)

- log in to the Windows Messenger with the given username and password
  - E-mail (username): kicki2004@hotmail.com
  - Password: Hp90twK
- You are working on the computer and as you are working you are logged on to messenger
- It is Monday, week 2
- You decide to see, what there is for lunch in your favourite restaurant Smökki this week. The menu can be found at http://www.unicafe.fi/ruokalistat_show/?db=ruokalistat_show&subpage=servinmokki_english

Kicki's calendar

Week 1

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Seminar on ship Important lectures from 8-19	Swimming 18-19 in Mäkelän-rinne			Karate class 20:30-21:30		

Week 2

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Working on the computer all day!	Swimming 18-19 in Mäkelän-rinne	Lunch with Mom	Meeting at 10-12 in Otaniemi	Karate class 20:30-21:30		

Lars' calendar

Week 1

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Get movie tickets for tomorrow! (Catwoman)	Movie at 17:45, 19:00, or 20:30	Go to Joensuu	In Joensuu	Back from Joensuu at 22:00		Fly to Russia

Week 2

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
In Russia	Fly back from Russia at 17:20				Go to Tavastia to see The Rasmus at 23:00	

Appendix C

Usability Evaluation Questionnaire

Please fill out this form. The questions are about the usability test you just conducted.

How familiar are you with instant messaging?

- ☐ I have used instant messaging more than 100 times
- ☐ I have used instant messaging more than 10 times
- ☐ I have used instant messaging less than 10 times
- ☐ I have never before tried instant messaging

What level of expertise would describe your instant messaging skills best?

- ☐ Beginner
- ☐ Novice
- ☐ Intermediate
- ☐ Expert

How familiar was MSN Messenger to you in advance?

- ☐ I have used it several times before and had no problems with it
- ☐ I have used it before, but I am not an expert in using it
- ☐ I have tried it a couple of times, but I don't use it regularly
- ☐ This was the first time I used it

You tested instant messaging in three (3) different set-ups. Please, rate the set-ups if you can (1 = best, 3 = worst).

- ☐ Set-up A
- ☐ Set-up B
- ☐ Set-up C
- ☐ I didn't see any difference

Please explain the differences you noticed if you can.

Did you find the use case scenarios realistic?

- ☐ Yes
- ☐ No

Comments:

____

Continues on the other side →

Was your use of the Messenger in this evaluation natural?

- ☐ Yes, usually the conversation flows like this
- ☐ No, usually the dialog is quicker
- ☐ No, for some other reason

Comments:

____

In the end, some background information

Gender: ☐ Female ☐ Male

What do you do during the week?

- ☐ Study
- ☐ Work full-time
- ☐ Work part-time
- ☐ Something else

Your education (currently ongoing or last one concluded)

- ☐ Elementary school
- ☐ High school
- ☐ Polytechnic (*ammattikorkeakoulu*)
- ☐ University

Appendix D

Results from the preliminary survey about IM and presence service use.

Age category	**%**
18-25	81.8 %
26-35	18.2 %
over 35	0.0 %
How often do you use instant messaging?	
Daily	50.0 %
Weekly	8.3 %
Couple of times a month	33.3 %
Couple of times a year	0.0 %
Never (if you answer this, you can go straight to the last question)	8.3 %
Which client application(s) do you use/have you used?	
AOL Instant Messenger (AIM)	9.1 %
ICQ	63.6 %
MSN Messenger / Windows Messenger (they are almost the same)	45.5 %
Yahoo! Messenger	18.2 %
Some other, what?	45.5 %
Why do you use instant messaging?	
It is a fun way to communicate	18.2 %
It is a cheap way to communicate	63.6 %
It is a fast way to communicate	72.7 %
It is an easy way to communicate	81.8 %
I use it, because all my friends use it	18.2 %
Some other reason, what?	9.1 %
For what do you use instant messaging?	
to chat with friends	72.7 %
to solve problems at work	36.4 %
to play games	9.1 %
to do other stuff, what?	27.3 %
Where do you use instant messaging?	
At work	45.5 %
In school	18.2 %
At home	81.8 %
Some other place, where?	18.2 %
When do you use instant messaging?	
When I have to ask/tell something to someone who uses instant messaging	81.8 %
When I have spare time during the day	36.4 %
Some other time, when?	0.0 %

How long are you usually "online"?	
I just sign in to check if someone is there	27.3 %
I stay online while I am doing stuff on the computer	54.5 %
I stay online almost all the time	18.2 %
If you have a mobile phone, how often do you send text messages (SMS)?	
Daily	27.3 %
Weekly	45.5 %
Couple of times a month	27.3 %
Less than couple of times a month	0.0 %
Never	0.0 %
If you could use instant messaging from your mobile phone, would you use it?	
Yes, definately	0.0 %
I might, but depending on the situation	81.8 %
No, probably never	18.2 %
I don't really understand how it differs from SMS	9.1 %