



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

VOIP V JABBER KLIENTU

VOIP IN JABBER CLIENT

SEMESTRÁLNÍ PROJEKT

TERM PROJECT

AUTOR PRÁCE

AUTHOR

Bc. VOJTĚCH KULIČKA

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. JOZEF MLÍCH

BRNO 2011

Abstrakt

Práce se zabývá možnostmi přidání funkcionality do existujícího XMPP programu se sdílenou tabulí. Analyzuje možnosti využití současných technologií pro podporu VoIP. Cílem je port klienta na komunikační architekturu telepathy a implementace VoIP.

Abstract

This thesis tackles the issues of implementing a VoIP support into an XMPP based IM application. The state of the art is analyzed to find a suitable technology to base the VoIP on. The work's goal is to port the existing client application to network framework telepathy and implentation of VoIP.

Klíčová slova

VoIP, IM, sdílená tabule, XMPP, telepathy

Keywords

VoIP, IM, Shared whiteboard, XMPP, telepathy

Citace

Vojtěch Kulička: VoIP in jabber client, semestrální projekt, Brno, FIT VUT v Brně, 2011

VoIP in jabber client

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením pana inženýra Jozefa Mlícha

.....

Vojtěch Kulička

May 22, 2011

Poděkování

Děkuji svému vedoucímu Ing.Jozefu Mlíchovi a Ing.Jaroslavu Řezníkovi za odbornou pomoc.

© Vojtěch Kulička, 2011.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Contents

1	Introduction	2
2	Makneto	4
2.1	Current architecture	4
2.2	Motivation for porting and adding VoIP support	6
3	XMPP	8
3.1	XMPP protocol	8
3.2	Jingle	13
4	Telepathy	15
4.1	Telepathy architecture	16
4.2	Empathy	22
5	Existing solutions	23
5.1	Applications with whiteboard and VoIP support	23
5.2	Multimedia streaming protocols	26
5.3	VoIP	27
6	Port to Telepathy	31
6.1	Connection	31
6.2	Whiteboarding	33
6.3	Contact list	33
7	VoIP implementation	35
7.1	GStreamer	35
7.2	Phonon	37
7.3	Farsight	37
7.4	VoIP in Makneto	38
8	Conclusion	39

Chapter 1

Introduction

Human is a social creature and likes to chat, share feelings and ideas. At first we managed to do so by making simple sounds. Those sounds later on developed into words. Then much later the human race started to feel the need to record what we were thinking. We made up symbols and started to write. As the society grew and spread, we wanted to communicate with people from other tribes and villages. At first we would travel and use spoken words, but as the distances grew we figured we can have our thoughts delivered in writing. Mail was born. In 1844 telegraph was invented by Samuel Morse followed by telephone in 1874 by Alexander Graham Bell. And finally in 1969 the Internet was created. All of these inventions aimed to provide means of communication to satisfy the needs of the evolving society.

In the early days of the Internet email was the main means of communication. And just like regular mail people would have their electronic mailboxes to which the emails were delivered. Email was a huge step forward for it provided a way to almost instantly deliver text from one place to another regardless of the distance for free. The main disadvantage of email is that people had to check their mailboxes read new mail and then reply. It is just neither fast nor convenient enough for team cooperation when team members are far apart. For those and other purposes like chatting Instant Messenger programs were introduced.

An IM program offers realtime communication between two people via text messages that are delivered from one user to another instantly. Instant messengers became very popular and started adding on features like multiuser chat, various games and most importantly VoIP (Voice over IP) support. VoIP capable IM like skype have become extremely popular at first for making it possible for people to call each other for free over the internet. Later video conferencing capability was added, so you could talk and see you colleague at the same time. One more thing comes in extremely handy when working in a team - a whiteboard.

At this time there is no usable IM providing VoIP and shared whiteboard for GNU/Linux. This thesis aims to add VoIP support to an existing XMPP client with shared board called Makneto. Makneto was created by Jaroslav Řezník as a master's thesis in 2008. At this point it is using iris library for XMPP communication. The shared board data is also transferred over XMPP. One of the goals of this thesis is to port Makneto to telepathy, which is now a very reliable and robust library for communication for numerous protocols.

The following chapter 2 gives a detailed description of the current version of program Makneto. We will find out about it's architecture, strengths and weaknesses.

Chapter 3 focuses on XMPP/Jabber communication protocol. It talks about it's features and limitations. The chapter contains concrete examples that give the reader an idea

38 of how it works and is it so popular.

39 Chapter 4 is about Telepathy communication framework, how it works, what it consists
40 of. There is also a description of a IM client application Empathy based on Telepathy.

41 Current audio and video streaming protocols are listed and discussed in chapter 5.
42 Based on this chapter a suitable protocol is chosen for the implementation. This chapter
43 also talk about VoIP and what should the implemantator have in mind when writing a
44 VoIP application.

45 The port to Telepathy and new Makneto's architecture are in chapter 6. It talks about
46 design decisions and reasoning behind them.

47 Chapter 7 explains how was the VoIP implemented, what problems were encountered
48 and what is the result.

Chapter 2

Makneto

The collaborative application Makneto is the main topic of this chapter. First we take a look at the architecture and what it is based on. The following section explains why port Makneto to Telepathy and add VoIP. The main sources are [41, 25, 22].

2.1 Current architecture

Makneto is an instant messenger with a shared board capability. It was written by Jaroslav Řezník as his master's thesis in 2008. Makneto is written in C++ using Qt version 4 and KDE 4 libraries. Qt is a application framework. It extends C++ and adds dynamic functionality like signals and slots mechanism or property system. Signals and slots are used instead of callback function, because they are as opposed to callback functions type-safe. The dynamic property system allows objects to have properties added at runtime without previous specification in their header file. In order to use the extended functionality the objects must directly or indirectly inherit QObject, which all Qt classes are derived from. These mechanisms also require to be compiled by the Meta Object Compiler before any standard C++ compiler.

Qt runs on all major computer platforms like GNU Linux, Microsoft Windows and Mac OS X. Moreso it is supported by mobile device platforms such as Symbian and Microsoft Windows Mobile. There is also a port to Google Android called Lighthouse. It was developed by company named Trolltech and was available under two licenses. First was a commercial license allowing companies to write proprietary applications for a fee. Second was GNU General Public License, which was of course free. Thanks to the commercial license the Qt documentation is one of the best among any software available under GNU GPL. In june 2008 Troltech was acquired by Nokia. Nokia decided to make the source code available so that anyone could contribute. In January 2009 with the release of Qt 4.5 another licensing option was added - Lesser General Public License [25, 26].

Besides Qt Makneto utilizes functionality provide by KDE 4 [20] libraries, which makes Makneto unable to run on a different operating system than GNU Linux. KDE is a desktop environment created by Matthias Ettricht in 1996 for he felt the need for a good quality window manager for Linux. KDE is currently in version 4, which brought great features. There is a new desktop called Plasma, which allows users to display widgets called plasmoids directly on the desktop. That allows users to have a TODO list, calendar, translator, weather forecast, system monitor and much more right in front of them on their desktops without having launch any of those to get the information they need. It makes users'

83 work easier and more efficient. KDE and GNOME are the most common window manager
 84 in Linux. GNOME offers stable useful environment and is influenced mostly by large
 85 businesses using it. KDE is more flexible and works more with the look and feel.

86 Makneto communicates using XMPP/Jabber protocol implemented in Qt-based C++
 87 library called Iris. All of the Iris is primarily used by Psi instant messenger. It's development
 88 is still quite active and it supports all of Jabber's key features. Iris' downside is very poor
 89 documentation. It's wiki is very brief and all to all gives one example. The only way to
 90 find out how it works is browsing through Psi code [30].

91 Makneto's shared whiteboard is based an official extension of XMPP SVGWB by Joonas
 92 Govenius [11]. SVGWB is implemented in Psi and Makneto utilizes their code. SVGWB
 93 defines all the necessary actions like whiteboard session initiation including invitation and
 94 mainly how to send and receive information about the graphical objects using XMPP text
 95 messages. The actual graphical object representation is defined by SVG [21](Scalable Vector
 96 Graphics) by W3C. SVG describes graphics objects using vectors in XML format. Makneto
 97 uses just a subset of SVG called SVG Tiny [22] as it does not need all of the features of SVG.
 98 Tiny SVG describes two-dimensional vector graphics and raster graphics and multimedia.
 99 To sum up Makneto's whiteboard features, it allows you to draw lines, rectangles, ellipses,
 100 circles, sketch using a paintbrush and input images in jpg and png formats. Graphical
 101 objects are resizeable, can be rotated and copied.

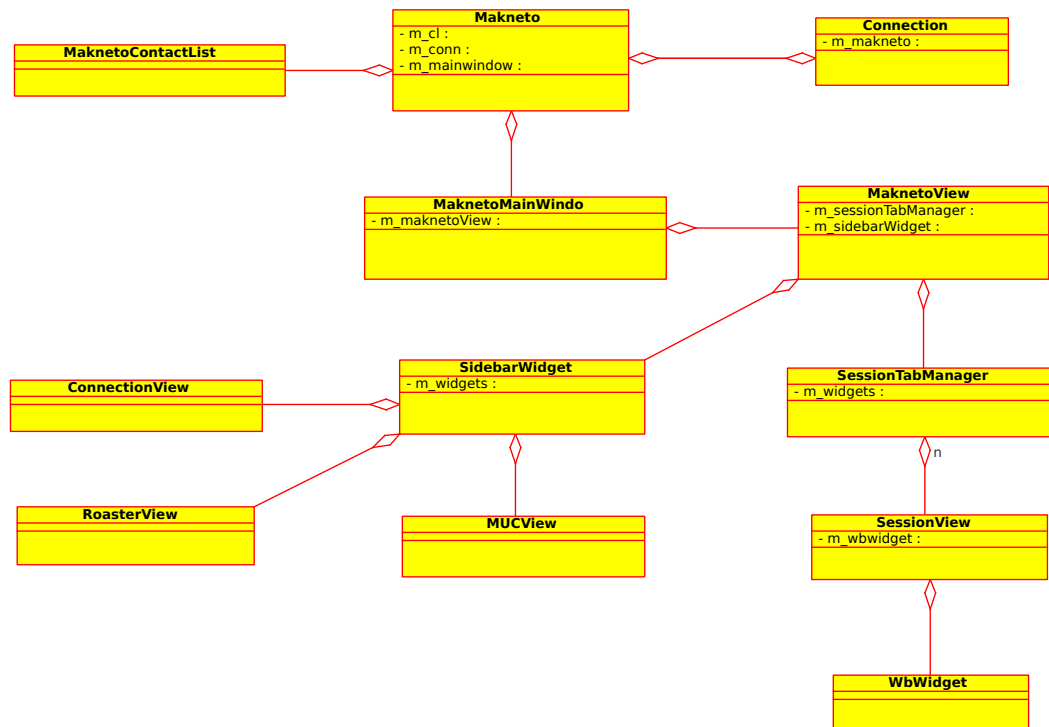


Figure 2.1: Makneto class diagram

102 Makneto runs in one window, which you can see in figure 2.2 and is represented by class
 103 **MaknetoMainWindow** as shown in the class diagram in figure 2.1. User's contact list is on a

104 panel on the left hand side as a part of stacked widget along with MUC¹ tab and tab for
 105 controlling presence. The whiteboard and chat session are initiated through the contact
 106 list. Makneto handles more session at once each in a separate tab. Each tab is represented
 107 by a `SessionView` object and the whiteboard within it by `WbWidget`. The subject to
 108 change is class `Connection` that is built on top of iris library. Using the application is very
 109 comfortable although it from dies time to time.

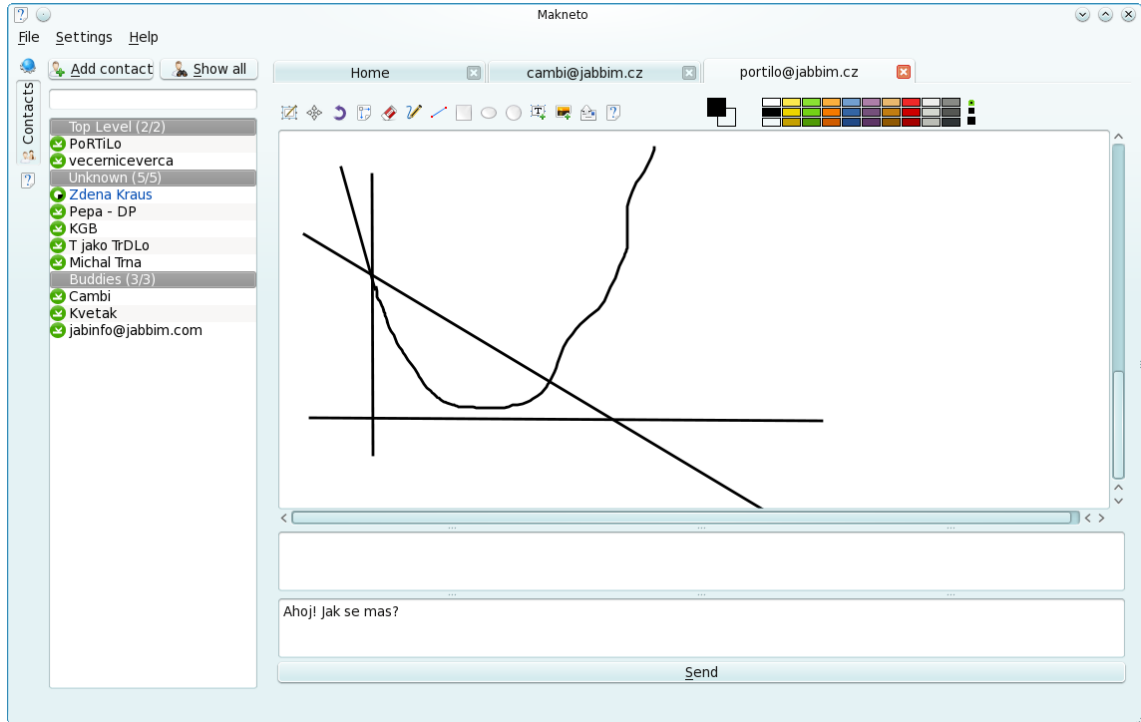


Figure 2.2: Makneto before any work was done on it.

110 2.2 Motivation for porting and adding VoIP support

111 Makneto has a potential to become a great collaboration tool. The tasks of today are more
 112 and more difficult so the need to solve the task in teams is greater and greater. Especially in
 113 computer science the teams are often from all over the world and it is important to discuss
 114 current issues. Meeting in person is not an option and that is where Makneto comes in.
 115 The shared whiteboard is very useful for sharing thinkmaps and visualising problems and
 116 solutions, but the days when people had the time and patience to write are gone. Everybody
 117 wants to talk these days.

118 Makneto is at this point has couple design flaws. First it uses iris library for com-
 119 munication in XMPP network. Iris is very poorly documented and every change in the
 120 implementation would reflect in Makneto. Second Makneto depends on KDE libraries,
 121 which at this point is not necessary. Getting rid of this dependency will help increase
 122 Makneto's portability.

¹Multi-User Chat

123 The current implemenation of Makneto is UI and backend in one large application.
124 The goal is to separate the backend and UI. Backend will take care of communication.
125 That means text messaging, shared whiteboard and after this thesis is finished also VoIP.
126 Backend will use communications framework telepathy described in the following chapter.
127 Telepathy supports various protocols as well as voice and video calls. Makneto will be able
128 to use XMPP, ICQ, IRC, MSN etc. with one implementation of a client.
129 Frontend shall be a subject to change based on target device. Makneto for PC will have
130 a Qt4 frontend and Makneto for smartphones and tablets will use the Qt Quick UI. With
131 rapidly increasing number tablets and smartphones made and sold it would be shame not
132 to plan to support them.

Chapter 3

XMPP

This chapter talks about Extensible Messaging and Presence Protocol [8], it's history, key features and usage. The information is mainly acquired from [34, 32, 33].

In 1999 Jeremie Miller created protocol called Jabber. Jabber was an open protocol based on XML as opposed to existing protocols like ICQ [13] or AIM [27], which were proprietary and owned and controlled by private companies. The first attempt to make Jabber a standard failed. The second attempt did not use the name Jabber, but eXtensible Messaging and Presence Protocol instead. IETF¹ approved and XMPP was standardized in 2004 in RFC 3920 and RFC 3921 called XMPP Core and XMPP IM respectively. Development of XMPP is still active as it is based on XML it is possible to add functionality without jeopardising the compatibility of existing implementations.

XMPP is a robust, scalable, secure, open and extensible protocol that has been well tested over the years passing all of the test without any sign of trouble. It has been very well thought out and altogether it is a great protocol. The description given below is in some of the parts simplified. Full description is out of the scope of this thesis.

3.1 XMPP protocol

XMPP is defined in [32] and [33], which describe all the key features of the protocol. Authors of XMPP aim for a scalable, extensible and in every way powerful protocol. Years later with XMPP still around and more and more popular we can safely say they succeeded. The key features of XMPP include:

- **Decentralized architecture** - XMPP network does not rely on one server. The network consists of servers and clients. Every client may run his or her own server. XMPP server registers clients and communicates with other servers much like with email(see figure 3.1). A client often wants to communicate with another client on a different server. In such case other client's server is looked up and the message forwarded. If a client is not satisfied with the server he or she registered with he or she may simply register with a different server or run his or her own. There is also no way to take the whole network out with DoS² or DDoS³ attack. There simply is not a small number of target to attack.

¹Internet Engineering Task Force

²Denial of Service

³Distributed Denial of Service

- **Open nature** - XMPP is an open standard that can be used by anyone without having to pay, sign any agreement or behave by any restrictive policies. Also it's fate is not in hands of one company but rather in the hands of everyone. Anyone who wishes to contribute can do so by cooperating with XMPP Standards Foundation.
- **Extensibility** - thanks to XML based nature of XMPP it is very simple to add new features without discontinuing support of the old ones. This feature makes XMPP very flexible for it can easily add new functionality and it has been already done couple times.
- **Security** - with built-in support for TLS and SSL there is no need to worry that the messages might be read by a third person using man in the middle attack. Companies using XMPP for communication inside their network might run their server locally with no access to the outside world.

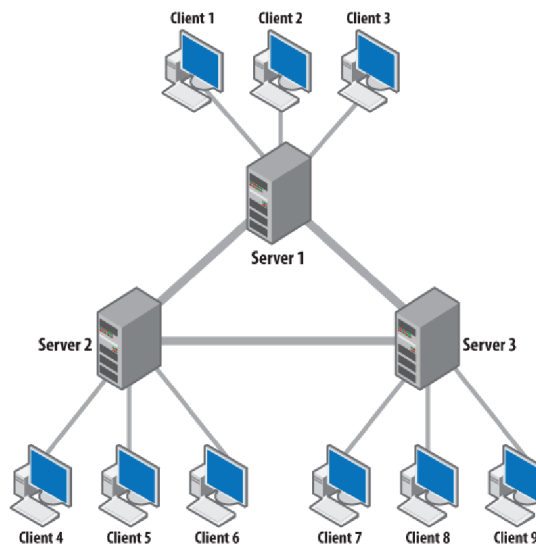


Figure 3.1: XMPP decentralized architecture [34]

Now let's look at the protocol and how to use it.

Client-server communication

XMPP is basically streaming XML documents. The stream is an unbounded XML document which contains another XML documents from both client and server. The XML documents enclosed in the stream tags with a depth of 1 are called stanzas. Stanza is a basic unit of communication like a packet. The following stanzas are defined:

- **message** - used for getting information from one place to another in a push manner. The message stanza is not acknowledged and no answer is expected. It is used for instant messages, alerts, notifications, groupchat etc. The nature of the message is specified by **type**.

- 185 • **presence** - stanza for acquiring another user's presence. XMPP honors users' privacy
186 and to get someone's presence status he or she need to authorize it first by adding
187 the querier to his or her contact list. There are several types of presence just like in
188 any other IM protocol.
- 189 • **IQ** - is an abbreviation for Info/Query. This stanza is used for everything else. IQ
190 works on question-answer basis. It is used for example for getting remote client's
191 capabilities. IQ stanza must always receive a reply.

192 First a TCP connection is established between client and server. Once established a
193 stream is opened by client by sending the server `<stream>` tag. There is an example of
194 opening a stream below:

```
1 <?xml version='1.0' encoding='UTF-8'?>
2 <stream:stream
3   xmlns='jabber:client'
4   xmlns:stream='http://etherx.jabber.org/streams'
5   to='\"jabbim.cz\"'
6   version='\"1.0\"'>
```

Example 3.1: Opening communication with server

195 Server answers with a second stream back to the client, which is shown in the piece of
196 XML code that follows.

```
1 <?xml version='1.0'?>
2 <stream:stream
3   xmlns='jabber:client'
4   xmlns:stream='http://etherx.jabber.org/streams'
5   id='856661962'
6   from='jabbim.cz'
7   version='1.0'
8   xml:lang='en'>
```

Example 3.2: Server's reply to client's opening stream

197 The next step is to negotiate properties of the stream. The server sends an XML
198 enclosed in `stream:feature` tags, informing the client about features it supports. The most
199 important property is by far encryption and authentication. Preferred encryption method
200 is TLS, but SSL is also an option. However some servers may require usage of TLS. TLS is
201 recommended for both client-server and server-server communication. Authentication is a
202 key responsibility of the server. The server must ensure that users attempting to connect
203 to it are who they say they are. The server acts as a gateway to the entire network and
204 must not allow identity spoofing. Authentication is done via SASL⁴ and options supported
205 by the server are enclosed in `mechanism` tags. The example, where server supports TLS
206 and SASL using PLAIN text or MD5 is shown below:

⁴Simple Authentication and Security Layer

```

1 <stream:features>
2   <starttls
3     xmlns='urn:ietf:params:xml:ns:xmpp-tls' />
4   <mechanisms xmlns='urn:ietf:params:xml:ns:xmpp-sasl'>
5     <mechanism>
6       PLAIN
7     </mechanism>
8     <mechanism>
9       DIGEST-MD5
10    </mechanism>
11  </mechanism>
12  ...
13 </stream:features>

```

Example 3.3: Server sends supported encryption options

207 Once stream parameters are set, client can request presence, set his or her own presence
208 and communicate with other clients. Sending a message to a fellow user is accomplished
209 via **message** stanza and may look like this:

```

1 <message
2   from='vtheman@jabbbim.cz'
3   to='kuba86@jabber.cz'
4   xml:lang='en'>
5   <body>Hey, how are you?</body>
6 </message>

```

Example 3.4: Chat message

210 It must include **to**, **from** attributes and **body** opening and closing tags. The first specifies
211 who the message is addressed to using a Jabber ID. Jabber ID consists of
212 **user name @ domain name [/ resource]** and many people mistake for email address be-
213 cause the resource is optional. Resource was added to JID to allow user to connect from
214 different device/locations without having to log out. An example of resource is **home** or
215 **work**. A simple Jabber ID is **vtheman@jabbbim.cz**. Attribute **from** naturally contains a
216 jabber ID as well, but this time an ID of the sender. The **body** element contains the actual
217 message.

218 Next stanza is presence and it is needed for following presence of the contacts in the
219 users' roster. To get presence of a user he or she must approve of it. Once user has been
220 authorized to get another users presence they have both subscribed to get each other's
221 presence. After connecting to the server a user sends initial presence stanza: **<presence/>**.
222 From that moment on the server takes care of the presence. Whenever the user's presence
223 changes, it sends notification to all subscribers in the user's roster. Similarly if someone
224 else's presence changes the user gets notified by his or her server.

225 Presence stanza contains elements **show** and **status**. **show** can be either **chat** meaning
226 available and ready for chat, **away** meaning the user is not at the PC at the moment, **xa**
227 indicates the user will be gone for a longer period of time and finally **dnd**, which stands for
228 do not disturb.

```
1 <presence from="vtheman@jabber.cz" to="kuba86@jabber.cz">
2   <show>dnd</show>
3   <status>Working on my thesis!</status>
4 </presence>
```

Example 3.5: Presence publication

229 The last of stanzas is Info/Query shortly IQ. IQ is very similar to HTTP in methods
230 and in the query-answer nature. IQ queries use method **get** for requesting information and
231 method **set** for making requests based on provided information. Answer to a query is IQ
232 stanza of type **result** and contains information requested by **get** method or acknowledge
233 in case of **set** method. The last type of IQ is **error** and is used to indicate that something
234 went wrong. A good example of an IQ stanza is acquiring a roster:

```
1 <iq type="get">
2   <query xmlns="jabber:iq:roster"/>
3 </iq>
```

Example 3.6: Requesting roster from server

235 And the server replies with:

```
1 <iq type="result">
2   <query xmlns="jabber:iq:roster">
3     <item jid="kuba86@jabber.cz"/>
4     <item jid="rezza@jabber.cz"/>
5     <item jid="imlich@jabber.fit.vutbr.cz"/>
6   </query>
7 </iq>
```

Example 3.7: Roster sent by the server

236 The stream ends with `</stream>` tag and that means end for the communication.

237 XMPP Extensions

238 As mentioned earlier XMPP is easily extensible due to its XML based architecture. XMPP
239 extensions are published by XSF⁵ as XEP⁶. Basic XMPP functionality is defined in the RFC
240 3920 and RFC 3921 and every XMPP server and client ought to implement it. Functionality
241 described in XEP however is optional. These are the most popular extensions:

- 242 • **Multi-User Chat** defined in XEP-0045 [31] allows users to create virtual rooms just
243 like in IRC, invite their contacts to join the room and thus communicate with multiple
244 people.
- 245 • **Service Discovery** registered as XEP-0030 [17] defines a way of findit out what
246 capabilities have one's contacts.

⁵XMPP Standards Foundation

⁶XMPP Extension Protocol

247 • **Entity Capabilities** defined in XEP-0115 [18] adds client's capabilities to presence
248 information, so that if a XEP-0115 capable client requests presence it receives a list
249 of supported features as well.

250 There are tens of extensions either standardized or waiting for becoming a standard
251 defining various handy features.

252 3.2 Jingle

253 XMPP Extension defined in XEP-0166 [36] known as Jingle is a signaling protocol that ini-
254 tiates, manages and terminates media sessions via XMPP. Jingle was first used in Google
255 Talk [10] for Voice call signaling in 2005. The idea was to use an existing XMPP co-
256 munication channel to setup a peer-to-peer media session that uses a different means of
257 transporting data, e.g RTP⁷ for voice or video and TCP for file transfer.

258 Figure 3.2 shows how the protocol works. Session initiation starts when the initiator
259 sends **session-initiate** with Application type, e.g voice call, and Transport method, e.g
260 UDP are described. As you can see in the figure Jingle uses an IQ stanza so the initiator
261 immediately receives a IQ result acknowledging the invitation reception from the responder.
262 Next all the necessary application type and transport type parameters of the session are
263 negotiated. In case of voice call the application type parameters might be audio codec
264 and sampling frequency. Transport type parameters include the peers' IP addresses and
265 ports and transport method. When all the parameters have been setup the responder
266 either accepts or declines the invitation. If accepted the data start flowing between the two
267 peers just as it was agreed during the initiation phase. Jingle can also be used to adjust
268 parameters of an existing session if necessary. And finally one of the peers sends the other
269 **session-terminate** to end the session.

⁷Real-time Transport Protocol

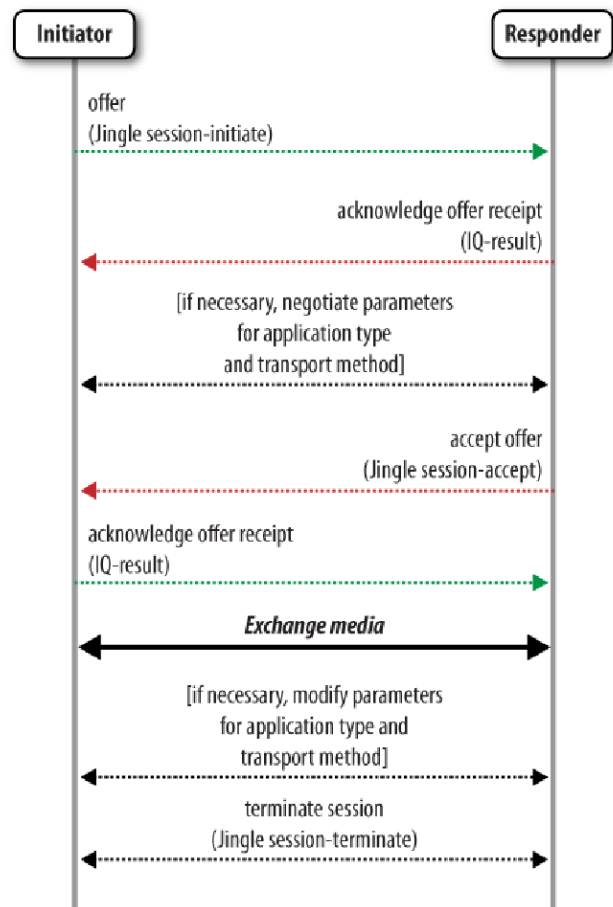


Figure 3.2: Data flow in media session initiation, management and termination using XMPP Jingle [34]

Chapter 4

Telepathy

In this chapter the framework to which Makneto shall be ported is looked at. First basic terms and basic idea of how is the framework designed what can it do. Then we describe D-Bus and the relationship with Telepathy and have a detailed look at components of Telepathy. Finally we talk about an existing client built on Telepathy. The chapter takes information mainly from [7, 4].

Telepathy [6] is a modular communications framework for building real-time communication applications. It supports numerous communication protocols as pluggable backends e.g XMPP/Jabber(telepathy-gabble), SIP(telepathy-sofiasip), MSN(telepathy-butterfly), etc. Each of telepathy's components runs in a separate process as desktop service and communicates via D-Bus. The components are shared by telepathy clients. For example if there are two clients using XMPP they both use the same instance of telepathy-gabble. To get a better idea of how this concept works take a look at figure 4.1 [7].

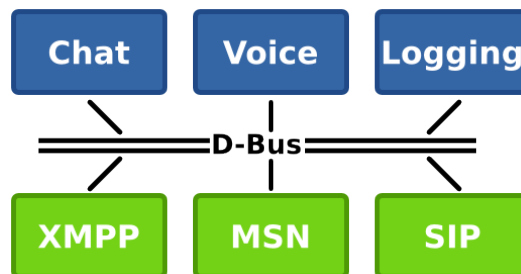


Figure 4.1: Telepathy architecture [7]

There are several features making telepathy very useful as a communications framework.

- **Robustness** - all the components are independent. If one crashes, others will not be affected
- **Ease of development** - the components can be replaced without having to stop the service
- **Language independence** - since telepathy components use D-Bus for communication among themselves, any language that has D-Bus binding might be used to write them

- 292 • **Desktop independence** - D-Bus is present in both main Linux window managers
293 GNOME and KDE, so the same telepathy components could be backend for appro-
294 priate frontends.
- 295 • **Code reuse** - the client applications do not have to worry about protocol specifics,
296 which are handled by Telepathy. The client can use more protocols by making no or
297 small alterations to the code.
- 298 • **Connection reuse** - more than one Telepathy client can use the same connection
299 simultaneously:

300 4.1 Telepathy architecture

301 Telepathy is a very powerful framework and as such it is also complicated. To successfully
302 write programs using Telepathy we need to know what telepathy consists of and what it is
303 based on. This section tries to put all the terms into context of this thesis.

304 D-Bus

305 D-Bus is a kind of inter-process communication. It allows two applications running in
306 different processes, written in different programming language communicate. More so these
307 applications may communicate directly, without having to go through message bus daemon.
308 There two types of D-Bus. First is a system bus used for events such as “USB device
309 disconnected” or “printer out of paper.” Second type is per-user-login-session bus, which
310 is used by user applications. D-Bus low level API is represented by libdbus and it requires
311 XML parser(libxml or expat) to work. Higher level language bindings such as Qt, GLib,
312 Java etc. are built on top of libdbus and offer more convenient way of using D-Bus, although
313 they add more dependencies [4, 7].

314 Each process that wants to communicate over D-Bus will need to use most the following
315 depending on it’s nature:

- 316 • **Unique name** - is an unique id (e.g. 2.1) assigned by D-Bus daemon to the client
317 application. Unique name is similar to a public IP address.
- 318 • **Wellknown name** - is similar to a DNS name. If a process wants to make a service
319 available to other processes it requesets a wellknown name. If another process wants
320 to access the service it uses the wellknown name to do so. Wellknown name might
321 look like this: org.freedesktop.Telepathy.ChannelDispatcher.
- 322 • **Object path** - is a path to an object that is exported by process running a service.
- 323 • **Interface** - is a way of requesting a service using signals or methods. Each D-Bus
324 client must register at least one interface and each interface provides at least one
325 method or signal. Every interface needs to have to name like a wellknown name.
- 326 • **Method** - is impleneted in the object specified by object path and exposed in the
327 interface for that object for other processes to use.
- 328 • **Signal** - is a D-Bus signal client process can connect to it’s callback function. If a
329 signal is invoked the callbacked function is called.

330 • **Property** - is used for exposing D-Bus object's properties. To do so the object must
 331 implement org.freedesktop.DBus.Properties interface.

332 The following figure 4.2 shows an example of two programs connected to D-Bus to be
 333 able to communicate with each other. Program B provides a service called well-known
 334 name (org.freedesktop.foo.Bar) and it's id is 1.3. Program A does not provide any
 335 service and thus does not need any well-known name. It just needs an id 1.2 to use other
 336 programs' services [7].

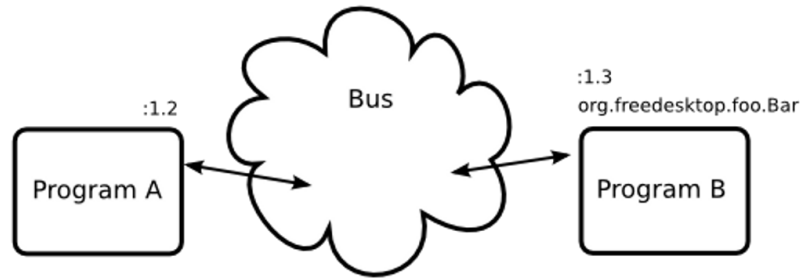


Figure 4.2: D-Bus id and wellknown name example [7]

337 The figure 4.3 shows an overview of all of the terms described above in a simple diagram.

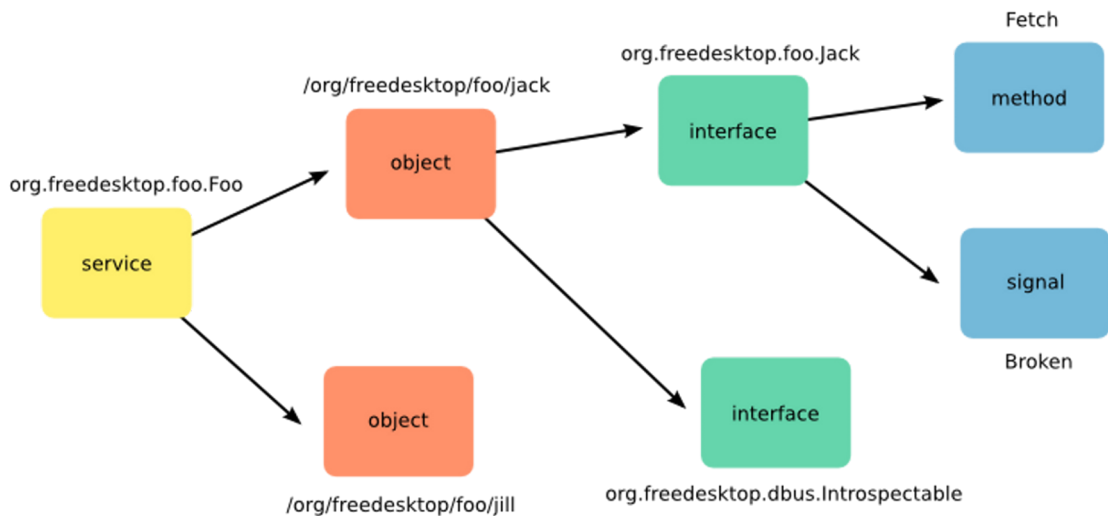


Figure 4.3: D-Bus architecture [7]

338 D-Bus is a key component of Telepathy framework. Telepathy supports many protocols
 339 all of which might provide different capabilities. For example IRC does not support avatars
 340 while XMPP does. Eventhough avatar feature is supported by XMPP protocol it might
 341 not be supported by the server we are connected to or by the opposite client in case of
 342 peer-to-peer connection. The available features are exposed by D-Bus Properties Interface.
 343 That is an easy way of determining the protocol, server or client capabilities [4].

344 Mission Control

345 Mission Control is a Telepathy component that implements Account Manager and Channel
346 Dispathter and it's primary purpose is to encapsulate those two. There always only one
347 running on the system. It is very important to check version running for they are backwards
348 incompatible [7]. For better understanding of the role of this and the following components
349 have a look at figure ?? . It illustrates the relationship among the the key components.

350 Account Manager

351 Account Manager is responsible for handling accounts(e.g. XMPP, ICQ, MSN etc.). It is
352 accessible by well-known name on D-Bus - `org.freedesktop.Telepathy.AccountManager`.
353 A client application first creates an account using the provided methods, supplying it with
354 Connection Manager, protocol and display name. Account Manager creates and then as
355 long as the Account is active maintains Connection to that Account. The Accounts are
356 persistent and the next time the application runs it does not need to create them again.

357 To create a Connection a Connection Manager is called by Telepathy. An Account may
358 be valid or invalid and also enabled or disabled. Valid ones may establish a Connection,
359 whereas invalid can't. Enabled and disabled property are user's way of telling the applica-
360 tion which one should be ignored and which not. The lists of valid and invalid accounts is
361 accessible via appropriately named methods as are the enabled and disabled. [7].

362 Account

363 Accounts are created via Account Manager. They are only created once and then stay on
364 the system and can be accessed by any Telepathy clients. An active Account object registers
365 with D-Bus and has an object path `/org/freedesktop/Telepathy/Account/CM/PROTOCOL/ACCDN`.
366 CM stands for Connection Manager(e.g. gabble, salut, butterfly, etc.), PROTOCOL is substitu-
367 tion for a protocol name and ACCDN is Account's Display Name. The Account object im-
368 plements `org.freedesktop.Telepathy.Account` interface. Features supported by this
369 interface depend on the protocol used and the server-side software.

370 The Account settings are done via `iorg.freedesktop.Telepathy.Properties` inter-
371 face. All available attributes can be obtained by calling a single method provided by the
372 properties Interface. The method is very convenient for it returns all the propeties at once
373 in signle D-Bus call. Setting all properties at once works exactly the same. Some features
374 are available via specified interface, e.g. avatar. Avatar used to be a property, but now it
375 has it's own interface. The Interfaces property of the account lists all interfaces of additional
376 features [7].

377 Connection Manager

378 Connection Manager supplies Account Manager with Connections. It is not directly used
379 by the client program. Account Manager requests connection for active accounts. There is
380 always only one Connection Manager for each protocol running at any time.

381 Connection Manager is a protocol-dependent Telepathy component. Different proto-
382 cols need different Connection Managers. For example if the client application wants to
383 communicate using XMPP/Jabber it has to use Telepathy-gabble and for MSN Telepathy-
384 butterfly is required. Some Connection Managers can communicate via more than one

385 protocol, for example Telepathy-haze. To see what protocols are supported there is an
386 appropriate method implemented by the Connection Manager [7].

387 Connection

388 Connection represents an active protocol session. It is associated with an Account and
389 is created by Connection Manager based on a request of the Account Manager. Connec-
390 tion implements org.freedesktop.Telepathy.Connection interface and additional interfaces
391 depending on the protocol. List additional interfaces available can be retrieved by checking
392 the Interfaces property. The most common interfaces are listed below[7]:

- 393 • **Contacts** - used to get as much information about a contact as asked in one D-Bus
394 call.
- 395 • **Aliasing** - serves for setting aliases for contacts and checking if the contacts have
396 changed their alias themselves.
- 397 • **Avatars** - interface to one of the most popular protocol features. Allows users to set
398 their avatars and retrieve other users' avatars.
- 399 • **ContactCapabilities** - retrieves capabilities of contacts' Clients to see what features
400 they support. Checks for example for VoIP or file transfer support.
- 401 • **Location** - lets user publish his or her current location as well as find out his or her
402 contacts' whereabouts.

403 Channel Dispatcher

404 This component handles Channels incoming from active Connections of valid Accounts.
405 Channel Dispatcher monitors available or activatable Telepathy Clients through D-Bus.
406 Clients register with user's session D-Bus and provide a CLIENT_NAME.client file. Both
407 of those serve as a way to publish Client's properties including a channel filter. The Channel
408 Dispatcher based on these properties knows what kind of a client it is and what type of
409 channels it is interested in (channel filter). If the Client is running then the properties are
410 acquired via the Client interface. If the client is not running and is activatable then the
411 .client file is used by Channel Dispatcher to pre-look up the properties and if they match
412 the incoming Channel, the Client is activated. So providing the .client file only makes sense
413 for activatable Clients.

414 When a Channel comes in from one of the Connections Channel Dispatcher notifies
415 appropriate Clients. There are three kinds of clients - Observer, Approver and Handler(see
416 section 4.1). The Channel is dispatched to all Observers and all Approvers with a matching
417 channel filter. The Approvers choose Handler to handle the Channel. Should the Client
418 fail, Channel Dispatcher may recover from such error and look for another Handler [7].

419 Channel

420 Channel allows the local client to exchange various kind of data with a remote server. It
421 is associated with a Connection and always implements at least two interfaces. The first
422 is org.freedesktop.Telepathy.Channel and the second depends on the Channel type.
423 Channels for text messaging will be of type Text and will implement

424 `org.freedesktop.Telepathy.ChannelType.Text` interface. The following list shows most
425 common types of Channels [7]:

- 426 • **ContactList** - used to get information of contacts in user's contact list.
- 427 • **Text** - designed for exchanging text messages.
- 428 • **Call** - used for VoIP and video calls.
- 429 • **FileTransfer** - Channel for sending and receiving files.
- 430 • **ContactSearch** - is used when a user wants to find a contact on a server.

431 Channels are created using two methods - `CreateChannel()` and `EnsureChannel()`.
432 These methods are implemented by both Channel Dispatcher and Connection. When calling
433 either of those methods on Channel Dispatcher the resulting Channel will go through the
434 procedure of looking for handler as described above. When using directly the connection
435 the calling application must handle the Channel itself as the Channel Dispatcher will not
436 interfere. It is also possible to supply the Channel Dispatcher with a preferred handler and
437 thus achieve the same effect. It is better to use the Channel Dispatcher for if the client
438 should fail it may dispatch the Channel to another handler [7].

439 Both of the methods provide a Channel. The difference between the two is that
440 `CreateChannel()` creates actual new Channel whereas `EnsureChannel()` will attempt to
441 reuse an existing Channel with the same properties. The first is typically used for file trans-
442 fer and contact search and the latter for text, streamed media and contact list Channels
443 [7].

444 Client

445 Client is an application that wants to use Telepathy. It needs to register a well-known name
446 in `org.freedesktop.Telepathy.Client` namespace, e.g. Empathy registers
447 `org.freedesktop.Telepathy.Client.Empathy`. Then it provides a `.client` file where purpose of
448 which is described above. Telepathy defines three types of clients - Observer, Approver
449 and Handler. All of these need to provide appropriate channel filter, e.g. Observer provides
450 `ObserverChannelFilter`. Based on the published filter the Channel Dispatcher dispatches
451 an incoming Channel to the Client or not.[7]

452 Observers are called upon a creation of a new Channel. They monitor Channels and
453 provide the acquired information to user. The observers have different functions based on
454 the type of observed Channel, e.g. Text Channel observer might serve as a logger and
455 FileTransfer observer as a file transfer progress monitor. Observer is must not interfere
456 except for when the user interaction like hitting the cancel button in a file transfer progress
457 window.[7]

458 Approver is a Telepathy Client that is supposed to accept the incoming Channel and
459 decide, which Handler it is dispatched to. The Channel Dispatcher provides Approvers with
460 a list of possible Handlers. Approver notifies the user of a new Channel and lets him or her
461 decide whether to accept or reject it. Similarly the user is allowed to choose which Handler
462 will handle the Channel. Handler might also be chosen by the Approver itself. Approver
463 does not call methods just like the Observer. Calling methods is up to the Handlers. For
464 example if there is an incoming file transfer the Approves lets user decide whether to accept
465 it or not, but the `AcceptFile` method will be called by the chosen Handler [7].

466 The last client is Handler. Handler does all the interaction with the Channel. A typical
 467 example of a Handler is chat-window. It displays messages and allows the user to send text
 468 messages back [7].

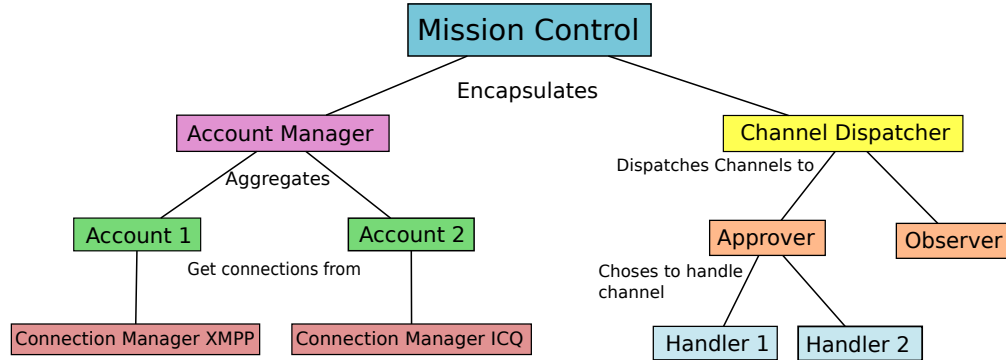


Figure 4.4: Simplified Telepathy architecture

469 The figure 4.5 represents typical setting using Telepathy. There is one Account Manager
 470 and one Channel Dispatcher for those are unique in the system. Next we can see Connection
 471 Managers, which allow the clients use numerous protocols. In general the number of
 472 supported protocols equals the number of present ConnectionManagers. Finally there are
 473 clients, which may an Observer for logging, Approver and a Handler. But it as well may
 474 be three handlers. Also one client may represent Handler, Approver and Observer at the
 475 same time. All of these entities communicate via D-Bus.

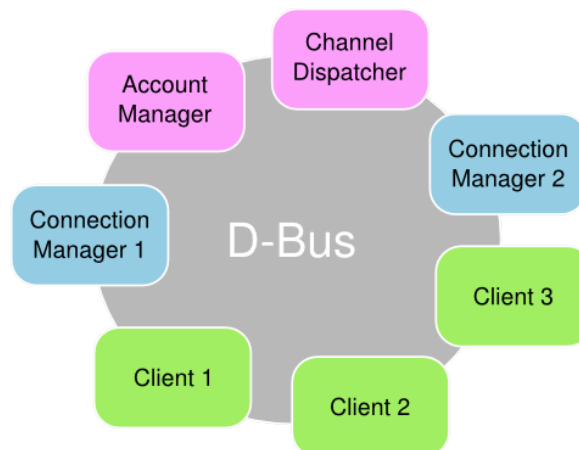


Figure 4.5: Telepathy components registered with user session D-Bus [7]

4.2 Empathy

Empathy is a multiprotocol instant messaging application based on Telepathy. In the terms described above Empathy is a Telepathy client. It is written in python using Telepathy-python bindings. Empathy registers a well-known name with D-Bus and communicates with Telepathy components to provide the communication services. Empathy supports text messaging, file transfer, voice and video calls over various protocols. Supported protocols include Google Talk, XMPP/Jabber, MSN, IRC, AIM Facebook, Yahoo!, Gadu Gadu, and ICQ. For some of those protocols like Google Talk and XMPP voice and video calls implemented. The support of the protocols depend on Telepathy Connection managers installed. Additional functionality includes sharing users' whereabouts among themselves, automatic reconnection when internet connection is reestablished and automatic changes of presence to away and extended away.[\[29\]](#)

The current stable version of Empathy is 2.32.2 and it is a default communication application in GNOME releases since version 2.24 instead of Pidgin. Empathy also replaced Ekiga - program for voice calls and video-call. It became an ultimate free communication tool. Empathy's GUI is takes after Gossip, which is an older IM application for GNOME.

Chapter 5

Existing solutions

The first section of this chapter talks existing applications supporting both VoIP and whiteboarding. The listed solutions were chosen based on relevance in the context of this work. The level of technical details of each of the applications depends on whether it is an open source or proprietary. The next section explains how to transfer multimedia over the internet. Finally the last section aims at VoIP and what must be taken into account when implementing it. The chapter is based on information from [35, 9, 23].

There is a number of existing communication applications offering voice calls and shared whiteboard. Some more popular, more advanced, more user-friendly, offering more features or better support than others. Some of those program support video call and some even conference calls. We shall go over the existing solutions that implement both shared whiteboard and voice calls. Among the described are Skype [37], Windows Live Messenger [24], Brosix [3], Yahoo! Messenger [40] and AIM [27].

5.1 Applications with whiteboard and VoIP support

Skype

Skype is the most popular and most used common VoIP application of all. Skype was released in 2003 and was developed by KaZaa[19]. It is available for all major computer platforms (Microsoft Widnows, Mac OS X and GNU Linux) as well as mobile platforms like Android and even Apple's iOS. Although skype is available for Linux it is not well supported. The latest version of skype for Linux is 2.1.0.81 Beta while skype for Windows is of version 5.1. Finally skype is built-in on more and more TVs.

Skype communicates using Skype protocol, which is proprietary. Recently much effort is being put in reverse engineering the skype protocol although the first attempt dates back to 2004 and was done in [35]. Skype encrypts the communication end-to-end with 256 bit AES so the amount of information acquired by packet sniffers is very limited. The motivation for recent efforts are simple. Skype is used by tens of millions of users every day, but the support for Linux is at this point almost nonexistent. Linux users have had enough and plan to create an open source client capable of communicating with skype. The wikipedia skype protocol page[38] is filling up with details.

While most of IM programs utilize client-server communication scheme, skype uses peer-to-peer model. The skype network consists of nodes, supernodes and login servers. See figure 5.1, which wasa taken from [35]. Nodes are clients. Each client keeps addresses of a number of supernodes. Supernodes are clients with good-enough bandwidth, public

526 IP address and enough memory and CPU power. Supernodes forward traffic to clients
527 behind device with network address translation or restrictive filters. It is believed that
528 skype uses something like STUN, which helps overcome NAT and was first defined in [15]
529 and then superseded by [16]. It seems that nodes themselves determine whether they are
530 behind address translating device or firewall. If two nodes want to communicate and either
531 of them or both is behind NAT then a supernode is used to forward traffic between them.

532 Skype call signalling is done via TCP and UDP is primarily used for the voice transfer.
533 If a skype client finds out it is behind a firewall that forbids UDP, the speech is transferred
534 using TCP. The codec used is unknown although there are couple candidates. What is
535 almost certain is the fact that skype uses wideband codec.

536 The decentralized architecture seems to be working well although there was an outage
537 on December 22nd 2010. Skype officials claim it was due to lack of supernodes[2].

538 Features of skype include calling landlines and cell phones and sending text messages
539 and vice versa - SkypeOut and SkypeIn. The feature list continues with voice and video
540 calls, multi-user chat, conference calls, voice mail and screen sharing. The newest and long
541 expected feature of video conference was introduced in may of 2010 in skype 5.0.

542 Though closed program, skype provides an API for developers who want to create
543 extensions called skype extras. Skype extras include a wide range of utilities that might be
544 plugged into skype. The extras uSeeToo, TalkAndWrite, WhiteBoardMeeting and Sketch
545 Pad all provide shared whiteboard each in their own way.

546 **Windows Live Messenger**

547 Windows Live Messenger was first released in July 1999 as MSN Messenger and offered just
548 text messaging with users of AOL Instant Messenger[27]. Due to AOL's constant effort to
549 block Microsoft from it's network Microsoft gave in and removed the feature. Since then
550 MSN Messenger could only connect to MSN Messenger Service. In 2001 with the release
551 of Windows XP, MSN Messenger 4.6 came out with voice call support. The last version
552 of MSN Messenger, version 7.5, featured video calls. Windows Live Messenger 8.0 was
553 released in June 2006 and that was the end of the name MSN Messenger.

554 Windows Live Messenger utilizes client-server model and communicates using Microsoft
555 Notification protocol over TCP. The first 7 versions of MSNP were disclosed to public, but
556 since version 8 the details have been kept a secret. MSNP does not use encryption so
557 even though MSNP's description was not published it was not hard to put it together using
558 packet sniffers.

559 The live Messenger is available for Windows, Mac OS X and recently was integrated into
560 Microsoft's game console Xbox 360. It features social network integration, offline messaging,
561 games and applications, voice and video calls and standard IM features. An interesting
562 feature is Multiple points of presence allowing user to be connected on two devices. Shared
563 whiteboard is available as an extra application and is not capable of multi-user session.

564 **Brosix**

565 An award winning application first released in 2006. Brosix features voice and video chat
566 and multi-user chat, basic IM functions and couple advanced functions. Brosix's whiteboard
567 is an Microsoft Paint like window shared among the participants and won Best IM Feature
568 2009 award from about.com. Next great feature allows users share screen, including mouse
569 and keyboard much like VNC. Finally Brosix implements co-browsing where users share a
570 browser window.

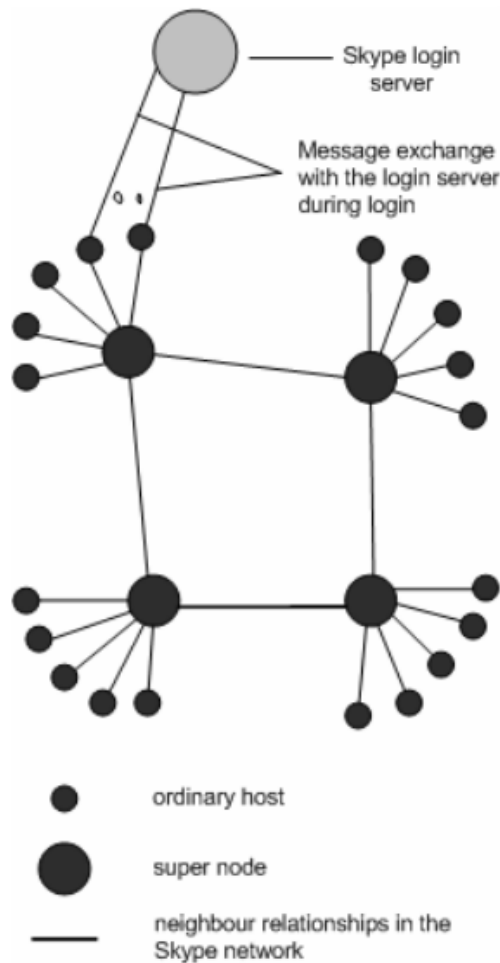


Figure 5.1: Telepathy components registered with user session D-Bus.[35]

571 Brosix is available for Windows, Max OS X and GNU Linux in commercial and per-
 572 sonal(free) version. There is little known about used technologies and it is almost impossible
 573 to reverse-engineer using packet sniffers as Brosix uses 256 bit AES encryption.

574 **Yahoo! Messenger**

575 Yahoo! Messenger is just like all of the above a closed program though some information
 576 has leaked[39]. Yahoo! Messenger protocol(YMSG) uses TCP on port 5050 or a different
 577 one if default is unavailable. To get to clients behind firewall HTTP is utilized. Video and
 578 voice supposedly use SIP and H.323.

579 Besides the standard set of IM functions Yahoo! Messenger can call PSTN, send SMS
 580 and handle voice conference. Whiteboard feature is called Scribbler and is plugin. Linux is
 581 missing in the list of supported platforms while Windows and Mac OS X are not.

582 **AOL Instant Messenger**

583 AIM is yet another IM with proprietary communication protocol. Though AIM is a bit of
 584 an exception for it supports two protocols. First is just for simple text messaging called

585 TOC and has been disclosed to public. Second protocol that supports all of the advanced
586 features is being kept a secret.

587 AOL's messenger is available for Windows and Mac OS X and features voice and video
588 calls as well as whiteboarding. Whiteboard is available as a plugin for AIM called IM
589 Whiteboard.

590 5.2 Multimedia streaming protocols

591 The following section lists the most popular protocols for VoIP work and explains how
592 it works. The first part is about signaling protocol SIP and the second about streaming
593 protocol RTP.

594 SIP

595 Session Initiation Protocol is standardized by IETF and was first defined in RFC 2543.
596 The latest definition is in RFC 3261. SIP is used in VoIP for negotiating the details of the
597 call. The parameters of the call are described using Session Description protocol described
598 in RFC 4566. Though designed for VoIP it can be used for establishing or terminating any
599 kind of session whether it is between two users or it is a multiuser session. Among the
600 features of SIP is also instant messaging, presence or any kind of event notification. SIP
601 uses primarily UDP on port 5060, but can also use TCP on the same port and 5061 for
602 TLS secured SIP. The syntax is similar to HTTP.

603 SIP clients are identified by URI which usually looks like this: `sip:username@hostname`,
604 to be concrete `sip:bob@biloxi.com`. First the client must register with a SIP proxy, which
605 in Bob's case is `biloxi.com`. Now let's say Alice wants to call Bob. To do that she needs
606 to know his URI. She sends an `INVITE` to her SIP proxy, by which it is forwarded to Bob's
607 proxy and finally delivered to Bob and his phone or computer start ringing. If he picks up
608 an appropriate numeric code is sent to Alice. The following example[14] shows the scenario
609 presented:

```
1 INVITE sip:bob@biloxi.com SIP/2.0
2 Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bK776asdhds
3 Max-Forwards: 70
4 To: Bob <sip:bob@biloxi.com>
5 From: Alice <sip:alice@atlanta.com>;tag=1928301774
6 Call-ID: a84b4c76e66710@pc33.atlanta.com
7 CSeq: 314159 INVITE
8 Contact: <sip:alice@pc33.atlanta.com>
9 Content-Type: application/sdp
10 Content-Length: 142
```

Example 5.1: SIP invite from Alice to Bob

610 Since SIP serves only to initiate the session it needs to cooperate with a protocol that
611 does transfer the data. That protocol is Real-time Transfer Protocol.

612 RTP

613 Real-time Transport Protocol is an IETF standard for transporting data that need to be
 614 delivered in real-time rather than reliably. Therefore UDP is used on the transport layer.
 615 First was RTP defined in RFC 1889 in 1996 and then later updated in RFC 3550 in 2003.
 616 It is primary protocol for streaming audio and video over the internet. It is used for VoIP
 617 for transporting the voice while SIP or anoteher signaling protocol negotiates parameters of
 618 the transport. More and more TV stations have been converting to the internet and they
 619 use RTP as means of distribution. RTP uses unicast as well as multicast when streaming
 620 to multiple subscribers.

621 RTP is used in conjunction with Real-time Transport Control Protocol. RTCP monitors
 622 QoS, statictics of the transfer and help with synchronisation when streaming to multiple
 623 destinations. The volume of RTPC traffic should be around 5% of the volume of the stream.

624 Unlike the circuit switched network, where the QoS is ensured by it's nature, the packet
 625 switched network does not have a way of ensuring short or not even constant delay or
 626 sufficient bandwidth (it is possible, but very rarely and in private network with proper
 627 SLA¹). RTP defines mechanisms for making the most of the packet switched network. It
 628 is important to note that in the internet packets of the same stream might take different
 629 path to their destination. That and network congestion are the reasons for varying delay
 630 commonly reffered to as jitter. RTP labels all the packets with sequence numbers. The
 631 implementation of RTP at the destination has a buffer to compensate for jitter and out-of-
 632 sequence delivery. If the packet arrives too late it is dropped. Dropping packets to some
 633 point might not even be noticable by the user.

634 RTP sends and receives data on even port numbers and the associated RTCP uses the
 635 next higher odd port number. An example on an RTP packet follows.

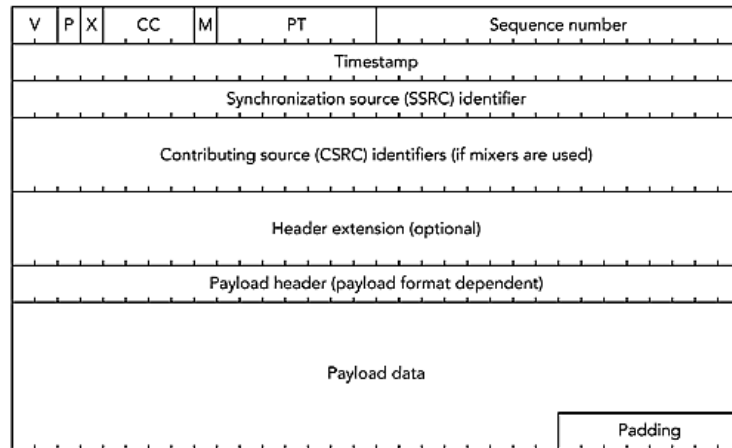


Figure 5.2: RTP packet.[28]

636 5.3 VoIP

637 Realization of Voice over IP deals with many problems. Factors effecting voice quality
 638 are choice of voice codec, echo cancelation, packetization, packet loss, delay, jitter (delay

¹Service Level Agreement

Voice codecs		
Codec name	Bitrate (kb/s)	MOS
G.711	64	4.1
G.723.1	6.3	3.9
G.726	32	3.85
G.729	8	3.92
Speex	8/16/32	unknown

Table 5.1: Voice codecs[?]

variation). First the voice needs to be digitized. That is done by a microphone, which consists of analog-to-digital converter and a mechanism for generating current based on sound in it's proximity. It is important to note that human ear can hear sounds of frequency between 16 to 16000Hz, but only fraction of this spectrum is used when speaking. It is 400 - 3500Hz to be concrete. By recording only a part of the spectrum lower sampling frequency might be used resulting in lower bitrate. The sampling frequency is determined based on Nyquist-Shannon sampling theorem [?]. The AD conversion in digital telephone networks is for example in digital telephone networks done by PCM², which takes 8000 samples - discrete values on the AD converter. Each sample has 8 bits so the resulting bitrate is 64kbit/s. PCM is for it's high bandwidth consumption and low level of captured information not suitable for VoIP [23]. There are alternatives specialised for representing speech, which has certain characteristics whereas PCM is used for any type of sound.

The next step is compressing the recorded voice. There are several compression algorithms. First difference is in length of voice chunks they compress. G.729 takes 10ms as opposed to G.723.1 that takes 30ms. The longer the chunks of voice are the less overhead (IP and Ethernet headers) is sent over the transport media. On the other hand the shorter the length the smaller impact on the conversation should the packet get lost. The compression algorithms used by the voice codecs should ideally lower the data size significantly, while causing short algorithmic delay and taking up insignificant amount of CPU time. Unfortunately there is no codec that would excel in all of the above. Good choice is thus a suitable compromise of the qualities. Since codecs are very hard to compare based on those qualities MOS³ as means of comparison attribute. The MOS is a subjective evaluation of quality of voice encoded/decoded by the particular codec on a scale 1-5. The table 5.3 shows the most popular voice codecs with bitrates and MOS.

Once the speech is digitized and encoded it needs to be sent over the network. Since VoIP is an interactive service the latency should be less than 150ms. If the delay is too long the users experience the two conversations effect. The types of delay that need to be taken into account at all times are processing delay, packetization delay serialization delay. The processing delay represents time needed for voice digitization and compression and is codec specific. Packetization delay occurs when the encoded voice is being loaded into packets and depends on how long chunks of speech the particular codec uses. Serialization means sending off the data through the network link and the delay is dependent on the available bandwidth on that link. For links with low bandwidth it is recommended to use codecs with low bitrates. G.729 for example uses voice activity detection and silence suppression.

²Pulse-Code Modulation

³Mean Opinion Score

673 The caller on the other side then is sent carefully generated comfort noise so that he does
674 not think the call went down. The comfort noise however has much lower bitrate requiring
675 lower bandwidth.

676 Last and largest portion of the total delay is time of network delivery. We need to
677 account for switching and routing delays, link transmission delays and also for jitter buffer
678 delays. The routing protocols work based on destination address which may, and often
679 does, result in packets of one conversation taking different routes to the caller. Different
680 routes mean different delays - jitter. Jitter buffer is used to accumulate arriving packets
681 and presenting them to the user in the correct order. Since RTP uses sequence numbers
682 it is an easy task to reorder packets. The problem is when packets are delivered too long
683 apart. In that case the missing packet is assumed to be lost for the resulting delay would
684 have a worse effect than the missing packet. Network administrators may influence the
685 network transmission delay using QoS mechanisms such as DiffServ⁴. DiffServ allows the
686 administrators to give the interactive data like voice and video a higher priority than traffic
687 like emails or HTTP. In order to do so the traffic must be classified and then marked. IPv4
688 provides means for marking traffic and thus allows using QoS throughout the internet.
689 The edge router through which the traffic enters the autonomous system either trusts the
690 marking of data from a neighboring network or (re)marks the traffic itself. Although most
691 of the networks utilize QoS there are still some that either don't or have incorrect setup.

692 Overcoming NAT and firewalls

693 First what is Network Address Translation (NAT) and what is it good for. The internet
694 is based on Internet Protocol version 4. The IPv4 was designed in 1981, when internet
695 was just a pack of machines mostly owned by universities and noone could imagine IPv4
696 address space⁵ could ever be used up. The enormous expansion of the internet cause that
697 the IPv4 addresses were all used up in February of 2011[12]. This is where IPv6 comes
698 in offering using 128 bit address and thus offering $3.4 \cdot 10^{38}$ cardinality of address space
699 allowing anyone to have public IP address. The transition from IPv4 to IPv6 is however
700 costly and complex due to the decentralized nature of the internet.

701 To connect new devices to the internet without having to wait for/transition to IPv6
702 NAT is used. It basically "hides" an entire network behind one IP address. This solves the
703 problem with insufficient amount of addresses, but creates another problem. The
704 devices behind NAT obtain a private address and can not be connected to directly from
705 the internet. The only device that is directly visible from the internet is one with public
706 IP address. Figure 5.3 shows the typical home network setup.

707 If a user on the home computer wants look at a website, the home computer sends the
708 request to the router, which performs translation of the home private address to a ISP's
709 network's private address. And the same happens on the edge of provider's network. The
710 routers performing NAT create a mapping of the source IP and source port to translated
711 address and port. The web server sends the reply back to the ISP, with it's address and
712 the translated port number. The edge router looks at mapping and translates back the
713 address and port number and sends it the home router. Same action is performed by the
714 home router and the packets are sent the home computer, which made the request.

715 Accessing the home computer from the outside can not be done for two reasons. First
716 private network is not present in the internet's routing tables. And second even if we knew

⁴Differentiated Services

⁵IPv4 address is a 32 bit number - 4,294,967,296 (2^{32}) addresses

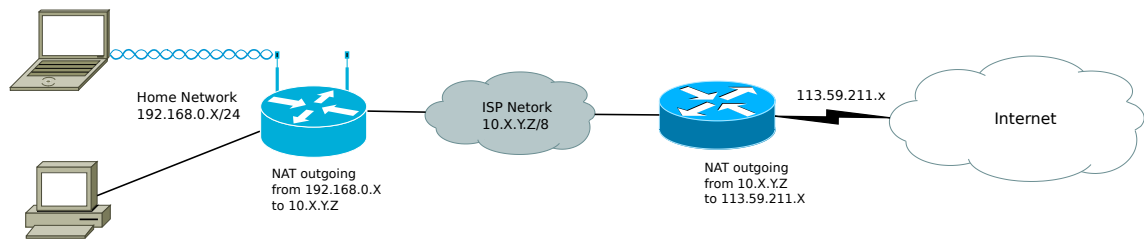


Figure 5.3: Typical connection to the internet

717 the address of the ISP's edge router, without the dynamic mappings the router would not
 718 know which device in it's network to forward the packets to. To summarize it is only
 719 possible to send data to a device in a private network if it initiates the communication.
 720 Moreover that is possible only if the peer has a public address. Two computers behind NAT
 721 can not communicate directly⁶.

722 There are several solutions for overcoming NAT like STUN or TURN⁷ (RFC 5766).
 723 STUN is a client-server protocol where the server has a public address and client is behind
 724 NAT. Client contacts the server with a request and the server then tries to acquire port and IP
 725 address, which translate to the client's address and a certain port number. STUN however
 726 is not always successful as it can't work with all types of NAT. Next there is TURN, which
 727 acts as a relay server and relays traffic from one client to another in case no other mechanism
 728 works. It is clear that this is a last resort. Maintaining such servers costs money for they have
 729 to be connected through links with high bandwidth and the traffic first goes in and then out.

730 Both of the above mentioned protocols are not universal enough for all kinds of address
 731 translation systems. Interactive Connectivity Establishment (ICE) by IETF defined in RFC
 732 5245 describes a methodology of taking the best of each NAT overcoming protocol to allow
 733 the clients with private addresses to communicate with each other either directly or through
 734 a third party relay server.

⁶It is possible through port forwarding, which is intentionally omitted

⁷Traversal Using Relay NAT

735 Chapter 6

736 Port to Telepathy

737 Telepathy communication framework described in detail in chapter 4 is a logical choice of
738 communication architecture if one wishes to create a powerful and easily maintainable IM
739 application. A proof that porting to telepathy is a step in the right direction is for example
740 a new client that is being developed by the KDE community and is supposed to become
741 a default KDE IM application. The first release is expected any time now. At that point
742 both main linux window managers KDE and Gnome (Empathy) will have their default IM
743 program based on telepathy. That fact alone promises great and lasting support of this
744 architecture. This chapter talks about the porting of Makneto from iris library to telepathy
745 and explains the design of the application and the reasoning behind it.

746 The first important change in the architecture of makneto is separation of user interface
747 from the communication backend. There were several reasons supporting this decision.
748 First is that since Telepathy is quite complicated and it takes time to get familiar with
749 it would be best if user interface used a simplified abstraction. And that is where the
750 communication backend comes in. Moreover the backend is the only piece of code that will
751 have to be adjusted should Telepathy undergo any design changes or adjustments. The
752 plan is to make the backend as portable as possible so it could run on most platforms
753 possible. The user interface will be created specifically for the targeted platform. With that
754 in mind it makes sense to separate backend and frontend. For instead of adjusting user
755 interfaces for all the platforms to the changes the backend will absorb it and frontends will
756 stay unchanged.

757 One of the aims of this work is to strip the current application of the iris library and
758 implement the backend based on Telepathy. I chose to use Qt4 bindings for Telepathy
759 simply because the rest of the application was implemented using Qt4 and it is overall very
760 convenient to use with high level of abstraction. The following figure 6.1 shows a diagram
761 of the new Makneto architecture.

762 6.1 Connection

763 As mentioned earlier Telepathy communicates over D-Bus, which means that if the appli-
764 cation has to have a D-Bus well-known name in order for Telepathy to be able to address
765 it. Namely the backend needs to implement `Client Handler` interface by subclassing
766 `AbstractClientHandler` and register with D-Bus. Along with wellknown name (in this
767 case Makneto) the application lets telepathy know what channels it is capable of handling.
768 Makneto backend is a Telepathy handler and at this point is able to handle text chat, mul-

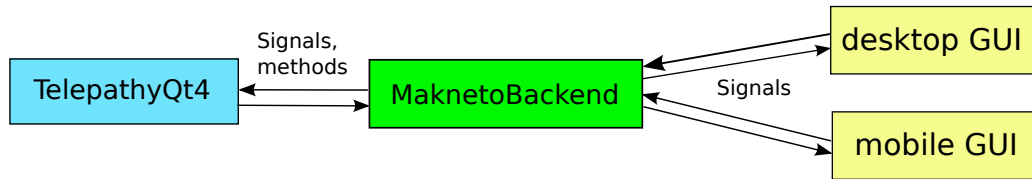


Figure 6.1: Simplified schema of Makneto components

769 tiuser chat, audio and file transfer. It is also prepared for video calls, which is functionality
 770 outside of the scope of this thesis. More on that in the following chapter.

771 Client handler needs to be initialised and that is where the **TelepathyInitializer**
 772 class comes in. It merely needs a name of the client and sets up features for contact,
 773 connection, account and channel factories. The factories ensure that if the managed object
 774 like connection for example is ready it supports all the features specified upon the creation of
 775 the connection factory. That way the factories hold the desired features and once the object
 776 emits **ready()** signal there is no need to query the available features. The factories are
 777 supplied to constructor of Account Manager. Account Manager contains **Account** objects
 778 representing accounts of the user. To be able to use Account Manager **becomeReady()**
 779 method must be called and then wait for **finished()** signal. Then the accounts may
 780 be accessed. The accounts we get from the Account Manager must also be made ready
 781 the same way as the manager. Once everything is ready to use the **Initializer** emits
 782 **finished()** signal.

783 At this point nothing stands in the way of having the accounts go online. This is handled
 784 through a context menu of a contact list, see section 6.3 on the frontend and by . When
 785 an account is brought online the Connection Manager creates a Connection with features
 786 specified in the **ConnectionFactory**. If there is another telepathy application already running
 787 with the account online the Connection for that account will be shared to prevent wasting
 788 resources.

789 Figure 6.2 shows classes of the Makneto backend that participate on the communication
 790 with Telepathy and through it with anyone else. **TelepathyClient** handles both incoming
 791 and outgoing Channels of types asked for upon registration with D-Bus. Once a Channel
 792 arrives the **TelepathyClient** checks if a session with the contact already exists and if it
 793 does the Channel is handed over to it. If not than new session is created and signal is
 794 emitted to inform the user interface. **Session** class encapsulates all of the communication
 795 of the user with the outside world. This is the main difference between iris and telepathy.
 796 Iris uses a single point of entry for all communication whereas with telepathy it is done
 797 via Channels. The session represents all kinds of interaction the user might have with the
 798 outside world through Makneto.

799 It is important to note that outgoing channels eventhough requested via **Session** invoke
 800 **handleChannels()** method of the **TelepathyClient**. There is a flag present with the
 801 channel indication whether it has been requested or is coming from the outside.

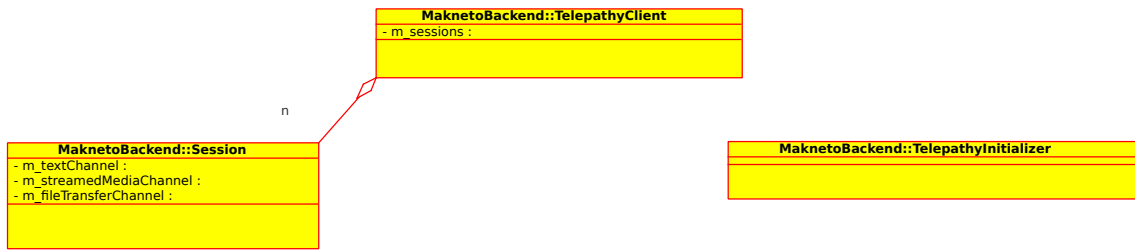


Figure 6.2: Makneto backend class diagram

6.2 Whiteboarding

Whiteboarding at this point is not supported by Gabble, connection manager for XMPP. Makneto itself the feature supports for it is Makneto's main feature. Until a plugin for whiteboard support is written for Gabble, the whiteboard messages will be tunneled through text messages. The main problem was sending SVG data to clients that would simply display the data to the user. Luckily XMPP offers resource parameter, which is used to check if the contact on the other side is using Makneto and thus can interpret the data correctly. Unfortunately this meant losing the ability to whiteboard with clients implementing SVGWB by Joonas Govenius, though the JEP has not yet been approved. Whiteboard classes belong to the backend and once the functionality is added to Gabble it will be a simple task to make use of it and restore the lost compatibility.

The main whiteboarding classes remained unchanged except for the SVG data delivery. The previous implementation based on the iris library supplied conveniently a XML element containing the whiteboard information. Since telepathy does all the XML parsing and interpretation for us, the only option was reconstruct the `wb` element from string.

6.3 Contact list

Makneto's contact list has been ported to telepathy and utilizes Qt model/view architecture. The model is a part of the backend and the view is a part of the user interface. The contact list model is based on a model from Telepathy-Qt4-Yell project[?]. The model is connected to all the possible signals stating the item's properties have changed. Once the item has changed, the model's slot receives the signal and emits `changed()` signal, letting the view know it should update information showing to the user.

One of the greatest benefits of telepathy - multi protocol support - is now present in Makneto. The new contact list model has two types of nodes (items) as shown in the class diagram in figure 6.3. Since it is a tree model I will start from the nodes. Nodes are individual accounts registered with telepathy. These accounts may used in other applications built on telepathy. Leaves are items representing contacts under each account. Although whiteboarding is supported only through XMPP, the other of the features like text chatting or file transfer are available over the other protocols. It must be of course supported by the client of the peer.

As you can see from the class diagram the `TelepathyClient` has an instance of the contact list model. This is useful for requesting sessions with contacts from the frontend

834 - **RosterWidget**. It simply specifies the contact or account by sending index of currently
 835 selected item of the model. Any other way would require implementing special Makneto
 836 classes for contact and account in order to keep the frontend independent on telepathy

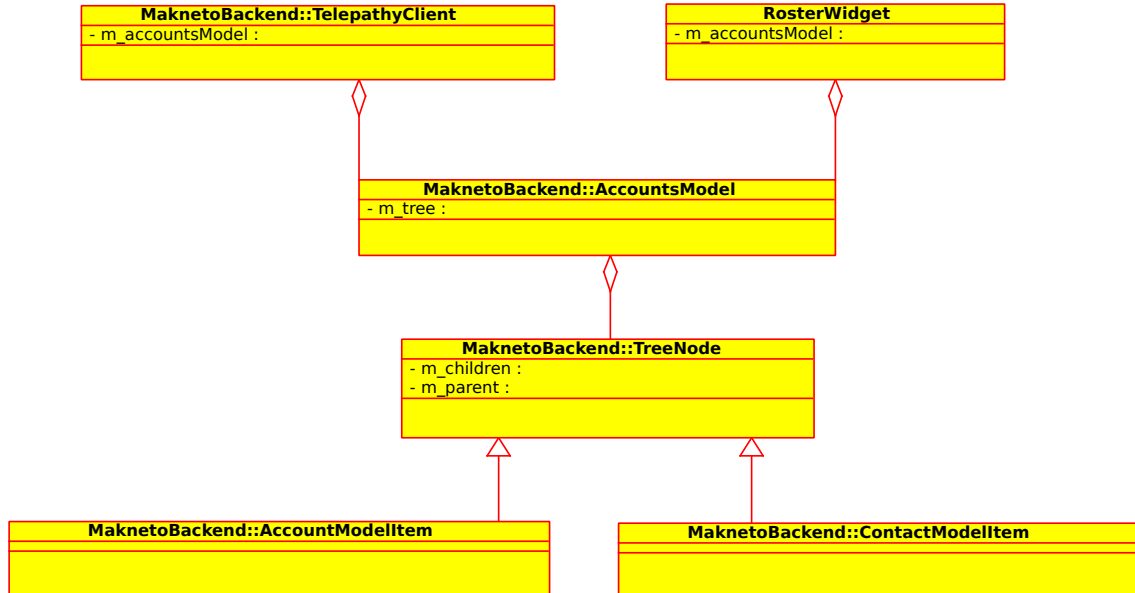


Figure 6.3: Makneto contact list class diagram

837 The user interface was kept almost the same eventhough both of the communication
 838 libraries have very different usage. Moreover this work does not aim at the frontend. One
 839 major change of the GUI is worth noting. The main class **Makneto** was an attribute in most
 840 of the classes and was passed along trthrough all those classes. I have utilized the singleton
 841 design pattern, for only one instance of the class exists. Makneto's **Instance()** method
 842 is static making it possible to get instance of the class anywhere within the application.
 843 **TelepathyClient** is also implemented as a singleton.

Chapter 7

VoIP implementation

The final chapter of this work contains details about implementation of VoIP into the application Makneto. Frameworks used for handling multimedia are listed with short description. And finally I talk about the implementation itself. Information in this chapter were found at [1, ?, 5]

Implementing an application handling multimedia means working with devices like sound card, camera, etc. There are several ways of doing so. First would be low-level access, which is platform and device specific. In GNU/Linux it requires using either Open Sound System (OSS) or Advanced Linux Sound Architecture (ALSA). Both of those offer abstraction of hardware beneath, but very low and work only on UNIX like systems and GNU/Linux respectively. Since I am aiming at portability they are not a good choice. Next is using a multimedia framework with higher level of abstraction supported on multiple platforms such as Gstreamer [1] or Phonon [?].

7.1 GStreamer

GStreamer is a multimedia framework for writing streaming multimedia applications. It is based on GLib 2.0 and offers API for writing plugins and thus extending it's functionality. Programmers use API for creating applications handling audio, video or both. Recently a stable bindings for Qt have been released making the development even more convenient. Have a look at figure 7.1 to get an idea of what the framework consists of.

Building an application using GStreamer is done by chaining the basic objects - elements. Each element performs a specific function like encoding audio stream, displaying video on the screen on reading from a network stream. It has data inputs and outputs called pads. Depending on the element it might have just an output pad called source¹. That might be for example one representing an audio file. If it contains just an input pad - sink, then it might be a sound card or screen. Elements containg two pads are filters, encoders, decoders, etc. and more than two would be muxers, demuxers, etc.

Chaining elements to perform certain function means connecting their sources and sinks. The chain is called pipeline and there is an example of one in figure 7.2. There are several elements linked together starting with the file-source. The file contains both audio and video and thus a demuxer is used with two sources, one for each stream. Next there is decoding and presenting the result to the user. With a large number of plugins supporting most of known codecs and devices this framework becomes a very powerful tool.

¹Source because it is a source from the next element's point of view

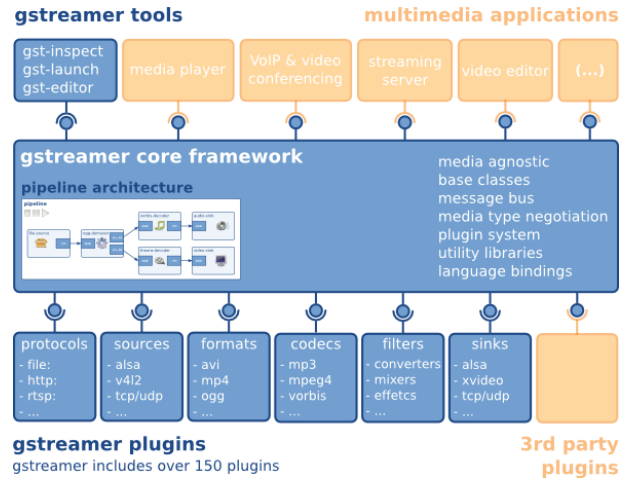


Figure 7.1: GStreamer architecture

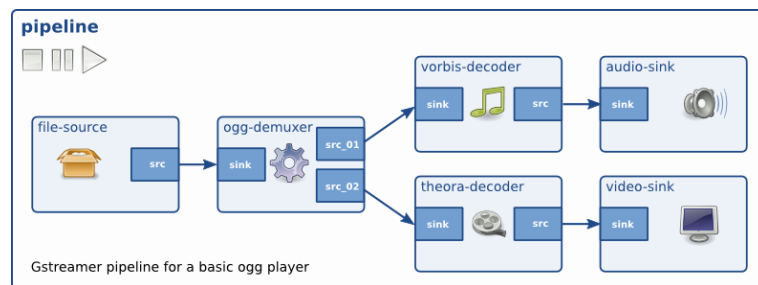


Figure 7.2: Example of a GStreamer pipeline

877 Pipeline is a special case of bin, which is a collection of elements. Since bin is also an
 878 element the programmer can create a pipeline consisting of several elements and performing
 879 certain function and use it as element in a larger and more complex pipeline. This design
 880 allows abstraction on a very high level. When linking elements together the media type
 881 that one produces and the other consumes must be compatible. That is determined through
 882 elements' capabilities in a process called "caps negotiation". GStreamer provides a way of
 883 automatic linking of elements called autoplugging based on capabilities.

884 When an element is created and linked there is no data flowing until it's state changed
 885 to play. The default state in which no resources are allocated is null. Next state is ready,
 886 when the element has all of it's resources allocated, but the stream is not opened yet. It is
 887 during this phase when the user finds out that the resource is unavailable or incompatible.
 888 From ready the elements go to paused by opening the stream a buffering it's contents. At
 889 this point the elements should be ready to switch state play instantly. Finally the play
 890 state is same as paused, but the data is actually flowing.

891 GStream is designed for high performance. Data between elements are copied only when
 892 absolutely necessary, otherwise pointer dereference is used. Moreover the stream processing
 893 is running in separate threads.

7.2 Phonon

7.3 Farsight

Farsight2 [5] is library implemented in C GLib based on GStreamer architecture. It eases transporting multimedia over the internet. It is meant to be used instant messenger application for audio and video calls. Farsight is used by Telepathy via telepathy-farsight library for which there are even Qt bindings. Empathy utilizes farsight for it's audio and video calls and so does for example Nokia Internet Tablets.

This library provides GStreamer elements for multimedia streaming protocols. The supported protocols are UDP, RTP and MSN. One of the farsight's features is NAT and firewall overcoming ability. It implements ICE for reaching into private networks and HTTP tunneling for breaching restrictive filters. Moreover farsight supports dynamic codec negotiation and switching. It also makes it possible to bridge incompatible protocols and conference with for example a GTalk and Yahoo user.

Using farsight is consistent with using GStreamer, because it implements GInterface. **FsConference** represents a conference that can contain one or more sessions. **FsSession** corresponds to one RTP session and offers a sink and source pad. The sink pad should be connected to either microphone or video camera, generally an audio or video input device. The source pad would be linked to audio or video output device. One session has one media type and one sink pad, but can have more source pads in case of more participants². For audio calls with video there will be two **FsSession** objects each for one type of multimedia. Each session has one or more transmitters. **FsTransmitter** abstracts the network layer. One transmitter serves for one transport method (UDP with ICE, UDP, HTTP tunneling, etc.). Finally there is **FsStream** object that represents a connection to one participant. The streams correspond with source pads in the session.

If used with telepathy, farsight utilizes **StreamedMediaChannel** for signalling. The protocol used for signalling depends on the used connection manager. In case of Gabble Jingle is used and SIP in case of Sofiasip. Farsight first creates a session and it's type is established. If there are more than just media types then sessions within this conference are created. Streams are built next. If the client allows user to specify his or her preferred codecs it should be supplied to farsight when it emits **stream-get-codec-config** signal. Once farsight has the preferences it creates the stream. In case of VoIP the streams are bidirectional. At this point farsight negotiates the codecs. It goes through gstreamer plugins and looks for encoding and decoding elements. List of available codecs is sent to the remote machine. With codec negotiation finished the only thing left to is arranging a peer to peer transfer of the data through the stream. This is accomplished by using ICE. First remote candidates followed by local candidates are acquired. Candidate is an IP address and port where the participant might be able to receive data. Local and remote candidates are put in candidate pairs. Thanks to ICE farsight is able to determine whether the peers are on network and transfer data locally. When a working and suitable candidate pair is established the streaming begins. Farsight lets the application know by emitting signal **src-pad-added** of the stream object. That means the src pad of the stream is ready to supply data to a prepared GStreamer pipeline. Until this moment all the data transferred went through the signalling protocol.

²one source for each participant

7.4 VoIP in Makneto

The implementation of VoIP in Makneto is based on KDE Telepathy Call UI [?] by George Kiagiadakis. It uses telepathy farsight qt4 bindings to create a `TfChannel`. The application communicates with the farsight channel via GLib signals. Events the application checks for are creation of a session followed by creation of a stream, channel closure and codec configuration request. Once stream is created more signals need to be connected. These include stream source pad ready, stream closed and resource requested. By emitting `resource-requested` signal farsight lets the application know that the GStreamer pipelines for processing incoming and outgoing data should be created and put into playing state.

The GStreamer pipeline used for audio data transmission is in figure 7.3. The chain starts with `AudioInput` representing either Pulse, ALSA or OSS input devices depending on what is available. The volume element is used for adjusting volume of the raw audio input. There is a volume widget available within the backend. Volume is linked to an audio converting element, which converts the raw audio between different formats. The following element handles audio resampling. And the final element represents the RTP stream encapsulated by farsight.

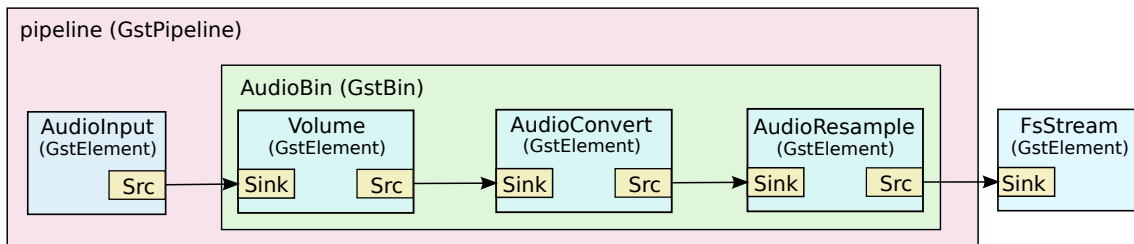


Figure 7.3: GStreamer pipeline used for VoIP in Makneto

The receiving pipeline is very similar. The `AudioBin` is used unchanged. The data however flow in the opposite direction. Moreover instead of audio input an audio output must be use. The process of acquiring it is the same.

956 Chapter 8

957 Conclusion

958 The world of today can be described as one enormous network, to which everyone is or
959 soon will be connected. One of the Internet's greatest features is bringing people together.
960 People who hundred of years ago would have to travel sometimes months to see each other
961 or wait for reply get an opinion on an idea from a colleague. Today it is just a few clicks
962 away.

963 Numerous application exist featuring whiteboard and VoIP and some of them even video
964 conferencing. All of the programs mentioned in chapter 5 unfortunately used proprietary
965 protocols and do not provide any information about it at all.

966 There are several ways to implement a VoIP solution. RTP with SIP seems to be a
967 feasible solution as it has been proven to work by numerous applications. Though after much
968 thinking it seems the technologie to with will be XMPP Jingle. It has been implemented
969 by Collabora and is now part of Telepathy Gabble - XMPP connection manager. Though
970 still under development the presented result look promising.

971 There are several tasks to be done. First is porting Makneto to Telepathy. Then a
972 thourough examination of voice encoding must be done to decide what codec shall be used.
973 And finally implementing VoIP support to Makneto.

974

Bibliography

- [1] Gstreamer 0.10 core reference manual, 2008. [online; accessed 10 March 2011].
- [2] Terry Brock. Skype outage today. [online; accessed 27 Dec 2010].
- [3] Brosix. Brosix - seruce corporate instatnt messaging for companies. [online; accessed 30 Dec 2010].
- [4] Collabora. D-bus. [online; accessed 19 Nov 2010].
- [5] Collabora. Farsight2 wiki. [online; accessed 4 April 2011].
- [6] Collabora. Telepathy. [online; accessed 12 Nov 2010].
- [7] Collabora. Telepathy wiki. [online; accessed 18 Nov 2010].
- [8] XMPP Standard Foundation. Xmpp standards foundation. [online; accessed 11 Dec 2010].
- [9] Bur Goode. Voice over internet protocol, 2002.
- [10] Google. Google talk. [online; accessed 12 January 2011].
- [11] Joonas Govenius. Svg whiteboarding, 2006-06-19. [online; accessed 11 December 2010].
- [12] IANA. Iana address space report. [online; accessed 23 February 2011].
- [13] ICQ. Icq. [online; accessed 29 Dec 2010].
- [14] J. Rosenberg et al. Sip: Session initialization protocol, 2002. [online; accessed 7 January 2011].
- [15] j. Rosenberg et. al. Stun - simple traversal of user datagram protocol (udp) through network address translators (nats), 2003. [online; accessed 6 January 2011].
- [16] J. Rosenberg et. al. Session traversal utilities for nat (stun), 2008. [online; accessed 6 January 2011].
- [17] Joe Hildebrand, Peter Millard, Ryan Eatmon, Peter Saint-Andre. Xep-0045: Multi-user chat. [online; accessed 21 January 2011].
- [18] Joe Hildebrand, Peter Saint-Andre, Remko Tronçon, Jacek Konieczny. Xep-0115: Capabilities advertisement. [online; accessed 23 January 2011].

- 1002 [19] KaZaA. Download music - music downloads and mp3 downloads from kazaa.com.
1003 [online; accessed 26 Dec 2010].
- 1004 [20] KDE. Kde. [online; accessed 3 Dec 2010].
- 1005 [21] Kolektiv autorů. Scalable vector graphics(svg). [online; accessed 18 Dec 2010].
- 1006 [22] Kolektiv autorů. Svg tiny 1.2 specification. [online; accessed 18 Dec 2010].
- 1007 [23] A. M. Kondoz. *Digital Speech*. John Wiley & Sons Inc., 2004.
- 1008 [24] Microsoft. Windows live messenger 2011. [online; accessed 4 Jan 2011].
- 1009 [25] Daniel Molkentin. *The Book of Qt 4*. O'Reilly, July 2007.
- 1010 [26] Nokia. Qt - a cross-platform application and ui framework. [online; accessed 15 Nov
1011 2010].
- 1012 [27] America OnLine. Aol instant messenger. [online; accessed 2 Jan 2011].
- 1013 [28] Colin Perkins. *RTP: Audio and Video for the Internet*. Addison-Wesley, June 2008.
- 1014 [29] The GNOME Project. Empathy - gnome live! [online; accessed 01 Dec 2010].
- 1015 [30] Psi. Iris library. [online; accessed 17 Nov 2010].
- 1016 [31] Peter Saint-Andre. Xep-0045: Multi-user chat. [online; accessed 21 January 2011].
- 1017 [32] Peter Saint-Andre. XMPP Core. [online; accessed 27 December 2010].
- 1018 [33] Peter Saint-Andre. XMPP IM. [online; accessed 27 December 2010].
- 1019 [34] Peter Saint-Andre, Kevin Smith, and Remko Troncon. *XMPP: The Definitive Guide*.
1020 O'Reilly, April 2009.
- 1021 [35] Henning Schulzrinne Salman A. Baset. An analysis of the skype peer-to-peer internet
1022 telephony protocol, 2004. [online, accessed 26 Dec 2010].
- 1023 [36] Scott Ludwig, Joe Beda, Peter Saint-Andre, Robert McQueen, Sean Egan, Joe
1024 Hildebrand. Xep-0166: Jingle. [online; accessed 25 January 2011].
- 1025 [37] Skype.com. Skype. [online; accessed 26 Dec 2010].
- 1026 [38] Wikipedia. Wikipedia, skype protocol. [online; accessed 26 Dec 2010].
- 1027 [39] Wikipedia. Wikipedia, yahoo! messenger. [online; accessed 27 Dec 2010].
- 1028 [40] Yahoo! Yahoo! messenger. [online; accessed 30 Dec 2010].
- 1029 [41] Jaroslav řezník. Sdílená tabule. Master thesis, 2008.