Department of Communications and Networking

# Business and technical aspects of mobile peer-to-peer social networks

**Marcin Matuszewski**

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
**DOCTORAL DISSERTATIONS** 136/2013

# Business and technical aspects of mobile peer-to-peer social networks

**Marcin Matuszewski**

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Electrical Engineering for public examination and debate in Auditorium S1 at Aalto University, School of Electrical Engineering (Espoo, Finland) on the 4th of October, 2013, at 12.

**Aalto University**
**School of Electrical Engineering**
**Department of Communications and Networking**

**Supervising professor**
Prof. Raimo Kantola

**Preliminary examiners**
Prof. Schahram Dustdar, The Vienna University of Technology, Austria
Prof. Mika Ylianttila, Oulu University, Finland

**Opponent**
Prof. Hassan Charaf, Budapest University of Technology and Economics, Hungary

Aalto University publication series
**DOCTORAL DISSERTATIONS** 136/2013



ISBN 978-952-60-5314-1 (printed)
ISBN 978-952-60-5315-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5315-8

http://lib.tkk.fi/Diss/

Unigrafia Oy
Helsinki 2013

Finland

441 697
Printed matter

**Aalto University**

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

# Abstract

**Author**
Marcin Matuszewski

**Name of the doctoral dissertation**
Business and technical aspects of mobile peer-to-peer social networks

**Publisher** School of Electrical Engineering

**Unit** Department of Communications and Networking

**Series** Aalto University publication series DOCTORAL DISSERTATIONS 136/2013

**Field of research** Networking Business/Technology

**Manuscript submitted** 26 June 2013 **Date of the defence** 4 October 2013

**Permission to publish granted (date)** 1 July 2013 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

**Abstract**

Introduction of mobile high-capacity networks, as well as the widespread penetration of powerful mobile handsets provide a good basis for the development of peer-to-peer applications in the mobile environment. However, it is not certain if the P2P services that were well adopted in the fixed In-ternet can also be successfully launched in the mobile environment.

This article dissertation presents research results of mobile community service provisioning using P2P technology. The study was carried out on two levels: technical and business. On the business level the dissertation discusses the mobile P2P service provisioning ecosystem including analysis of stakeholder needs as well as potential scenarios for mobile P2P services. The dissertation presents the results of a user survey and a literature study. The presented material reveals that there is room for P2P services in the mobile environment, however user requirements are different than in the fixed environment. The dissertation also presents a scenario planning methodology that proposes the Schoemaker's variant of scenario planning as a suitable method for evaluating emerging mobile services. Consistent and coherent learning scenarios that were developed using the proposed methodology are also presented.

On the technical level, the dissertation presents P2P system architectures, protocols, and algorithms that enable the provision of community services in the mobile environment. In particular, the dissertation describes the world first resource sharing system that works on top of SIP networks. The system enables mobile phone users to share resources with each other and does not require any changes to the basic SIP infrastructure. A Social DHT architecture that allows for efficient formation of mobile communities is also presented. The dissertation shows how the P2P infrastructure can become a feasible cost efficient replacement for a mobile infrastructure by presenting a Distributed IP Multimedia Subsystem as well as a pioneering new mobile P2PSIP system for real-time communication services. The dissertation discusses an implementation of a P2P system that allows mobile phone users to search for knowledge in their trusted social communities overcoming the problems identified in the business study of the dissertation. The results of measurements and trials conducted show the technical feasibility of mobile community service provisioning using P2P technology.

**Keywords** Social networks, end-user research, scenario planning, mobile services, peer-to-peer, P2PSIP, content sharing, voip, IMS

**ISBN (printed)** 978-952-60-5314-1 **ISBN (pdf)** 978-952-60-5315-8

**ISSN-L** 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

**Location of publisher** Espoo **Location of printing** Helsinki **Year** 2013

**Pages** 185 **urn** http://urn.fi/URN:ISBN:978-952-60-5315-8

# Acknowledgements

I would especially like to thank Professor Raimo Kantola for his support, guidance, and discussions. I would also like to thank Professor Schahram Dustdar and Professor Mika Ylianttila for their pre-examination comments. My gratitude goes to Professor Hassan Charaf for devoting his time and expertise to act as my opponent. While working at Nokia and Aalto University (previously Helsinki University of Technology) for several years, I have had many interesting and inspiring discussions with numerous people. Special thanks to Miguel Garcia, Henry Sinnreich, Nicklas Beijar whom I have been privileged to work with in close cooperation and for their excellent and inspiring discussions. I wish to thank my colleagues and coauthors for their valuable participation in this work (list in alphabetical order): Sergey Balandin, Salman Baset, Heikki Hämmäinen, Mikko Heikkinen, Tuomo Soinio (previously Tuomo Hyyryläinen), Markus Isomäki, Timothy Jasionowski, Janne Kari, Imre Kelenyi, Esko Kokkonen, Juuso Lehtinen, Jukka Nurminen, Olli Rantapuska, Henning Schulzrinne, Hannes Tschofenig. I am grateful for the opportunity to work with the skillful and creative people at the IETF, IEEE and from the Hong Kong University of Science and Technology, Ericsson, Nokia Siemens Networks, T-Mobile, Microsoft, Oulu University of Technology. I would also like to express my gratitude for the thesis work funding from Nokia, TeliaSonera Foundation. Many other persons deserve thanks – you know who you are. Last but not least, I would like to thank my parents and my wife Karolina for their support.

# List of included publications

[P1] M. Matuszewski, N. Beijar, J. Lehtinen, and T. Hyyrylainen, “Understanding Attitudes Towards Mobile Peer-to-Peer Content Sharing Services”, IEEE International Conference on Portable Information Devices, Orlando, FL, USA, March 2007.

[P2] M. Matuszewski, N. Beijar, J. Lehtinen, and T. Soinio (previously T. Hyyryläinen), “Content sharing in mobile P2P networks: myth or reality?”, International Journal of Mobile Network Design and Innovation, Vol. 1, No. 3/4, pp. 197–207, January 2006.

[P3] M. Matuszewski, M. A. García-Martín, N. Beijar, and J. Lehtinen, “Resource Sharing and Discovery on Top of IMS”, IEEE Consumer Communications and Networking Conference: "Seamless Consumer Connectivity", Las Vegas, NV, USA, January 2007.

[P4] M. Matuszewski and M. A. García-Martín, “Social Distributed Hash Table (SDHT). Share and collaborate in your social network.”, IEEE Wireless Communications and Networking Conference, Hong Kong, PRC, March 2007.

[P5] M. Matuszewski and S. Balandin, “Peer-to-Peer Knowledge Sharing in the Mobile Environment. Leveraging the strengths of social networks.”, Fifth International Conference on Creating, Connecting and Collaborating through Computing, Kyoto, Japan, January 2007.

[P6] M. Matuszewski, “Distributed IMS”, IP Multimedia Subsystem (IMS) Handbook, Chapter 11, CRC Press 2008, ISBN 1420064599

[P7] M. Matuszewski and E. Kokkonen, “Mobile P2PSIP - Peer-to-Peer SIP Communication in Mobile Communities”, IEEE Consumer Communications and Networking Conference, Las Vegas, NV, USA, January 2008.

[P8] M. Heikkinen, M. Matuszewski and H. Hämmäinen, “Scenario Planning for Emerging Mobile Services Decision Making: Mobile P2PSIP case study”, International Journal of Information and Decision Sciences, Vol. 1, No. 1, pp.26-43, January 2008.

# Authors contribution

[P1] The author started and was leading the research presented in the paper, being responsible for designing and conducting the survey, analyzing results, as well as editing the paper. The survey was designed and implemented by Nicklas Beijar and the author. Juuso Lehtinen and Tuomo Hyyrylainen helped us in the analysis of the survey results. The author wrote Sections 1, 2, 3, 5 and co-authored Section 7 with Nicklas Beijar.

[P2] The author started and was leading the research presented in the paper. The author was responsible for the system and software architecture design, signaling, analysis of results, as well as editorial tasks. The author wrote Sections 1, 2, and co-authored Section 3 with Nicklas Beijar and Tuomo Hyyrylainen, Section 5 with Juuso Lehtinen and Section 6 with Nicklas Beijar.

[P3] The author started the research and provided the first version of the architecture and signaling that was modified together with Miguel Garcia to achieve the final result presented in the paper. Writing of the paper was mainly done by the author and Miguel Garcia (equal share). The author edited the document. Juuso Lehtinen and Nicklas Beijar contributed measurement results presented in Section 5.

[P4] The author started the research and developed SDHT architecture and content sharing service presented in the paper. Miguel Garcia contributed knowledge about implementation of the system in IMS. The author wrote and edited most of the document.

[P5] The author was responsible for user studies and incentive models as well as for development of the system architecture. The author was also responsible for developing mobile software that enabled mobile phone users to search knowledge in their social network. The author edited the paper and wrote most of the text.

[P6] The book chapter is an extended version of the paper[1] that the author wrote with Miguel Garcia. The author brought knowledge about P2P systems and Miguel Garcia contributed knowledge about IMS. Together we designed the Distributed IMS system. We equally shared writing of the paper. The author edited the paper.

---

[1] M. Matuszewski and Miguel A. Garcia Martin, Distributed IP Multimedia Subsystem, IEEE International Symposium on a In World of Wireless, Mobile and Multimedia Networks, 2007 (WoWMoM 2007), Helsinki, Finland

[P7] The author started and was leading the research presented in the paper. The author wrote and edited the paper. Esko Kokkonen implemented the system presented in the paper. The paper received the best paper award among 6 workshops at the IEEE CCNC 2008 conference.

[P8] The author contributed knowledge about P2PSIP systems and the P2PSIP business ecosystem. Mikko Heikkinen contributed knowledge about scenario planning techniques. Together with Mikko Heikkinen, we analyzed the mobile P2PSIP service using Schoemaker's model as well as developed and evaluated the potential mobile P2PSIP scenarios. Most of writing and edition was done by Mikko Heikkinen. The author wrote Sections 3, 4.6, 4.8. Heikki Hämmäinen offered advises on general techno-economic matters and guided the research work.

# Table of contents

# List of figures



# List of tables

# Abbreviations

| | |
|---|---|
| 3G | Third Generation (Mobile Communication System) |
| 3GPP | 3rd Generation Partnership Project |
| AOR | Address-of-Record |
| ARPU | Average Revenue Per User |
| AS | Application Server |
| CAPEX | Capital Expenditures |
| CPU | Central Processing Unit |
| DHT | Distributed Hash Table |
| DIMS | Distributed IP multimedia Subsystem |
| DTLS | Datagram Transport Layer Security |
| HLR | Home Location Register |
| HSS | Home Subscriber Server |
| I-CSCF | Interrogating-Call/Session Control Function |
| IETF | Internet Engineering Task Force |
| IMS | IP Multimedia Subsystem |
| IP-CAN | IP Connectivity Access Network |
| IPR | Intellectual Property Right |
| IPsec | Internet Protocol Security |
| KM | Knowledge Management |
| MVNO | Mobile Virtual Network Operators |
| NAT | Network Address Translator |
| OPEX | Operating Expenditures |
| P2P | Peer-to-Peer |
| P2PP | P2P Protocol |
| P2PSIP | Peer-to-Peer Session Initiation Protocol |
| P-CSCF | Proxy-Call/Session Control Function |
| PDA | Personal Digital Assistant |
| RPC | Remote Procedure Call |
| SAS | Short Authentication Strings |
| S-CSCF | Serving-Call/Session Control Function |
| SDHT | Social Distributed Hash Table |
| SigComp | Signaling Compression |
| SIP | Session Initiation Protocol |
| SIPS URI | Secure SIP URI |
| SMS | Short Message Service |
| SRTP | Secure Real-time Transport Protocol |

| | |
|---|---|
| STUN | Simple Traversal of UDP (User Datagram Protocol) through NATs |
| TLS | Transport Layer Security |
| TURN | Traversal Using Relay NAT |
| UA | User Agent |
| URI | Uniform Resource Identifier |
| VoIP | Voice over Internet Protocol |
| XML | Extensible Markup Language |

# 1. Introduction

Peer-to-Peer overlay networks have gained tremendous public attention during recent years. Peer-to-peer (P2P) networks have enabled Internet users to freely share resources such as music files, digital movies and pictures, to share knowledge as well as communicate with other people with the minimum assistance of centralized servers. Nodes in peer-to-peer networks called peers act both as servers and clients in contrast to the client-server model where the relatively low number of servers provide a service to a large number of clients. In the P2P networks any peer may provide services to other peers utilizing its own resource such as bandwidth, storage space, computing resources, and power. P2P technologies have gained popularity because of their high scalability, low OPEX and CAPEX, self-configuration, and identity protection features.

The peer-to-peer revolution was launched in October 1999 with the introduction of the Napster content sharing service. Its easy to use interface, that enabled access to unbounded free music resources, induced its widespread popularity and an extremely rapid growth.

The increasing number of Napster users provoked a concern about the future of the music industry. Napster became a threat to labels and the Recording Industry Association of America (RIAA). A couple of months later RIAA sued Napster for copyright infringement. The industry argued that Napster was contributing to massive copyright violations, as Napster users shared tens of thousands of songs every day. On February 9th 2001, a Circuit Court decided that Napster violates copyright law and ordered Napster to install filters and blocks to prevent transfer of copyrighted material. The Napster trial, however, has not finished this music industry headache. New services such as Gnutella, Kazaa, and BitTorrent whose decentralized architecture makes it more difficult to shut them down, have been launched. Moreover, they allowed the sharing of content with other Internet users without revealing user identities.

Napster has not only changed the conditions under which the copyright law is applied, but what is more important, it has altered the landscape of music retailing. New possible business models that have emerged together with free file sharing services have caused big changes in the economics of music distribution. It was only the beginning of the Internet revolution in content distribution.

P2P technology also entered other market segments such as real-time communication, Internet TV and video distribution bringing fundamental changes to both fixed and mobile markets. SKYPE is an example of P2P real-time communication service developed by the inventors of Kazaa, a content sharing service. Since August 2003, more than 300 million copies of Skype's VoIP software have been downloaded.

Skype allows Internet users to freely communicate over the Internet. Skype provides similar services to traditional voice service operators but it does not charge for calls inside its network even if they are international calls. Skype is able to offer free calls because the cost of operating the service (OPEX) and customer acquisition is smaller than in telecom networks that have to maintain various servers and assist users in software/hardware configuration. P2P technology allowed Skype to distribute voice servers among end-user devices. The scalable P2P architecture that provides self-configuration and automatic reaction to failures allowed Skype to easily increase the size of their network to millions of users.

P2P applications, on the other hand, significantly increased traffic in the operator networks forcing operators to invest in the upgrade of their infrastructure [1]. Some of the operators used other means to deal with increasing P2P traffic such as throttling, traffic shaping, policing, or started charging for excessive traffic [2][3].

## 1.1 P2P in the mobile domain

Such a spectacular success of the P2P systems would not have been possible without technological changes in the Information and Communication Technologies (ICT) industry. Faster residential Internet connections, more powerful desktops, and cheaper storage have been the main drivers stimulating the P2P growth.

A similar technological change can also be observed in the mobile industry. The introduction of high-capacity Third Generation (3G) and Wireless Local Area Network (WLAN) networks, more powerful handsets with larger memories and multimedia extensions provide a good basis for the development of novel Internet services in the mobile environment.

Most of the current mobile content distribution and real-time communication platforms are implemented using the client-server architecture. The importance of mobile person-to-person communication and popularity of P2P content sharing services in the Internet raises a question about applicability of scalable peer-to-peer architectures in the mobile environment.

In the legacy content distribution platforms, a user searches for content by navigating through an operator's portal using a standard web browser or by using a special application such as Apple iTunes. When the user locates the desired content, he may download the content directly from a server maintained by the service operator. Alternatively the content can be delivered to the user by the MMS messaging service. The legacy content distribution platforms were mainly designed to distribute professional content.

Another type of content distribution platform are mobile blogs. In contrast to the legacy content distribution platforms, mobile blogs allow mobile phone users to post photographs, audio, text and video to their blogs or albums that are stored in centralized servers. The content stored in a user's blog is immediately viewable using a standard web browser by other users. The blogs are an incremental innovation. They allow for sharing user created content but content distribution utilizes the standard client-server architecture.

Marketing platforms are yet another content distribution platform. A marketing platform is a mobile content sharing application that enables users to find, buy, and recommend mobile content to other mobile users. The system tracks what mobile content consumers are purchasing from centralized content servers and encourages them to recommend the content to their group mates by sending messages (SMS, MMS or instant messaging) to their buddies. The marketing platforms are similar to the legacy content distribution platforms. They use client-server architecture for content distribution. The main difference is that they stimulate demand for a particular piece of content stored in the centralized servers using word-of-mouth marketing.

Similarly, real time mobile communication systems such as the IP Multimedia Subsystem (IMS) or conventional Session Initiation Protocol (SIP) systems use the client-server model. They store user's registrations or presence information and handle communication logic in centralized servers.

The presented platforms utilize a relatively small number of powerful servers to provide services to users. In contrast to the described platforms, in Peer-to-Peer networks content as well as user registrations and presence data are distributed among P2P nodes that may be run on end user devices such as mobile phones.

Because of the distributed nature of the P2P technologies, bringing the P2P communication model to mobile networks is not straightforward. Mobile devices have different constraints than their fixed counterparts. Let us briefly discuss the constraints:

*Battery consumption*

In all kinds of battery-powered devices battery consumption is one of the main problems. In the P2P overlay networks, peers perform various functions necessary for maintaining the P2P overlay such as routing incoming messages, maintaining connections with other P2P nodes, updating the network topology, and storing data on behalf of other nodes. All of these operations increase battery consumption. We can distinguish two states in the operation of P2P software: *idle* and *active* state. The idle state represents a state when peer software runs in a battery-powered device but a user does not use any services provided by the P2P overlay network. As it will be discussed in Chapter 4 the battery of a mobile phone running P2P software may be drained in just 5 hours in the idle state. The active state includes all operations of the idle state and additionally operations related to the service usage by the user of the device. The battery consumption is even higher in the active state. However, the active state is limited in time. The time when a device is in the active state depends on service usage patterns. In contrast to P2P content sharing networks, VoIP sessions are rather short, typically lasting as long as phone calls. On the other hand, an average user is most likely to initiate more VoIP sessions than content downloads everyday.

*Bandwidth consumption*

The P2P overlay protocols and P2P overlay network architectures must be optimized: first to reduce potential charges incurred by transmitting data over the air interface and to limit the overall traffic in the mobile networks especially in the access networks.

Being connected to the P2P overlay nodes and performing all of the overlay functions requires transmission of data over the expensive air interface. Moreover, usage of P2P services requires transmission of additional data over the air interface. P2P software may transmit and receive even 1MB of data every hour of operation being in the idle state. This is unacceptable for many users.

*NAT traversal*

In contrast to servers operated by mobile operators or Internet service providers, typically mobile phones are behind Network Address Translators (NATs) that make them difficult to access by other network nodes. There are mechanisms to traverse NATs such as the one presented in [P7], but these mechanisms additionally increase traffic over the air interface and consume battery power. A NAT traversal mechanism may require a device to send keep alive messages to keep the binding in NATs and establish and maintain a UDP/TCP or Datagram Transport Layer Security (DTLS)/ Transport Layer Security (TLS) connections towards other peers or a relay server. Most peers are required to connect to more than one node. Typically the more connections a mobile device needs to handle the faster the battery is consumed and the more traffic is sent over the air interface [4][5].

*Intermittent connectivity*

Mobile devices may change their IP address by roaming to another access point or may lose wireless network coverage. When a peer that stores a particular resource, such as user registration information or content, loses connectivity with the overlay network the resource becomes unavailable for other P2P overlay network nodes. On the other hand, as it is presented in [P1] users of content sharing applications do not want to keep their P2P application running all the time. Many respondents indicated that they switch off their P2P applications after downloading the content. Moreover, because of the intermittent connectivity of a mobile device, a mobile P2P overlay network must allow P2P nodes to join and leave the overlay easily with minimum overhead signaling for the whole network.

*Limited resources*

In addition, some small capacity devices such as low-end mobile phones may have limited CPU and/or memory capacity that does not allow them to run sophisticated P2P algorithms. A slow CPU may not allow a device to execute P2P processes and handle transmitted messages. The memory can be too small to store resources on behalf of other network nodes. Some mobile devices may also have a narrow-band interface that is too slow for P2P services.

*Usability issues*

Mobile devices have typically limited screen sizes and input capabilities, such as a small keypad, that create different challenges in designing a user interface than in full size personal computers. For example, displaying the same amount of information on a small screen than on a 15-inch display requires a user to scroll extensively to find a piece of information.

## 1.2 Contribution to knowledge

Because of the constraints just presented, it was not certain if the P2P services that were well adopted in the fixed Internet could also be successfully launched in the mobile environment. Is there any real customer need to use such services? Is the idea feasible from both business and technical perspectives?

In this article dissertation research results and practical innovations are presented on mobile community services that utilize P2P technologies. My research goal was to design practical innovations. To be practical, innovations need to go thru scientific research including theory development, user studies and verification prototyping. After this and in addition to this, also, a lot of engineering type of work is needed, before new innovations are adopted. In this dissertation research results are discussed on two levels: technical as well as business.

The article dissertation consists of an overview and 8 included publications. The overview links knowledge presented in 8 included publications with a larger body of my research work and engineering type of work presented in Internet Drafts, patent applications and other conference papers that were not included in the article dissertation but listed as references.

The main contributions presented in this thesis are:

1. I researched and designed architectures, protocols and algorithms that enable efficient provisioning of community services in the mobile environment. This work is presented in P2, P3, P4, P5, P7 as well as in my IETF drafts [6][7][8][9][10] and patent applications [11][12][13].
2. I researched and presented user requirements, needs and potential business scenarios for community services in the mobile environment. This work is presented in P1, P5, P8.
3. I designed and implemented three mobile community services that are based on P2P technologies namely: a content sharing service, a knowledge sharing service and a real-time communication service. This work is based on the results of my research on architectures, protocols and algorithms as well as my user and scenario studies discussed above. The services are presented in P2, P3, P5, P7. The services were demonstrated at research conferences [14][15] and at the Nokia World 2008 event in Barcelona.
4. I researched and presented how to distribute IP Multimedia Subsystem infrastructure utilizing P2P technologies. This work is presented in P6 as well as in the conference paper [16]. The patent has been granted on this invention [17].

I also conducted supporting research activities that formed an input to and complemented my research work in the main research areas presented above. Among others I worked on the following research items. The results of the supporting research are fully presented in publications that are listed as references and are not part of the included publications:

a) I researched technical constraints of mobile service provisioning in large scale P2P mobile networks such as battery, traffic, as well as mobile hardware constraints. The research led to the design of the mobile P2P network architectures and protocols. This work is presented in [18].
b) I also studied P2P traffic in operator networks and surveyed P2P traffic management mechanisms. This work is presented in [2].
c) I researched security aspects of P2P services and co-authored P2PSIP requirements draft in the IETF [19]. I also co-developed a P2P security mechanism for VOIP services utilizing the ZRTP protocol. The mechanism is presented in the conference paper [20]. I also designed a mechanism for secure data update in DHT based networks. The mechanism is presented in the patent application [21].
d) I researched and developed a mechanism and an application for maintenance of social network connectivity. The mechanism utilizes a SIP resource event package. My research on the mechanism, application and SIP resource event package are presented in [22][23].
e) I researched and developed NAT traversal mechanisms for P2P services that were published in P7 and the patent applications [24] [25].
f) I researched and developed mobility mechanisms for P2P services that were published as patent applications [26][27].
g) I co-authored an IETF draft about the P2PSIP protocol proposal called P2PP. The protocol was merged with two other P2PSIP protocol proposals into the P2PSIP protocol proposal accepted by P2PSIP WG in the IETF as a working group item. This work is presented in [7].
h) I designed SIP protocol extensions that enable for content and resource sharing. The extensions are presented in the IETF drafts [23][8][28][29][30].
i) I also conducted an analysis of a music distribution ecosystem and developed a model that helps in evaluating the strengths and weaknesses of various mobile content distribution platforms [31].
j) I analyzed a content sharing ecosystem taking as an example the music industry. This work is presented in [32].

The different aspects of my research presented above are discussed in more details in the following sections of the overview. The presentation is structured using a community services provisioning framework described in the 1.3 section.

## 1.3 Scope of the article dissertation

The rest of the article dissertation is organized around a mobile community services framework presented in Figure 1.

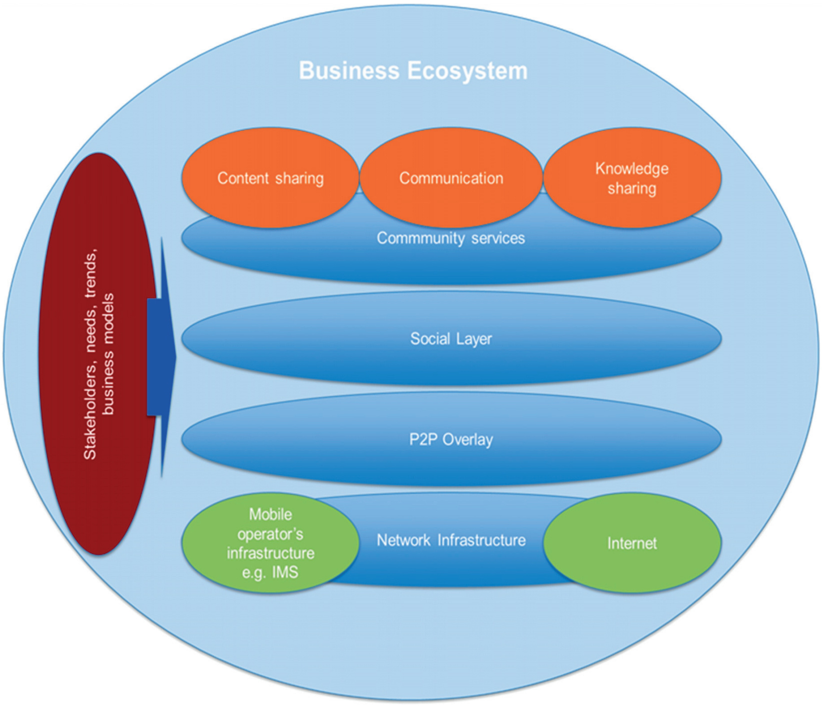


**Figure 1. Mobile community services framework**

The mobile community services framework presented in Figure 1 splits the provisioning of community services into four horizontal layers, namely: Network Infrastructure, P2P Overlay, Social, and Community Services layers and the two levels of technology and business.

The Network Infrastructure layer provides a network infrastructure that is used by the P2P Overlay layer to form P2P overlay networks. Network nodes can be for example dedicated servers, mobile phones or personal computers. Depending on technical realization and business aspects that will be discussed in the next chapters, a P2P overlay can be run on top of different network nodes and systems.

The P2P Overlay layer accommodates different technologies and networks provided by the network infrastructure layer. Figure 1 presents two different realizations of the P2P overlay. In the closed mobile network environment, such as 3G mobile networks, the P2P overlay can be run on an operator controlled infrastructure such as an IP Multimedia Subsystem (IMS) or

Distributed IP multimedia Subsystem (DIMS). In a loosely controlled environment such as the Internet, P2P overlay can be run on mobile phones or personal computers in a similar fashion to Internet content sharing networks or on servers maintained by P2P service providers. P2P technology can also be used to distribute a mobile network infrastructure, especially IMS as we described in [P6].

A P2P content sharing network that works on top of IMS in an operator controlled environment is presented in [P3]. A P2P overlay, where P2P network nodes are located in the Internet either in the end user devices, for instance mobile terminals, or dedicated network servers are presented in [P2] and [P7]. [P7] presents the example of a standalone multi-service overlay network as well as the implementation and measurements of the world's first mobile P2PSIP system. The description and results from pioneering implementation of a mobile P2P content sharing system that works on top of the SIP infrastructure is presented in [P2]. Mixed implementations where some of P2PSIP nodes are dedicated network servers and others are end user devices are also possible.

Separation of the P2P Overlay layer from the Network infrastructure layer allows the P2P overlay layer to accommodate different network technologies. Nodes that form the P2P overlay network may be located in different network domains, for example in operator controlled domains as well as in the Internet. Therefore, a P2P overlay may span different network technologies and network domains. The P2P Overlay layer will be discussed in Chapter 4.

The Social layer enables a P2P overlay network to support social networking such as group formation, social search, and social communication. An example realization of the Social layer is presented in [P4] and [P5]. The social layer will be discussed in Chapter 5.

The Community Services layer represents community services that are provisioned using socially enhanced peer-to-peer networks. This article dissertation presents three different mobile community services, namely: content sharing [P1][P2][P3] and communication [P7], as well as knowledge sharing [P5]. The community services will be discussed in Chapter 6.

As it is presented in Figure 1, the mobile community service provisioning ecosystem is shaped not only by technical enablers but also by end user needs, trends and business models. The business aspects of mobile community services provisioning and end-user needs are presented in [P1], [P5], and [P8], and will be discussed in Chapter 2.

# 2. Business implications

Change is constant in technology and network standards, the industry is continually developing new technologies that can minimize the capital and operating expenses (CAPEX and OPEX) of mobile networks. The 3G, Wi-Max, WiFi and Long Term Evolution technologies not only allow reducing the cost of network capacity but also facilitate mobile services that require increased network bandwidth. Mobile operators have invested heavily in 3G licenses and network equipment. The higher capacity of 3G systems was supposed to become an enabler for innovative data services that could bring new sources of revenue to help buck the trend of falling voice Average Revenue Per User (ARPU). Unfortunately, initial 3G adoption rates were below expectations. This was mainly because of inadequate pricing, hastily conceived services (video call or video streaming were not appealing to the customers), and the low penetration of 3G terminals. The situation is, however, changing.

The increasing penetration of more powerful mobile handsets with larger memories and advanced multimedia capabilities provides a good basis for the development of peer-to-peer applications also in the mobile environment. However, it is not certain if the P2P services that were well adopted in the fixed Internet can also be successfully launched in the mobile environment. Is there any customer need for such services? What business scenarios can we expect in the future? This chapter will address these problems, looking at user demands and requirements as well as the business aspects of Mobile P2P services.

## 2.1 Mobile P2P service provisioning ecosystem

Figure 2 presents an adopted Porter's Five Forces model representing the forces shaping the mobile services ecosystem. As can be observed, the mobile services ecosystem is shaped by many factors such as: economy development, regulation, Intellectual Property Rights (IPRs), the level of competition, new entrants, supplier's power, buyer's power and needs as well as by new products and services that can substitute for already available products.

Around the center of the figure, that represents a mobile service, reside entry barriers to the market. These are of two kinds. There are, firstly, the de-facto entry barriers such as the massive investments that are needed to build a mobile network or access copper to homes in a specific economic situation and, secondly the legal barriers such as frequency licenses and copyrights. The nature of these entry barriers that are typical for information services and network services is such that they help to create monopolies or oligopolies leading to slow innovation and high prices. This is detrimental to the society as a whole. Therefore, the governments exercise regulation and business area specific policies. The idea of these is to help to direct the suppliers of the goods in question to act as if they offered their goods on an open market.

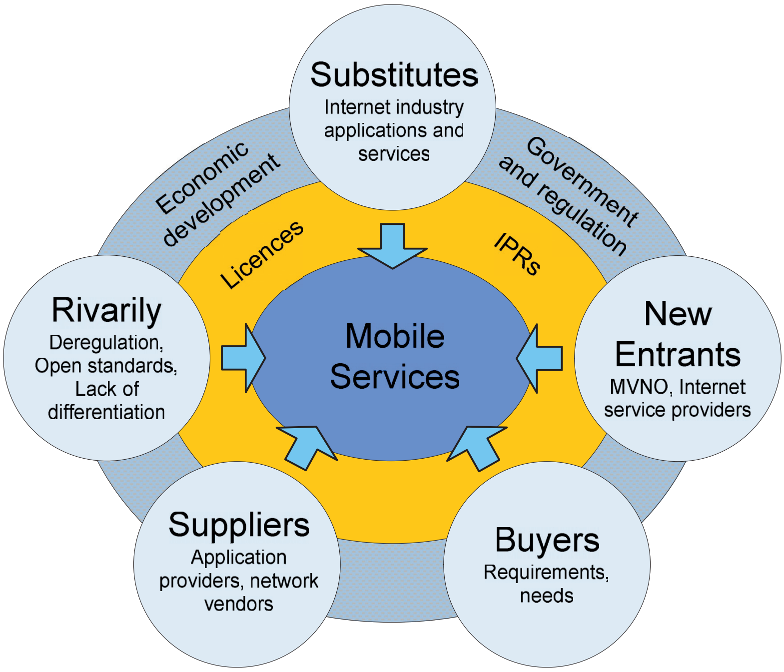


**Figure 2. Major forces shaping the mobile industry**

In recent years, it has been easy to observe increased competition in the mobile industry. Regulation and governmental policies opened a room for new players and increased innovation. The competition came from different industries. New players started testing new business models. Some players became Mobile Virtual Network Operators (MVNOs) and started offering mobile services by using the infrastructure of incumbent mobile operators. Many of them tried to differentiate their offering from the incumbent operators' product portfolio by changing pricing models, introducing new services as well as targeting niche markets and tailoring their offerings to the requirements of the groups targeted. The competition came also from players offering similar services over other technological platforms. Internet service providers that used to offer their services to PC users also started targeting the constantly increasing population of mobile phone users. Among many other services, they started testing P2P services for their potential viability.

In [P8] we present the analysis of stakeholders, trends and uncertainties in the mobile P2P service provisioning ecosystem. Also discussed are likely future business scenarios of mobile P2P services taking as an example a P2P real-time communication service. The P2P service provisioning ecosystem consists of the following stakeholders:

*Network operator*

A network operator operates the network where the service is provided.

*Service operator*

A service operator provides a mobile P2P service. New players such as MVNOs and Internet service providers entered the mobile service provisioning market offering services such as P2P real-time communication that has been very popular in the Internet domain but not available for mobile users.

As we already discussed there can be a separate network and service operator. The roles of operators may also be combined. A mobile operator may operate a mobile network and provide mobile P2P services. In some scenarios, a separate service infrastructure operator may be needed for operating a P2P overlay network for the mobile P2P services.

*Device provider*

A device provider provides mobile devices which are used to access the service. We have to remember that adoption of mobile services is driven by sales of handsets. Users perceive services through mobile terminals and their functionalities. Therefore, the development of the mobile P2P services business is highly dependent on the sales of multimedia terminals.

Network and service operators may be irrelevant if the network is assembled on an ad-hoc basis. In the ad-hoc set-up the mobile P2P service is operated by users themselves who run mobile P2P applications on their mobile devices.

*Application provider*

An application provider provides a mobile P2P application. As discussed in the previous chapters, the P2P applications may range from content and knowledge sharing applications to applications that enable real-time communication using the distributed resources of a P2P overlay network.

*Network vendors*

Network vendors supply the equipment necessary for providing mobile P2P services. A network vendor may provide a wireless access network infrastructure such as WiMax, WiFi or Long Term Evolution. The network vendor may also supply a service network infrastructure such as an IP Multimedia Subsystem that can be used for providing P2P services [P3][P4]. Network vendors may utilize P2P overlays for self configuration and scalability of IP Multimedia Subsystem as is presented in [P6].

*Software platform providers*

A software platform provider (for example, Apple, Google, Microsoft, or SUN) provide a software platform that is used to run P2P applications. In special cases, mobile P2P can be included in the software platform.

*Regulator*

A regulator imposes regulation on the other stakeholders. The regulator, for example, could be a national agency or a European Union level agency. Technology development is shaped by regulatory practices. Deregulation that encourages the efficient operation of markets intensifies competition in services and pricing models.

*End users*

End-users, are those people who actually use the mobile P2P services. Also, non-mobile users can participate in the same network.

## 2.2 Definition of the Product Feature Set

User studies and identification of necessary attributes, capabilities, characteristics, and quality of a product are vital in product design and concept validation. In [P1], [P5], [P6] we study the needs and requirements of the stakeholders of the mobile P2P services ecosystem. Our study includes an analysis of end-users (both consumers and business users) as well as service and network operators.

### 2.2.1 Consumer needs

In [P1] we studied the user needs for mobile P2P content sharing services. Our study concentrates on the mobile phone user view on P2P content sharing and complements conventional (non-mobile) user studies presented in [33][34][35][36][37].

Our paper presents results of a survey conducted among 98 potential users of a mobile P2P content sharing service in Finland. It studied the respondents' willingness to share content stored in their mobile phones as well as their interest in downloading content created by the content industry and by other mobile phone users. The paper also presents the respondents' attitudes towards mobile P2P content sharing services.

The results of our study suggest that mobile content produced by family members and close friends such as pictures, video clips and text files, is as popular as, or even more popular than, content created by the content industry.

The study revealed that there is a strong free riding attitude towards sharing content with unknown people. Concerning willingness to share personal content, respondents preferred to share their personal content with their family and close friends rather than with other social groups. Pictures and video clips were considered as more personal than text documents and other commercial content.

On the other hand, the respondents were concerned about whom they share professional content with. This may suggest that respondents are also concerned about the costs of sharing content such as access fees and battery life, or it might be a sign of increasing awareness about legal issues related to sharing commercial content.

The survey showed that the younger respondents are more interested in sharing content. Moreover, users of the current fixed peer-to-peer applications are more interested in the mobile P2P content sharing applications than other users.

These results show that there is a clear interest in sharing both private and professionally created content on mobile phones but there is a strong need for group management features in mobile P2P content sharing applications. On average the results show that groups of family and closest friends seem to be quite small in size. This suggests that P2P content sharing services have to support quite a fine granularity and flexibility in group formation. The mobile domain asks for modification of existing peer-to-peer applications that do not support privacy and group management or for creating totally new ones.

We also gathered respondents' opinions about mobile applications that could be implemented using peer-to-peer technology. The results show that a mobile P2P service can win new users because of the distinct usage environment, especially if group management was available.

Using the results of this study we designed and implemented the worlds first SIP based mobile content sharing system that will be presented in the following chapters.

### 2.2.2 Business users needs

The previous study [P1] indicated that willingness to share content drops dramatically in social groups other than close friends and family such as work and study colleagues. Respondents indicated that they are much less willing to share content and documents with work and study colleagues. This was a very interesting result especially since we live in a knowledge driven society where knowledge is one of the most important resources of our times [38][39]. Employees gain more and more knowledge that is specific to their work [40]. This knowledge needs to be made transparent in an organization. It has to flow from one organizational unit to another, from one employee to another filling up knowledge gaps. This transparency is required to avoid reinvention of the wheel and is priceless in stimulating innovation. We decided to investigate this problem in more detail to understand the reasons behind such behavior. The results of our studies are presented in [P5]. The article dissertation presents possible barriers for knowledge sharing in the corporate environment. We identified a number of barriers such as problems with interpretation of shared information and expression of ones knowledge [41][40][42], poor quality information that is shared with co-workers [43][44][41], preference of own solutions [40], unwillingness of knowledge workers to give away their knowledge because knowledge is perceived as power [45][46], fear of criticism or sharing poor quality information [47][41][48] as well as unwillingness to show particular individual competences to avoid unwanted assignments [49].

Our study showed that people are more likely to share their knowledge with members of previously established social communities [50][51][52][53][54][55]. When people get to know each other they are able to predict what to expect and how the other party will behave in certain situations [56]. When trust is strong, people are more willing to openly communicate and exchange knowledge without fear of "loosing face" or that someone will intentionally ridicule them in public [57][47]. This has implications both for knowledge sharing as well as usage of knowledge management systems as a source of information [44].

We also presented possible incentive models that can be applied in the corporate environment in order to stimulate knowledge sharing. We showed that financial incentives [50][46] achieve faster short-term results and may be useful to get a knowledge management project started [58], whereas soft incentives [59] and community based knowledge sharing that is based on trust have more of a long term positive effect.

Employees should not be forced to use knowledge sharing systems. The Knowledge Management (KM) systems must be designed to support their work and their social behavior. What constitutes a good system for one profession or employee does not necessarily have to be suitable to another one. Using knowledge about possible barriers and effectiveness of different incentive models, we have designed and implemented a peer-to-peer knowledge sharing system that is dedicated to mobile platforms such as cellular phones and inspired by the manner in which people interact in everyday life. The system is presented in [P5].

### 2.2.3 Network operator and service provider requirements

Looking more on the aspects of service provisioning we see that the competitive landscape of service provisioning is changing. In [P6] we studied how P2P can address the requirements of network and service operators in the fast changing mobile service ecosystem.

Average Revenue Per User (ARPU) from conventional mobile services is flat and in some markets is falling gradually. Commoditization occurs in voice and SMS services because of the lack of differentiation between the services offered by different operators. Moreover, many experts expect flat rate data plans to be a dominant charging model in the future. Because of the lack of differentiation in the traditional mobile services market and flat rate charging models, mobile operators are increasingly looking for cost efficient solutions. As it will be shown in Chapter 4, conventional service infrastructures can benefit from P2P technologies. P2P may improve scalability, self-configuration and automatic reaction to changes reducing capital and operational expenditures (CAPEX and OPEX) for service and network providers. On the other hand, as it was mentioned in Chapter 1, P2P may significantly increase traffic in operator networks forcing operators to invest in the upgrade of their infrastructure [1][2][3]. [P6] also addresses this problem by allowing an operator to control P2P traffic.

## 2.3 Business scenarios

Using information about trends, uncertainties and the value network, strategists and business development personnel evaluate emerging mobile services, or the future development of an emerging technology.

In [P8] we developed a scenario planning methodology for emerging mobile services decision making. We found the Schoemaker's variant of scenario planning [60] to be a suitable method for decision making. The main challenges in using the method are related to identifying the most relevant trends, uncertainties and stakeholders, and building consistent and coherent scenarios. However, the iterative nature of the method aids significantly in eliminating possible inconsistencies. The method is a systematic way to assess possible future outcomes. Scenario planning serves as a basis for a more detailed analysis, for instance, using quantitative methods.

Nine trends and six uncertainties were identified and consistent and coherent learning scenarios which depict a wide range of possible settings for emerging mobile P2P real-time services were developed using use proposed methodology. According to our analysis low firewall, NAT and battery constraints are important determinants in the success of real time services implemented using P2PSIP technologies. A potential stakeholder should seek settings where network and legal aspects are the most favorable. According to our analysis, the small-scale ad-hoc and private environments without interconnectivity to the Internet may benefit from P2PSIP technology. Revenue collection in those settings is possible by bundling devices and services in ad-hoc environments, and by charging installation and support fees in private environments. The stakeholders looking for global service provisioning should consider a public global semi-centralized scenario. In this scenario the service operator would gain its competitive advantage either by cost advantage over established network operators or by offering new features to users. The revenue model in this scenario could be transactional income from value-added transactions such as interconnections to other networks, or advertising. A subscription based revenue model is unlikely as consumers are accustomed to VoIP, IM and presence services provided for free.

## 2.4 Chapter Summary

In this chapter we elaborated on Mobile P2P service provisioning ecosystem, presenting stakeholders and their roles in the ecosystem. We conducted an analysis of end user needs and requirements regarding mobile P2P services. The analysis showed that there is a clear need for mobile community services, however, mobile phone users have different requirements than fixed Internet users. They are more concerned about privacy and the costs of sharing content such as access fees and battery life of mobile devices. We showed that there is a strong free riding attitude towards sharing content with unknown people. The uptake of mobile community services targeting business users is dependent on the right incentive models and very flexible group formation and social ties management functionality provided by mobile P2P systems. We also showed mobile operators requirements regarding P2P service infrastructure and a scenario planning methodology that can be used to create the most likely business scenarios for mobile P2P services.

# 3. Network infrastructure layer

As we have discussed in the previous chapters, mobile P2P technologies can be applied to both an operator controlled environment as well as to the Internet. Each environment has different characteristics and implementation of P2P in those environments can be quite different. In this chapter, we present a few examples of network infrastructure that can be used to provide mobile P2P services. As it will be shown in the next chapter, P2P technologies can also be used to create a distributed design of the mobile network infrastructure possibly minimizing OPEX and CAPEX.

## 3.1 SIP networks

Internet real-time communication is typically realized using the SIP communication model. Session Initiation Protocol (SIP) [61] is a signalling protocol that is used to establish, tear down and control multimedia sessions. SIP uses an Address-of-Record (AOR), to locate a user. An AOR is a SIP or SIPS URI that includes a user and a domain part. Sip:marcin.matuszewski@nokia.com is a typical SIP URI, where the user part is separated from the domain part with @. Although two SIP User Agents (SIP UAs) can communicate directly without any SIP infrastructure, most deployments require proxy and registrar servers that assist SIP UAs in multimedia session management. The proxy servers are used to help routing SIP requests to the location of other SIP UAs, authenticate and authorize users for multimedia services as well as implement call-routing policies. The registrars allow users to upload their current locations (typically in the form of IP address and port) to the location server. This information is used by the proxy servers to route a SIP request to a destination. In practice, registrars, proxies and location servers are all located on a single physical machine called a SIP server.

In a typical SIP communication model (called sometimes the trapezoid model) each SIP UA is registered in a particular domain. If a caller registered in a domain A wants to call a callee registered in a domain B, it sends a SIP request to its own proxy server. The proxy server resolves the domain part of the SIP URI of the callee (in this case domain B) using DNS and forwards the request to the proxy server of the domain B. The proxy server in the domain B forwards the received request to the terminal of the callee using contact information stored in a location server of the domain B.

SIP was standardised in the Internet Engineering Task Force (IETF)[2], being designed to support real-time communication in the Internet. SIP is mainly used by enterprises for real-time services in their enterprise networks and Internet SIP service providers.

## 3.2 IMS networks

The IP Multimedia Subsystem (IMS) [62] offers a platform where mobile and fixed network operators can provide IP services to their subscribers. IMS adopted the Session Initiation Protocol (SIP) protocol to establish, tear down and control multimedia sessions in operator controlled mobile networks.

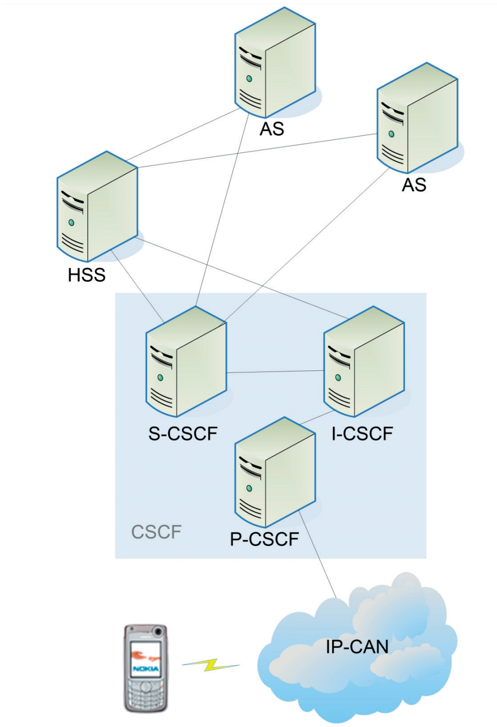


**Figure 3. IMS network**

The IMS architecture [62] was designed around the year 2000 as a technology that allows mobile operators to migrate towards IP-based packet networks. IMS inherited a few concepts from the GSM core network architecture. IMS defines among others the Home Subscriber Server (HSS), Proxy-Call/Session Control Function (P-CSCF), Serving-Call/Session Con-

[2] Internet Engineering Task Force (IETF), online: www.ietf.org, last time accessed on 12.05.2013.

trol Function (P-CSCF), Interrogating-Call/Session Control Function (I-CSCF), and Application Server (AS) functional elements. User devices can be connected to IMS through any IP connectivity access network (IP-CAN). Figure 3 presents a simplified IMS network.

HSS is an evolution of the GSM Home Location Register (HLR). The HSS is a general repository of persistent user data, such as user names, allocated identities, shared keys, allocated services, policies, permissions, etc. It is the central database in the IMS architecture, like the HLR is the central database in GSM.

P-CSCF is a SIP proxy that is the first point of contact for an IMS terminal (e.g. a mobile phone that is allowed to register to IMS). It is assigned to the IMS terminal during a registration process. It authenticates the user using authentication information stored in the HSS. Other nodes trust the P-CSCF, therefore, they do not have to authenticate the user by themselves. The P-CSCF stays on the path of all signaling messages originated or destined to the IMS terminal for the whole duration of the registration. It also compresses and decompresses SIP messages, using the SigComp [63] compression method, establishes IPsec [64] tunnels towards a terminal, and generates charging records.

S-CSCF is responsible for controlling SIP sessions and triggering service execution. The S-CSCF routes SIP messages to the correct destination. It also handles SIP registrations. The HSS assigns the S-CSCF to a user during the registration process. The S-CSCF uses the DIAMETER protocol [65] to communicate with the HSS. Based on service triggers, the S-CSCF forwards SIP messages to a particular Application Server (AS) or a collection of ASes that execute service logic.

I-CSCF performs different functions. It is able to hide the internal network of an operator from the outside world by modifying SIP messages before forwarding them to a foreign IMS network. It is also able to locate the S-CSCF allocated to a user to which a received SIP request is headed using the user's public identity.

The AS executes a service logic. There can be specialized ASes for certain services. The AS may be maintained by a mobile operator or by a third-party service provider that signed a service provisioning contract with the mobile operator. The AS communicates with the S-CSCF using the SIP protocol. If located in the home network, it can query the HSS using the DIAMETER protocol.

## 3.3 PlanetLab

A service operator may use geographically distributed servers for running P2P overlays. PlanetLab[3] is a group of computers available in the Internet. They are used as a test bed for computer networking and distributed systems research. PlanetLab can be used for running trials of new Internet services and applications as will be demonstrated in Section 4.2.4.

## 3.4 End user devices

The alternative to the above is running a P2P overlay on end user devices without or with minimum support from centralized servers such as an enrolment server. In the ad-hoc/private deployments the P2P overlay may be formed directly between end user devices such as mobile phones and personal computers. In larger deployments overlay networks may use resources of the end user devices as well as resources of centralized servers and the Internet. The centralized servers may assist end user devices in NAT traversal, interconnection with SIP, IMS or circuit switched networks, authentication and bootstrapping.

## 3.5 Chapter Summary

In this chapter we presented four examples of network infrastructures that may be used by the P2P Overlay layer to form P2P overlay networks. The SIP and the IMS infrastructure use the SIP protocol as a base signaling protocol. Both use the client-server architecture with a few servers handling a service logic centrally. In order to run a P2P overlay on top of the SIP based infrastructure, the P2P overlay must use the SIP protocol as a signaling protocol. SIP may require some modification in order to provide required functionalities for the overlay as it was presented in [23][8][66][9]. The third example of network infrastructure was PlanetLab. PlanetLab is a collection of hundreds of servers connected to the Internet. Researches may run various overlay networks on top of PlanetLab. Unlike the SIP based infrastructure, in PlanetLab there are no constraints regarding the signaling protocol. Some P2P overlay implementations running in the PlanetLab,

[3] PlanetLab, online: http://www.planet-lab.org, last time accessed on 12.05.2013.

such as OpenDHT[4] use web centric protocols such as the Extensible Markup Language Remote Procedure Call (XML RPC). It is also possible to run the P2P overlay on end user devices connected to the Internet such as mobile phones and personal computers. In this scenario the service provider only maintains a few servers that support end user devices in P2P network formation and interconnection with other networks. There are no constraints regarding the signaling protocol in this scenario except NATs and firewalls.

[4] OpenDHT, online: http://www.opendht.org, last time accessed on 12.05.2013.

# 4. P2P overlay layer

This chapter discusses the P2P overlay layer. Based on the results of our scenario planning studies we selected the most favorable scenarios for further research on the technical level, namely: the semi-centralized public global scenario and the ad-hoc/private scenario. In our research we put emphasis on access fees, firewall, NAT and battery constraints as important determinants of the success of services implemented using P2P technologies. Before presenting the mobile P2P overlay networks we will discuss the differences between various P2P overlay network architectures and the impact of the design decisions on the mobile P2P service provisioning.

## 4.1 P2P overlay network architectures

P2P overlay networks are typically classified according to the network structure and the degree of centralization. The first refers, roughly, to the P2P network topology construction and maintenance. The second is related to the existence of dedicated peers called sometimes "super-peers" that perform special services for other P2P overlay nodes. Moreover, the resource discovery works differently in different types of P2P networks. We can further classify P2P overlay networks as those that support partial match searches and those that allow only exact match searches. The choice of the architecture depends on the available network infrastructure, capabilities of network nodes, specific applications as well as the business aspects of mobile service provisioning.

### 4.1.1 Network structure

Nodes in P2P overlay networks can be organized in a structured topology, for example a ring topology, or every node can select peers it wants to connect to arbitrarily. The former are called structured P2P overlay networks, whereas the latter ones are unstructured P2P overlay networks. The structured P2P overlay networks are typically P2P networks implemented with a Distributed Hash Table (DHT) such as Kademlia [67], Bamboo[5], or Chord [68]. The unstructured P2P overlay networks are typically P2P networks

[5] The Bamboo Distributed Hash Table, online: http://bamboo-dht.org/pubs.html, last time accessed on 12.05.2013.

implemented with flooding algorithms such as Gnutella[6] or Kazaa [69]. Since the publication of the included articles semi-structured networks have also been proposed [70].

In unstructured P2P networks, flooding algorithms require a node that receives a search request to forward the request to all or a subset of nodes it is connected to except a node that sent the request. This way the search request traverses many nodes until the desired piece of information is found or a TTL (Time To Live) value stored in the request message prohibits a peer from forwarding the request further. Therefore, the likelihood of finding the desired piece of information is directly related to the distribution of that piece of information across the network. In other words, it is typically very easy and efficient to locate popular content in the unstructured P2P overlay network because the content is available at several peers in the P2P overlay network. Conversely, rare long-tail content shared by a few peers may not always be found before TTL expires The likelihood of finding the desired information is related to the probability that the requester (the peer searching for the information) is within a few hops of a peer holding a copy of the information.

Flooding-based systems are not very suitable for large scale mobile environments due to the large amount of traffic they require [71]. In an extreme case, a search operation requires flooding of the requests to all of the nodes in the P2P network in order to locate a desired resource. It is possible, however, to limit the scope of flooding, in order to minimize the maximum usage of network (especially radio) resources. However, this operation minimizes the chances of obtaining a successful search result.

On the other hand, building a structured P2P overlay in the mobile environment that remains consistent as peers enter and leave a network is also not a trivial problem. A Distributed Hash Table (DHT) is a structured overlay network architecture that allows for locating any piece of data stored in the overlay network using a limited number of messages, typically log($N$) messages, where N is the number of peers. Each peer in the DHT overlay gets a unique *peer ID,* sometimes also called *node ID*, when it joins the DHT. A Peer ID is typically a 128-bit or a 160-bit unsigned integer representing a position in the keyspace.

Taking as an example a ring topology that is presented in Figure 4, each node in the DHT ring has at least two direct neighbors: a predecessor and a successor peer. Sometimes, in order to increase the reliability of the network, a peer is linked to two, three or more predecessors and two, three or

[6] Gnutella, online: http://rfc-gnutella.sourceforge.net/index.html, last time accessed on 12.05.2013.

more successors, but, in general, the number of predecessors and successors a peer is linked to is small.

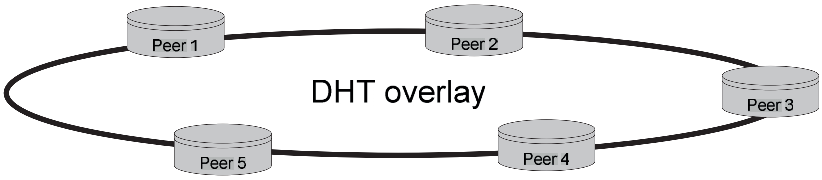

**Figure 4. An example DHT overlay with ring topology**

The place in the DHT ring is determined by the peer ID. The successor peer is a node whose peer ID is the next one in the ID ring when moving in a clockwise direction. The predecessor of a node is the next node in the ID ring when moving counter-clockwise. This allows a node to contact its predecessors and successors. We refer to the node's predecessors and successors collectively as neighbors.

Additionally, each peer also maintains a routing table, sometime called the finger table, which includes a list of peers and their IP addresses/port numbers (or/and other connectivity information) in the overlay network. The number of peers in the list is typically much smaller than the number of peers in the overlay. The finger table is a kind of collection of shortcuts to move throughout the DHT ring, so it is used to minimize the number of hops needed to locate a resource in the network.

Data stored in the DHT is indexed using data IDs. These data IDs are distributed among DHT nodes, each DHT node storing a subset of data IDs.

In the example DHT, the structure of the ring has to be maintained. This includes keeping track of predecessors, successors, its liveliness, and maintaining the finger table. Typically, neighbors send keepalive messages to each other checking if a neighbor is still online. If a peer leaves the overlay, the network has to adapt to the changes:

- a predecessor and a successor of a peer have to become immediate neighbors,
- one of the neighbors has to become responsible for the pool of data IDs that a peer that left a network was responsible for,
- if the overlay network uses a replication mechanism for data protection, the node that takes responsibility for the pool of data IDs that were stored in the node that left the network has to update its database storing data IDs and associated data,
- the peers that had a pointer to the peer that left the network in their finger tables have to update their finger tables.

This introduces an additional traffic overhead between peers and requires additional processing power needed to modify information about neighbors and finger tables. As it has been shown in [18][72], maintaining network DHT structures can have a significant impact on the battery life of mobile phones as well as the traffic sent over the air interface. This makes the implementation of a large scale DHT overlay running solely on mobile devices difficult in practice.

### 4.1.2 Resource discovery

From the end user perspective, P2P networks can be classified as those which allow only for exact match searches and those which can perform searches based on partial matches. The former are typically structured P2P networks implemented with a Distributed Hash Table. The latter are typically unstructured P2P networks such as Gnutella[7] or Kazaa [69] implemented with flooding algorithms.

In content sharing networks, users are not typically aware of a full keyword when they want to search for content such as a file, so, from the point of view of the user's requirements, the capability of performing partial match searches in content sharing networks is a must. This would lead to implementing flooding-based systems in P2P overlays in mobile environments. However, the search mechanism in the existing unstructured P2P networks may produce many irrelevant search results. Mobile phones have limited capabilities, such as a limited screen size, small keypads, and a limited battery lifetime. Displaying hundreds of irrelevant search results on a 2.4” screen is not an option. The search process has to be context aware and return results that are highly relevant to a particular user.

In contrast, partial match searches are not necessarily required in real-time P2P overlay networks such as P2PSIP [P7], [7] and Skype [73]. In those overlay networks, users typically search for contact information of other users using exact Uniform Resource Identifiers (URIs) such as sip:marcin.matuszewski@nokia.com. On the other hand, the real-time P2P overlay networks require a fast lookup process and high reliability. VoIP systems must be able to establish a call between a caller and a callee within a few seconds. Therefore, a callee must be located explicitly in a limited time period. Assuming that P2P VoIP networks would be implemented with a flooding algorithm, it would be very difficult to predict how much time and resources would be required to locate the callee. Structured P2P net-

[7] Gnutella, online: http://rfc-gnutella.sourceforge.net/index.html, last time accessed on 12.05.2013.

works are much more suitable for VoIP P2P networks since they are able to explicitly locate any resource using typically Log(N) messages.

As a consequence, the P2P overlay networks that support many different mobile services have to support a mixture of structured and unstructured P2P networks. A Mixed P2P network architecture is presented in [P4] and will be discussed in Chapter 5.

### 4.1.3 Degree of centralization

P2P overlay networks can also be classified according to the degree of centralization into three categories: centralized, pure, and semi-centralised P2P networks.

A *centralized P2P network* assumes the existence of a central entity that controls the network. In this architecture, the central node maintains the P2P overlay by keeping information about all peers in the overlay. This information may include their connectivity information, presence status and information about resources being shared. All peers connect to a central node in order to locate other peers or resources in the network. The main difference to the traditional client-server architecture is that in the centralized P2P networks communication or file transfer happens between peers directly without any intervention of the central node. Moreover, peers store data they want to share with other nodes thus sharing storage and computing resources with other peers.

In this architecture the central entity creates a single point of failure in the network. Moreover, in a situation when millions of P2P service users are distributed among different mobile operators, even different countries, this solution faces major problems with scalability and control placement. This network architecture was used in Napster, the first P2P content sharing system.

In contrast to the centralized network architecture, in a *pure P2P network* architecture there are no central entities that control the P2P overlay. In this architecture all devices are equal peers. All peers share the task of locating resources in the overlay, resource publication, and maintenance of the P2P overlay network. In the pure P2P overlay network architecture every device needs to be able to maintain the overlay by maintaining connections to a subset of peers and assist other peers in a search process. Because of this design decision, the pure P2P architecture is more resilient to node failures. However, it puts requirements in terms of battery consumption, processing power, memory, and bandwidth on all nodes in the P2P overlay network. These requirements can be quite challenging especially in the mobile environment where nodes are battery-powered devices such as mobile phones, PDAs, or WiFi phones [71].

Moreover, devices in the pure P2P architecture have to maintain many connections to other peers, even though some of the peers can be behind NATs. NAT traversal mechanisms may allow peers to establish connectivity with other peers that are behind NATs. However, NAT traversal mechanisms introduce additional traffic and processing overhead that has impact on battery lifetime.

There are also other reasons that may have an impact on the decision of a mobile device to become a peer in the pure P2P architecture. Low-end mobile devices may have too little memory, too slow CPU and may not support fast packet radio interfaces. A user may not be willing to assume the cost of increased bandwidth consumption related to the P2P applications. On the other hand, a mobile operator policy may not allow mobile devices to become peers. Some operators may want to provide high quality of service and therefore they may not want to risk that the search delay increases significantly when many mobile devices become unreachable or just leave the overlay network.

For these reasons, not all battery-powered devices even though having sufficient capabilities will be able to become peers in the pure P2P architecture. For those devices another P2P node type is introduced. The devices that cannot become peers will have to act as clients, sometimes called ordinary peers. A client interacts with the P2P overlay by contacting one or more peers using a client protocol. In contrast to peers, clients do not maintain the P2P overlay network and do not route messages to other peers or clients. The client protocol allows nodes that are not eligible to become peers to store, modify, delete and retrieve resource records stored in the P2P overlay. Requirements for the peer and client protocols are presented in [10].

A *semi-centralized architecture,* sometimes called a *hybrid architecture*, represents a compromise between the pure and the centralized architectures. The architecture allows the existence of both peers and clients. A semi-centralized architecture allows devices that cannot become peers to join the overlay as clients. A semi-centralized architecture with four peers and two clients is presented in Figure 5.

As presented in Figure 5, clients communicate with peers using a client protocol. The realization of the client protocol may take many forms. As it will be described in the following sections, the selection of the client protocol depends on network infrastructure and the required extensibility of the client protocol to support different mobile services.

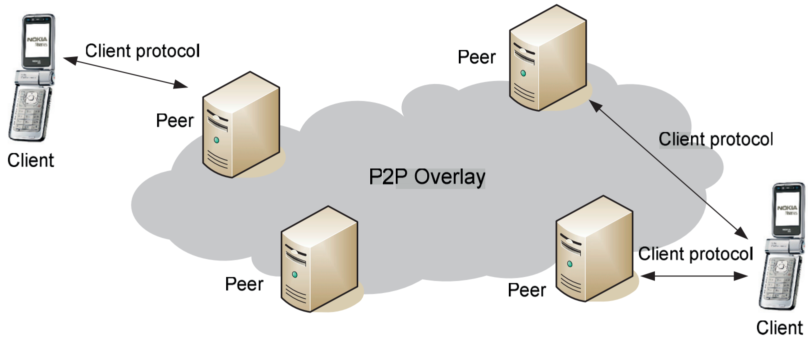


**Figure 5. Semi-centralized architecture**

Hybrid network architectures were discussed in [74][75][76][77]. Kato et al. proposed an XML-based mobile P2P protocol and algorithms for controlling the mobile P2P network. The authors have proposed a hybrid peer-to-peer architecture by introducing a control node. However, the paper lacks any evaluation of the proposed approach. It is limited to the presentation of the architecture and the related protocol.

Hoßfeld et al. analyzed the feasibility of porting eDonkey 2000, an Internet P2P file sharing system, into the mobile environment [75][76]. This work was carried out with the assumption that the popularity of the application is of greater importance than the simplicity of the system and interoperability with mobile network infrastructure and protocols. The authors introduced centralized elements to the eDonkey architecture. The conducted simulations showed that the proposed architectural changes lead to the increased performance of the mobile eDonkey system. However, in order to achieve a significant performance gain, the eDonkey protocol has to be modified causing interoperability problems with the original eDonkey system. Another issue was the feasibility of implementing the proposed system and making it interoperable with existing mobile systems.

Another paper discussed the Project JXTA for Java 2 Micro Edition (J2ME), which aimed to develop Java middleware that allows MIDP-compliant devices to participate in P2P activities with JXTA peers running on more powerful platforms [77][78]. The JXTA for J2ME architecture assumes that mobile terminals would act as windows into the network for their users while most of the computational tasks are performed in the fixed network by servers called JXTA relays. The JXTA relays appear as normal peers in the distributed P2P network. The JXTA project proposes a very interesting approach to the peer-to-peer networking.

None of the above papers fully investigated the complexity of the mobile eco-system and the mobile phone constraints discussed in the previous chapters. The papers concentrated on concepts rather than on applications and systems that can be deployed in cellular networks on wide scale for millions of users.

### 4.1.4 Analysis of P2P network architectures

In order to validate the different p2p architectures we analyzed battery constraints and message overhead in three large-scale mobile P2P configurations [18]. Kademlia was chosen as the DHT protocol of our analysis since then this was the only protocol that was used in deployed networks with more than one million users: the Mainline BitTorrent DHT[8] and the Azureus DHT[9]. We analyzed the following network configurations: the first configuration with all mobile nodes acting as peers (the pure P2P architecture), the second configuration with only a subset of the mobile nodes in the system participating in the DHT and the rest being clients (the semi centralized P2P architecture), and the third configuration where mobile nodes act as clients and they connect to the DHT network formed by network servers (the semi centralized P2P architecture with the usage of network operator's server farm).

*The first configuration: all nodes are peers (pure P2P architecture)*

The most significant benefit of the first scenario is that the load on each node is relatively low because it is evenly distributed among all nodes in the network. However, as the size of the DHT grows, the maintenance traffic and the routing latency also increase.

Since the DHT consists of mobile nodes, we assumed a high churn. High churn comes from a number of factors. First, mobile nodes run on batteries. Second, as the name implies, mobile nodes tend to move, and there could be frequent changes of IP address if the phone moves from one WLAN access point to another or switches from WLAN to cellular network when leaving a hotspot. Churn was modeled with p=0.7 (30%), which means that almost every third messages is lost.

As has been shown in [18], the message overhead per node ($R$) increases with the number of nodes ($n$). This is due to increase of the average routing length and the message overhead of lookup operation. The larger the DHT, the more hops are needed to reach the target of the lookup. With one million nodes, around one message per second is received by each node. Using the average energy cost per message value 0.3 J that were measured in [72],

---

[8] Mainline BitTorrent, online: https://www.bittorrent.com, last time accessed 12.05.2013.

[9] Azureus BitTorrent, online: http://sourceforge.net/projects/azureus/, last time accessed on 12.05.2013.

one message per second per each node corresponds to 1080 J of energy per hour, which means roughly 4-5 hours of uptime with a fully charged Nokia N95 smartphone. These results are inline with the results from our other experiments. The measurements of BitTorrent peer software running in a mobile phone being a part of the BitTorrent Mainline overlay network revealed that the battery of a Nokia N95 mobile phone drains in around 5 hours. 5-hours battery life is not acceptable by mobile phone users that are used to recharging their mobile phones every couple of days, not hours. The power consumption in the pure P2P architecture is affected by the maintenance of the routing state, as well as handling lookups and requests to store data on behalf of other users or nodes in the P2P overlay network.

On the other hand, the traffic introduced by the overlay protocols in the process of network maintenance and handling of lookups and store requests can also be significant. The measurement results presented in [18][72] show that the Nokia N95 mobile phone acting as a peer in the BitTorrent Mainline DHT received and transmitted around 1MB data over the air interface every hour.

*The second configuration: subset of the nodes are peers, the rest are clients*

Figure 6 shows a plot of the message overhead per node as the function of number of peers in the DHT serving a network with 1 million clients. It can be observed that having less than 10000 peers (1%) results in more than 2 messages per second, which is inacceptable in terms of energy consumption. Nevertheless, as the number of service nodes approaches the 3% limit, the overhead decreases below 1 message per second.

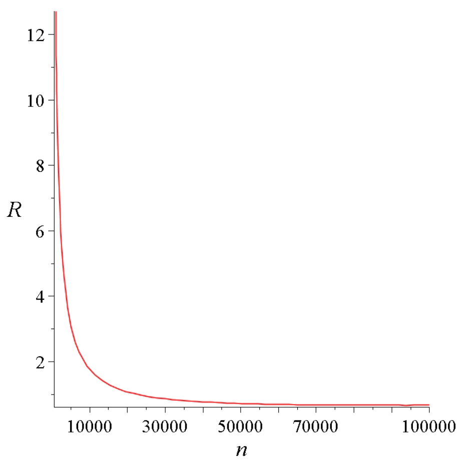


**Figure 6. Message overhead as a function of the number of peers**

If we look at the energy consumption, it turns out that we have an interesting trade-off. To minimize the long-term cumulative energy consumption, we would like to have as little traffic as possible within the DHT. This can be achieved if the minimum number of nodes participates in the DHT. However, Figure 6 shows that the less nodes participate in the DHT, the higher the traffic and energy per node becomes, which can results in exceeded capacity. Moreover, as it is discussed in [6], in P2P real-time communication overlay networks it is advisable to limit the number of peers in the overlay to reduce the lookup delay.

A possible modification is applying a strategy where by default everybody participates in the DHT but when the battery capacity drops below 50% the node leaves the network [12][27].

Another option is to assume that the DHT consists of powerful server computers. Thus the DHT is used to implement the server functionality in a distribution fashion. At the same time the distributed server could take advantage of the scalability and robustness aspects built-in in the DHT algorithms.

*The third configuration: DHT consists of powerful server computers*

Because the devices forming the DHT are server computers, we assume that the DHT is very stable. Churn is small and arises mainly from hardware faults and maintenance actions. We used p=0.99 (1%) for the plots, which means that the probability that a message is sent to an active contact is very high. We assumed that only 1% of sent messages are lost.

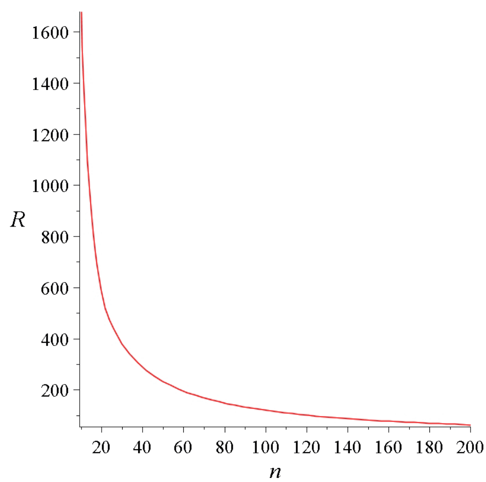


**Figure 7. Message overhead per node (Servers form the DHT)**

Figure 7 shows the number of messages received by a service nodes in a one million user DHT. Since servers can serve much higher number of requests than mobile devices, smaller number service nodes are required in the DHT.

We also analyzed traffic incurred in even larger P2P networks. We studied three scenarios with 10 million mobile nodes running a VoIP service in P2P networks. Table 1 classifies the traffic experienced by DHT nodes by type. It shows the quantitative predications for a few typical configurations. The first column shows the scenario with a single peer handling all of the traffic. The second column shows traffic in the semi-centralized architecture when we assume that the DHT consists of 50 nodes. The last column shows the data for the pure P2P architecture, having all the 10 million mobile phones function as peers.

**Table 1. Message overhead per second per service node in different size networks**

| **Traffic type** | **10 million users, 1 service node** | **10 million users, 50 service nodes** | **10 million users / service nodes** |
|---|---|---|---|
| Registration | 41200 | 1050 | 0.02 |
| Call | 14000 | 360 | 0.007 |
| Maintenance | ~0 | ~0 | 1.4 |
| Total | 55200 | 1410 | 1.427 |

As it can be seen the registration traffic is the main source of the load in the first two cases. Therefore it would be tempting to reduce the need for periodic re-registration. Unfortunately, a longer re-registration period increases the probability that the DHT contains outdated data.

The main difference in the third case is the increase of the maintenance overhead. While having 10 million service nodes greatly reduced the registration and call related message overhead per node, more DHT nodes resulted in much higher maintenance traffic.

Our analysis showed that even though the total traffic per service node is much smaller in the third case than in the first two cases, in the third case every mobile phone would have to handle 1,4 messages per second which is inacceptable in terms of energy consumption. The optimal configuration for large scale mobile P2P deployments is the semi-centralized architecture with powerful servers running peer software and mobile phones acting as clients.

## 4.2 P2P overlay networks in the mobile domain

This section contains a discussion of various P2P overlay networks that are suitable for the mobile environment. We start with presenting two implementation scenarios in 3-rd Generation (3G) mobile networks. The first implementation scenario requires only changes in the network infrastructure but does not require any modifications in mobile terminals themselves. In contrast to the first implementation scenario, the second implementation that does not require any changes to the mobile network infrastructure. It allows running P2P services on top of IP Multimedia Subsystem. The main drawback of this implementation comparing to the first one is the necessity of modification of mobile phone software in order to access P2P services. Finally we present two implementations that does not relay on the mobile infrastructure. They assume that mobile operators provide access to the Internet and internet service providers provide P2P services for all Internet users.

### 4.2.1 Distributed IMS

We start the presentation of mobile P2P overlays from a solution that runs a P2P overlay in the IP Multimedia Subsystem (IMS) nodes.

As has been discussed in Section 3.2, the IMS architecture was designed around the year 2000 and inherited a few concepts from the GSM core network architecture. In particular, IMS provides for the existence of a Home Subscriber Server (HSS), designed as an evolution of the GSM Home Location Register (HLR).

A key concept of the HSS lies in the fact that all the data pertaining to a subscriber is stored in a single HSS. Redundancy is typically achieved by replicating the HSS data into a different (backup) secondary HSS. Due to design constraints, it is hard to plug an add-on to an existing HSS to scale it up to handle a bigger number of subscribers when this factor is several orders in magnitude. The opposite is also true: a large HSS is typically costly for small networks.

Furthermore, if a large network needs to add more than a single HSS to handle all its subscribers, the operator is forced to install a Subscriber Location Function (SLF). This is a kind of a Diameter redirect server that acts as a single point of contact for Diameter clients (e.g., S-CSCF, I-CSCF, AS). The SLF contains the range of address spaces that each HSS is allocated; thus, it is capable of locating the HSS that contains the user-related data. As a consequence, the larger the network grows, the more configuration is required.

While it is true that IMS provides for some sort of I-CSCF and S-CSCF distribution, it is also true that the system is not highly scalable. Bringing more capacity to the network typically involves bringing more I-CSCF or S-CSCF nodes into operation, which is a costly operation due to the high number of required configuration operations.

A solution to these problems is presented in [P6]. The paper introduces a Distributed IMS[10] architecture. The Distributed IMS architecture replaces the IMS core network with one or a collection of self-organizing DHT overlay networks. The idea is based on the fact that DHTs are known to scale very well to a very large number of nodes and allow for creating self configured networks with automatic replication capabilities. This work focuses on distributing the IMS functional elements, such as S-CSCF and I-CSCF, as well as HSS, and up to some degree, presence servers. The document proposes three different configurations of distributed IMS: distributed HSS (HSS DHT), distributed S- and I-CSCFs (I/S-CSCF DHT), and combination of HSS, I/S-CSCF and presence server elements in a single P2P overlay (IMS DHT). In the presented architecture, mobile devices act as SIP user agents and do not perform any P2P networking functions, which are instead executed in fixed network elements. Figure 8 presents a simplified conventional IMS network, a configuration with separate HSS DHT and I/S-CSCF DHT overlays, and a single overlay that performs functions of HSS, I/S-CSCF and a presence server.

The document compares all of the three configurations. The comparison shows that the registration process in a single IMS overlay offers advantages over the configuration with two DHT rings, namely an HSS DHT ring and an I/S-CSCF DHT ring. Additionally, it must be noted that once the IMS DHT is in operation, since the HSS is co-located with the I/S-CSCF, there is no need for a Diameter protocol that allows CSCFs to download or query HSS-stored data. Some other Application Servers may need to implement the Sh interface towards the Distributed HSS, but such interface could be easily replaced by an enhanced subscription/publication to the SIP *reg* event package [79]. For these reasons, a single DHT overlay (IMS DHT) composed of the combined enhanced HSS with I/S-CSCF and presence server nodes is the best of the two presented possible alternatives for the conventional IMS.

[10] Patent Granted in the United States Of America: "DHT-BASED CORE IMS NETWORK", Patent number: 7796990.

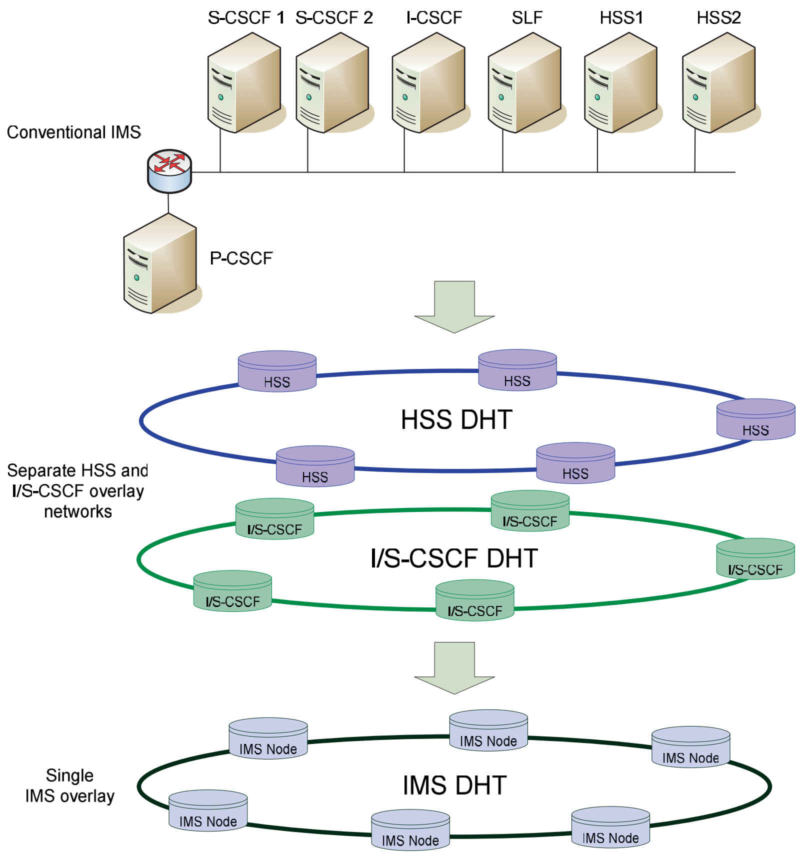


**Figure 8. Towards a scalable, self-configuring IMS network**

The comparison of the proposed solution with the conventional IMS core network shows that IMS DHT requires a similar message overhead in small networks and a slightly higher number of messages for reasonably large DHT networks. This is the price that needs to be paid for having a self-organized overlay network of servers. On the other hand, the additional messages in Distributed IMS are sent over a fixed transmission network that most likely is a fiber based network. Therefore, the additional messaging is a minor or even insignificant cost item. The robustness and self-organization properties of DHTs are able to reduce the OPEX in certain scenarios justifying the slight increase of messages. The paper also proposes optimizations to the proposed architecture that reduce message overhead in Distributed IMS and makes its operation more efficient.

Distributed IMS is targeted to operators and service providers that are looking for means to reduce their CAPEX and OPEX, effectively dealing with node failures while at the same time keep interoperability with operators that implemented the conventional 3G IMS infrastructure. Due to the fact that the DHTs are able to react themselves to changes, for example, a node comes up or a node fails, no human intervention is required. Bringing a new S-CSCF or HSS into operation requires almost zero configuration. This offers saving in OPEX. Moreover, distribution of the network functional elements in a DHT fashion leads to increased robustness of IMS network elements. However, the self-configuration of the system comes with the cost of additional overhead for the realization of self organization and replication mechanisms.

### 4.2.2 P2P Overlay on top of IMS

This section presents the second implementation scenario in 3G mobile networks. [P3] describes an architecture that allows operating a P2P overlay on top of IMS. The goal of this work was to design a P2P overlay that works on top of the conventional IMS and provides a minimum impact on underlying IMS networks. This was achieved by providing the SIP P2P Application Servers (SIP P2P ASes) as the sole network entities that have to be brought into operation for creating the overlay network.

According to Figure 9, one or more SIP P2P ASes are brought into operation in an administrative IMS domain. These nodes behave as regular SIP application servers from the point of view of IMS. Additionally, SIP P2P ASes behave as peers (or super peers according to the P2P terminology) in the P2P overlay. There can be one or more SIP P2P ASes in one administrative IMS domain.

SIP P2P ASes interact with other SIP P2P ASes located in the same or different administrative domain using SIP signaling. Connections between operators or service providers can be set up based on agreements. SIP P2P ASes may also interact with non-SIP overlay networks using the gateway functionality presented in Figure 9. These ASes effectively create the P2P overlay network. A SIP P2P AS behaves as a front-end towards IMS terminals (e.g. IMS capable mobile phones), make the resources of the network available to other peers, help IMS terminals to get results from searches, and maintain a P2P overlay network for resource sharing. The architecture of the proposed system allows for different P2P network structures and P2P algorithms. For example, the overlay network can either implement an unstructured network based on flooding, structured networks based on distributed hash tables, or a combination of both. A service provided can determine what the best algorithm for distributing the information is.

The proposed architecture is a semi-centralized P2P architecture. IMS terminals act as clients (or ordinary peers according to the P2P technology). They connect to and mostly depend on peers (the SIP P2P ASes), which are always supplied by the home network operator. Users authorized to use the service are provisioned with an initial filter criteria in the S-CSCF that allows proper routing towards one of the SIP P2P AS nodes.

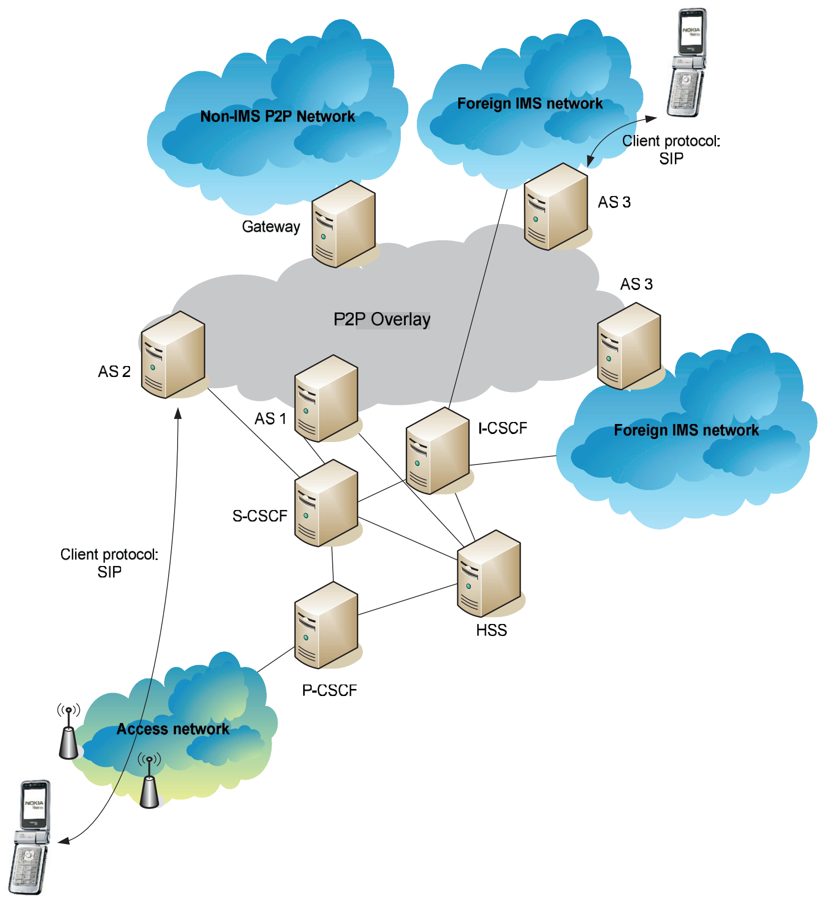


**Figure 9. P2P Overlay on top of IMS**

IMS terminals are responsible for publishing information about their shared resources and/or connectivity and presence information of a user. Then, other IMS terminals can perform searches of users and resources. The result of a search provides the user with a list of users and available resources and the SIP URIs where the resource or a user is located. Once the UE is aware of the SIP URI of the endpoint where the user or the resource is located, it can initiate a regular SIP session towards such a SIP endpoint and establish a VoIP session or fetch a picture file taken by the other user. This SIP session does not involve any SIP P2P AS, thus, it is a regular P2P session between two endpoints.

The architecture allows an operator to use the existing network nodes, such as (P-/I-/S-) CSCFs, HSS, etc., without any modification. Only small tuning of the configuration is required to provide the appropriate filter criteria to trigger the routing of certain SIP requests to the SIP P2P AS. However, this configuration is a normal IMS maintenance practice required anytime when a new service is taken into operation. The architecture may run both on top of conventional IMS as well as Distributed IMS networks that were presented in sections 3.2 and 4.2.1 respectively.

In many ways, it is beneficial to incorporate P2P in a controlled and standardized manner instead of only using an operator as a bit-pipe. The operator can implement any charging scheme for the usage of overlay services, including subscription based charging, pay per use, or a hybrid charging model. It can charge its users in the manner it wants. Furthermore, the number of commercial services that can be implemented over the overlay network is unlimited. IMS terminals, though, need to be upgraded with client applications that utilize the P2P overlay network.

The network architecture together with the signaling flow and the results of measurements are presented in [P3]. Our prototype shows that the device's capabilities (memory, CPU speed, etc.) or the signaling overhead do not impose any visible restrictions to the application in current mobile networks. This is mainly due to the system design that concentrates load to the SIP P2P ASes residing in the fixed network. A user may also specify the rate of notifications and which information about a resource he wishes to receive. Most of the delay is incurred on the transport layer of the wireless network and in search process in the P2P network. Table 2 presents delays incurred by publication and search processes.

**Table 2. Publication and search delay**

| Action | Request/Response | Delay (ms) |
|---|---|---|
| Publication | PUBLISH | 240 |
| | 200 OK | 150 |
| | **Sum** | **390** |
| Search | SUBSCRIBE | 260 |
| | Search delay[11] | 10 – 2000 |
| | NOTIFY | 340 |
| | **Sum** | **610-2600** |

We have also been working on an alternative solution to the one presented above that allows for publishing and searching for resources using the presence service. In [28] and [29] we describe how to use the presence service to locate resources stored in the overlay.

[11] The search delay represent a delay incurred in a file sharing network comprising of a few SIP P2P ASes in a single administrative domain.

### 4.2.3 P2P Overlay on top of a standalone SIP network

Some mobile operators decided not to deploy the IMS network because of its high cost and a questionable business model. They implemented a SIP based infrastructure instead. SIP networks allow for providing real-time services such as voice over IP, instant messaging, presence. However, end users and mobile operators may have an even greater interest in various non-real-time applications such as content sharing. [P2] presents an architecture that allows extending the current SIP infrastructure that was designed to provide real-time services with P2P resource sharing functionality.

The proposed architecture allows running a P2P overlay network on top of the standard SIP infrastructure. Figure 10 presents a P2P overlay that works on top of a SIP based network. There are two types of nodes in this architecture: SIP servers and mobile terminals. The SIP servers act as peers and are located in the fixed Internet. Mobile terminals act as clients. The goal was to concentrate load to the fixed parts of the P2P overlay to minimize the use of expensive radio resources, limit battery consumption as well as eliminate the barriers to joining the overlay for low end mobile devices.

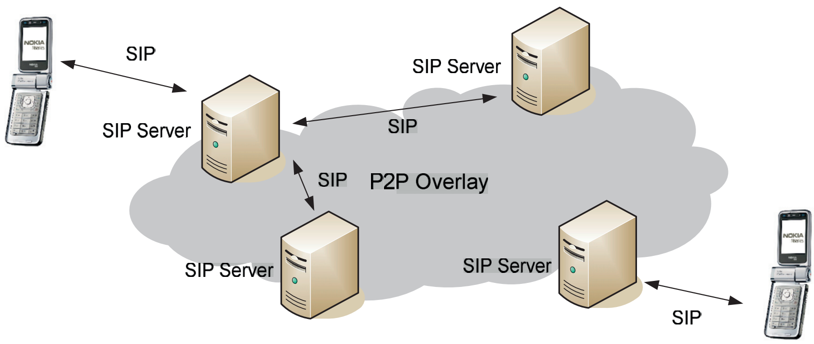


**Figure 10. P2P Overlay on top of a standalone SIP network**

The SIP servers can be maintained by a service provider or by users themselves. The P2P network between peers is a pure P2P network that operates in a fully distributed fashion. For improved scalability and reliability, P2P overlay may have several layers of P2P networks spanning the peers. All of the peers operating on a certain P2P network layer have an equal status. Peers interact with other peers using SIP signaling.

A client is an application running on a handset. Every client is connected to one or many peers. It may maintain a list of potential super-peers for bootstrap purposes. Alternatively, they may have a preconfigured address of a bootstrap server that stores all of the necessarily settings required to establish connectivity with the P2P network. Mobile terminals connect to the Internet using any type of access network such as WLAN or GPRS.

The client protocol between the client and the peer is the SIP protocol. In order to allow usage of SIP as a client protocol, we defined a new SIP 'resource' event package together with the syntax of an XML document that describes the generic resources available in the overlay. A resource can be, for example, files (e.g., images, video files, audio files), chat rooms, streaming content, printers, printer jobs, etc. This event package also allows SIP endpoints to subscribe to changes in the availability of resources, or perform searches for the availability and location of a given resource. Support for partial notifications and publications is also accomplished by using XML patch operations. In the proposed client protocol, the XML documents are carried in the body of SIP messages. The event package and XML syntax are presented in [23][8]. Based on the comments from the IETF community we also defined a new SIP event package and data format for describing files [66][9][30]. This event package and the syntax of an XML document are used for publication and searches for content files rather than more generic resources defined by the SIP 'resource ' event package.

A prototype implementation proves the feasibility of the presented architecture. The results of measurements show that the mobile phone capabilities such as memory and CPU, and a signaling overhead introduced by the system do not impose any visible restrictions on mobile phones running P2P applications.

The memory use of the client software varied between 200 and 350 kilobytes. The SIP stack and the SIP profile manager additionally consumed 170 kilobytes of memory. Altogether the application and SIP stack consumed half a megabyte of RAM at maximum. Based on the measurement results and subjective tests done by the project group, the search performance in the prototype system is more than adequate. When a user types in a search string, results are displayed almost immediately. In the test network the search process took around one second at maximum. However in more complex systems with many peers and millions of users the search process could be longer. The download signaling takes one or two seconds, but the actual content transfer takes time depending on the requested file size.

The implementation has a modular architecture, so that it is easy to implement a variety of multimedia services on top of the overlay. Detailed software architecture, prototype implementation and measurements are presented in [P2]. The format and semantics associated to a 'resource' event package that allows SIP endpoints to publish the availability of generic resources is presented in [23]. SIP framework used for advertising and searching for shared resources, such as services or files, within a given community is presented in [8][9].

### 4.2.4 Multipurpose P2P Overlay

The previous sections presented mobile P2P Overlay architectures that allow for providing non-real time services on top of real-time networks such as a standalone SIP network and the IP Multimedia Subsystem. However, content sharing networks have much higher customer penetration than VoIP networks in the Internet. Therefore, it might be more natural to extend content sharing networks with real-time functionality than vice versa. Moreover, in order to provide SIP based services the non-SIP service operator would have to deploy a SIP infrastructure such as SIP proxies, registrars and presence servers, voice mail servers, media servers, conference servers, session border controllers, media servers, policy servers, and network elements for QoS to name some. The SIP network infrastructure requires significant investments in the network servers and requires sophisticated network maintenance, engineering, and management techniques that impact OPEX and CAPEX. In addition, the SIP infrastructure does not scale well. Typically it does not support any self configuration or restoration mechanisms that allow for fast automatic reaction to changes in the network. Moreover, it requires integration and regression testing after every change.

We have been researching P2P overlays that could be suitable for both content sharing and real-time services. As it has been described in Section 3.1, SIP uses SIP URIs for locating the contact information of other users. In SIP networks, users have to publish any changes of their contact information in order to be reachable by other users. According to user studies, publication of SIP contact information has to be accomplished within 20 seconds and contact data retrieval must be accurate and done in no longer than 5-10 seconds. Moreover, the overlay must provide an adequate level of security.

One of the most popular P2P content sharing applications Bittorrent[12] , that uses the Kademlia DHT algorithm [67], does not allow for real-time data storage and retrieval. The problem is related to the design of the Bittorrent protocol that does not allow for fast and predictable contact information retrieval and does not provide adequate security.

Understanding these problems, we wrote a Peer-to-Peer Session Initiation Protocol (P2PSIP) security requirements draft [19], that describes security

[12] Bittorrent, online: http://www.bittorrent.org, last time accessed on 12.05.2013.

requirements in P2P overlay mechanisms standardized in IETF[13]. We also co-authored a P2PSIP Protocol Framework and Requirements draft [10], that specifies a P2P protocol framework, presenting design options and recommending specific P2P protocol functionalities. Finally, together with Salman Baset and Henning Schulzrinne from Columbia University we designed a multipurpose P2P Protocol (P2PP) that is suitable both for real-time communication and content sharing. The protocol is presented in [7].

We also designed a multipurpose P2P network architecture. Figure 11 presents a multipurpose P2P overlay. In this figure clients connect to the P2P network using XML RPC or P2PP protocols, whereas unmodified SIP devices use the standard SIP protocol to connect to peers that implement a SIP server.

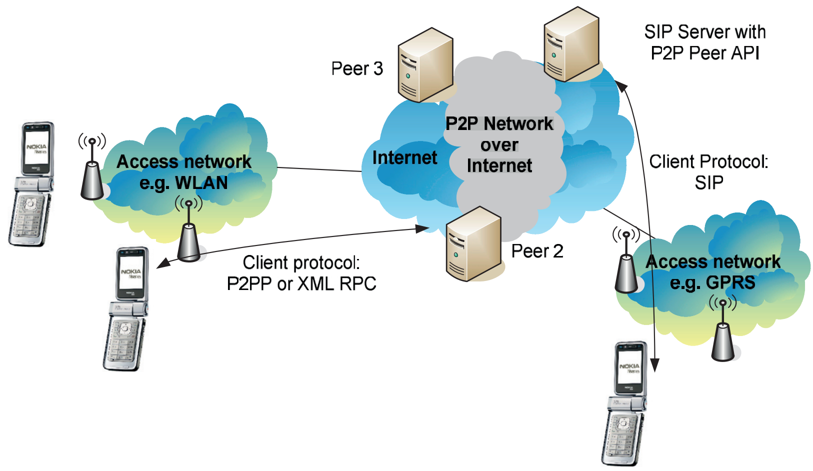


**Figure 11. A multipurpose P2P overlay**

In contrast to the architectures presented in the previous sections in the proposed architecture, the P2P clients communicate with the peers using a client protocol that supports generic methods for storing, updating, deleting, and retrieving data from the overlay.

RPCs are procedures called in a program on a client device, for example, a mobile terminal that goes over the network to an RPC server that implements the called procedures. The RPC server executes the procedures and sends the results back to the caller. The calling program then continues executing using data returned by the RPC server. This scheme allows less powerful machines to access high powered resources such as a distributed database of the P2P overlay network.

[13] Internet Engineering Task Force (IETF), online: www.ietf.org, last time accessed on 12.05.2013.

XML RPC that is a part of the web services framework, encodes the RPC requests into an XML data format and sends them over a standard HTTP connection to a RPC server. The server decodes the XML, executes the requested procedure, encodes the results in XML and sends them back over the HTTP connection to the client.

An alternative is to use a binary client protocol as it was suggested in [7] or binary XML. The main advantage of a binary protocol is that it can be optimized for the mobile interface.

In the proposed architecture the client protocol can also be a subset of the protocol running between peers, such as the P2PP protocol. This is an approach that we proposed in [7]. The advantage of designing a client protocol as a subset of the peer protocol is that a device that implements a single P2P protocol stack can act both as a client or as a peer. Depending on the situation, a device may change its role from a peer to a client and vice versa. We designed a mechanism that allows for a smooth transition between peer and client roles. The mechanism is presented in the patent application [12]. We also designed mechanisms for mobility support and monitoring in mobile P2P networks [26][27]. The main disadvantage of this approach is that the implementation of the protocol engine can be more time consuming compared to the implementation of the XML RPC based client protocol.

The multipurpose P2P overlay does not require peers to implement the SIP stack. Nodes store and retrieve their contact information from the overlay. There is no need for SIP servers. SIP is used for multimedia session establishment directly between mobile terminals without involvement of any peer node. The multipurpose P2P overlay may also store information about servers that assist clients and peers in NAT traversal such as Simple Traversal of UDP (User Datagram Protocol) through NATs (STUN)[14] or Traversal Using Relay NAT (TURN) servers. These servers may implement a full P2P protocol stack or use a client protocol to advertise their services.

A multipurpose P2P overlay may also support conventional SIP UAs (unmodified SIP devices) that are available on the market. In this option the P2P overlay would have to appear as a conventional SIP network to SIP UAs. In order to support conventional SIP UAs some peers in the multipurpose P2P overlay would have to implement a SIP server functionality which would include proxy, registrar and redirect server functionalities. The SIP server functionality would have to support a P2P interface that allows for accessing a generic distributed storage service of the P2P overlay. This solu-

[14] In the new specification RFC5389 STUN is not a NAT traversal solution by itself. It was specified as a tool to be used in the context of a NAT traversal solution. This was an important change from the previous version of this specification (RFC3489), which presented STUN as a complete solution.

tion does not require any changes to the existing SIP UAs, which can use SIP services provided by the overlay without even noticing its existence.

The research presented in [P7] shows a multipurpose overlay based on DHT that can be used as a replacement for conventional SIP networks. Our implementation used a Remote Procedure Call (RPC) protocol such as XML RPC and the P2PP protocol as the client protocols [14]. The presented P2PSIP application allows for locating users and establishing multimedia calls between mobile phones located in any place in the Internet and does not require any SIP servers. In our final implementation mobile phones may act both as peers and clients depending on the situation. The network can be constructed using only mobile phones without network infrastructure and our NAT traversal mechanism allows phones to establish connectivity in any mobile network. Our implementation was presented for the first time at CCNC 2007 conference and the final solution has been demonstrated publicly during Nokia World 2008 event.

The results of measurements of our P2P client software show that the registration delay that has two components: a relay registration delay and a AoR (Address of Record) insertion delay, is significantly higher in a 3G network than in a WLAN network. According to our measurements the average registration delay in the 3G network (excluding the peaks in the measurements that happened due to software bugs) was around 3 to 4 seconds and in the WLAN network was around 800 ms. Similarly the call setup delay including a contact information discovery delay and a signaling delay was around 7 seconds in the 3G network and 4 seconds in the WLAN network. Both delays were below the user acceptable thresholds in telephony namely 10 seconds for the call setup delay and 20 seconds for the registration delay. It is worth noticing that during our measurements the GSM circuit switched telephone call setup delay was around 6-7 seconds. Overhead traffic did not also impose any significant restrictions on the commercial implementation of a server-less mobile VoIP service.

Our measurements showed that the battery consumption of our P2PSIP client application is significantly better than the battery consumption of the mobile Bittorrent application discussed in Section 4.1.4. The results of the measurements shows that our P2PSIP application being in the client mode and running our NAT traversal mechanism drains the Nokia N95 mobile phone's battery in around 20 hours compared to only 5 hours in the case of the mobile Bittorent without any NAT traversal.

Our implementation allows a mobile phone to join the overlay as a peer. However, we have to remember that in large scale deployments (1 million overlay network discussed in Section 4.1.4) the phone battery may drain quickly. We recommend the peer mode for small ad-hoc or small private networks where traffic overhead is not much higher than in the client mode.

### 4.2.5 Combined P2P Overlay Architecture

It is possible to combine all of the presented architectures into a single P2P overlay network spanning mobile networks and the Internet. Figure 12 presents a P2P overlay that allows IMS terminals to access services provided by the P2P network over the Internet. The IMS terminals connect to the multipurpose P2P overlay network in the Internet using a gateway peer. As it is shown, peers provide community services such as content sharing inside communities, knowledge sharing and real time communication which will be described in Chapter 6. The next chapter presents a social layer that enables the P2P overlay layer with community features.

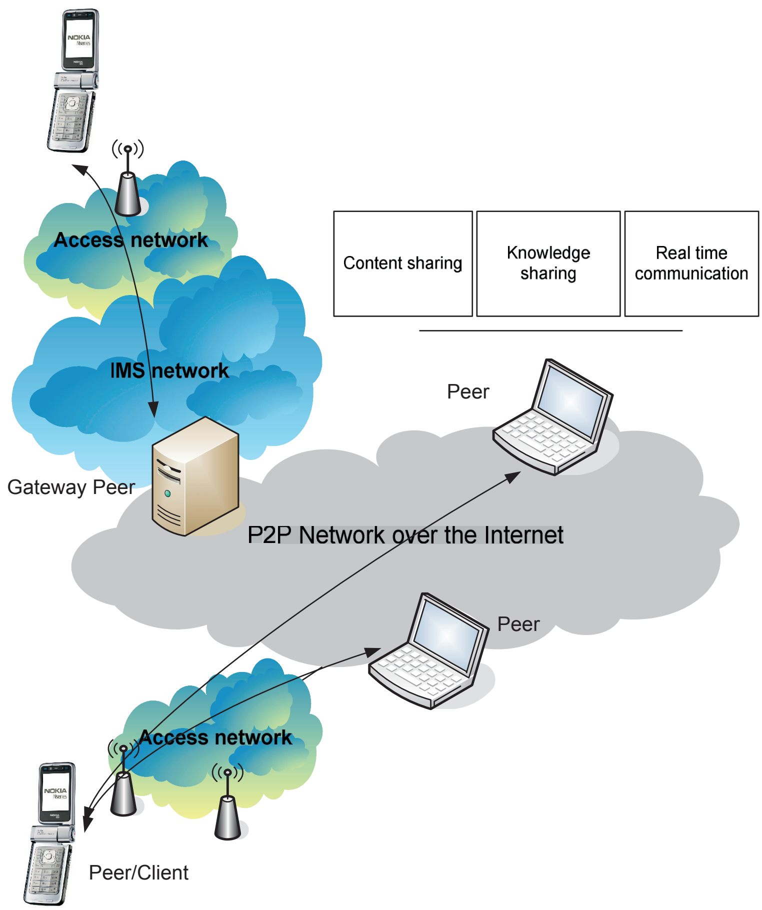


**Figure 12. Combined P2P Overlay Architecture**

## 4.3 Chapter Summary

In this chapter we have been discussing different aspects of community service provisioning on the P2P overlay layer. We presented different P2P network architectures and discussed their drawbacks and advantages. Our analysis and measurements showed that excessive traffic and battery consumption in the pure P2P architecture make deployment of large scale pure P2P mobile overlay networks impractical. Our study showed that semi-centralized P2P networks are the most suitable for the mobile environment.

We discussed mobile P2P architectures in the mobile operator domain and in the Internet. Two architectures in the mobile operator domain were presented, namely: Distributed IMS and the overlay on top of the IMS. Patented Distributed IMS technology places peers into the core IMS network elements making IP Multimedia Subsystem scalable and self-configurable. In contrast to the first architecture the second implementation does not require any changes to the mobile network infrastructure. It runs the P2P overlay on top of SIP P2P ASes, the standard IMS infrastructure elements. In our discussion on mobile P2P architectures in the operator domain we also mentioned a method for publishing and searching for resources using the presence service.

We discussed mobile P2P overlay architectures in the Internet. We presented the architecture that allows running a P2P overlay on the SIP infrastructure. We proposed new resource and file event packages and frameworks for publishing resources and files in the Internet. We also designed a new P2P protocol suitable both for fixed and mobile devices. The protocol can be used for running a multipurpose P2P overlay. In the overlay mobile phones may act as peers and clients and may switch between those roles smoothly.

All of the discussed architectures were successfully implemented and measured. The measurement results show that we managed to overcome problems with excessive battery consumption and excessive traffic also in networks where devices are behind NATs. We believe that the proposed architectures are suitable for commercial deployments. All of the presented architectures can co-exist forming a single overlay network.

# 5. Social layer

People live and communicate in groups, spending most of their lives interacting with other people. They share lives, emotions with people they love, care about, or simply work with. This aspect of human existence explains the popularity of person-to-person, person-to-group, group-to-person, and group-to-group communication services like voice calls, SMS, emails, instant messaging, chat rooms, and push-to-talk. These services are driving the mobile industry and will continue shaping it in the future. Most of the data exchanged in the mobile networks is highly positioned in the social context. The social network participants decide on their role in the network and how they like to interact with others. In this chapter we discuss mechanisms that allow enhancing P2P overlay networks with community management features.

## 5.1 The Social Layer

In [P4] we introduce the Social Layer that enhances the P2P overlay with mechanisms of social networking such as group formation, social search, and social communication. The social layer uses information about social connections between users and maps them to the P2P overlay layer. Social connections are modeled using the social network theory. The theory represents individual actors and their social relations using a social network abstraction. In the social networks we have nodes and ties. Nodes are individual actors, in our case mobile phone users or organizations, and ties reflect the relationships between the actors. The networks can have different shapes. The shape depends on actors and their social relations. An example social network with two communities (IMS experts and Foo fan club) is presented in Figure 13.

People maintain and expand their social networks when they meet other people, call them, or send them sms or e-mails. A trace of their interaction patterns is already available in mobile phones in the form of call, sms and email logs or geographical coordinates that can be used to create links between two people that were in the same place. The information about other people and groups is typically stored in contacts applications in mobile terminals. A contacts application stores business cards of other people. Business cards include different Uniform Resource Identifiers (URIs) like telephone numbers (e.g., E.164 numbers), SIP URIs, e-mail addresses, user identifiers in systems like Skype, Flickr, as well as employment information, and group membership.

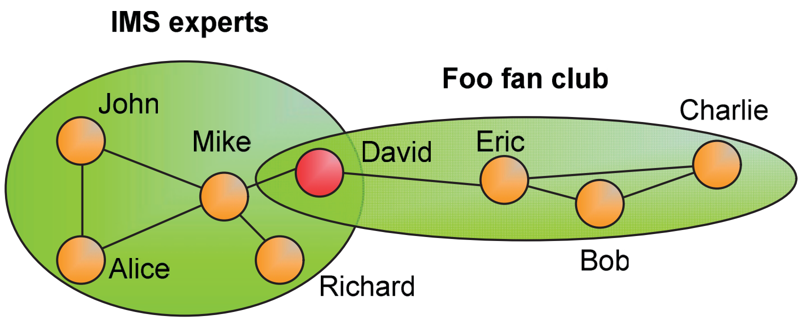


**Figure 13. An example social network**

For the purpose of coping with difficulties of maintaining the P2P network topology and for adapting the P2P network topology to social behavior of P2P network users [P5] proposes a P2P "requestor mediated" model. In this model each peer has a set of pointers to other peers which collectively define a potential P2P network topology. The pointers "virtually" connecting peers are defined by the list of "business cards" contained in contact applications in mobile terminals.

This method of topology creation possesses two main advantages. A person can be uniquely identified by some of the information contained in his or her business card. This information forms a universal address that is independent of the addressing mechanism of the underlying communication technologies. It also allows using a set of communication mechanisms to reach a particular person. It may also provide context information such as presence.

A mobile phone user can manually specify who is his friend and what are the social groups he belongs to, by assigning some attributes to the particular contact's profile in the Contacts application in the mobile device. For example, David may create a group called 'Foo fan club' and classify Eric as a member of this group. Similarly, he may create the 'IMS experts group' and add Mike to this group. The information about groups and group members may be used to facilitate search and information exchange mechanisms.

[P5] presents a routing mechanism in the social layer and an example data structure describing the social network of the overlay network participants and their knowledge including skills, experience as well as content files, web links, etc.

## 5.2 Maintaining social network connectivity

The effectiveness of the routing mechanism on the social layer depends on the social network connectivity. An accurate map of social relations is crucial for achieving high efficiency of the search process. The decision where to forward the search request and the search logic is based on information stored in the contacts application in mobile terminals. As our studies show, it is worth noting that information stored in the contacts applications contains, in many cases, partial contact data, so some important data might sometimes be missing. Additionally, the information stored in business cards may be not up-to-date.

Typically, users only store partial contact information, most likely, a telephone number. Contact information is in many cases exchanged on a piece of paper, for example, on a paper business card. The space reserved for the contact information in paper business cards is limited. Typically a telephone number and an email address are given.

The contact information is also changing from time to time. Users are changing email addresses, for instance, switching from a Gmail to Yahoo email account, changing a mobile or SIP services provider. Whenever an existing identifier changes, the user has to remember to inform known parties about the changes. Therefore, it can happen that finally not all known parties are informed. Moreover, people tend to live busy lives and every minute spent on boring updates is not appreciated. People tend to send SMS, e-mail with their new/updated contact information. The recipient of such information has to remember to update its list of contacts with new information. What happens if the contacts application stores data that is not up-to-date and a device owner urgently needs to contact another party? Moreover, as we have shown in [P5] users expect automatic collection and updates of data required for social networking.

In order to mitigate this problem, [22] proposes a user friendly mechanism for publishing, requesting, and updating information stored in business cards, such as URIs, names, addresses, photos of their contacts, across different User Agents such as mobile and fixed terminals. The mechanism can easily extend the functionality of Personal Information Management (PIM) software in cellular and fixed UEs. The mechanism requires minimum network support. It is transparent both in mobile and fixed networks and can be implemented as an IP Multimedia Subsystem (IMS) service. The following sections present two use cases of automated contact information update mechanisms with the use of the SIP protocol:

*Sending an updated job title and a phone number to other users (push operation)*

Let us take a fictitious user Bob. Bob will be changing his SIP URI in the near future, so he wants to update John with the new URI. In this case Bob's device sends a SIP MESSAGE request to John's device. The message includes a new job title and a new SIP URI, and also a date when these will be active. Upon reception of the SIP MESSAGE request John's device sends a 200 OK response message to Bob's device. John's device authenticates the sender via any of the existing SIP mechanisms for authentication, and authorizes the operation, for example, based on the fact that the sender has been authenticated as Bob, and that the sender's URI matches the URI currently stored in Bob's entry in the contact application running in John's device. When the requester is verified John's device updates Bob's entry in the contact application and sets a timer indicating when the information takes effect. The message flow is presented in Figure 14.

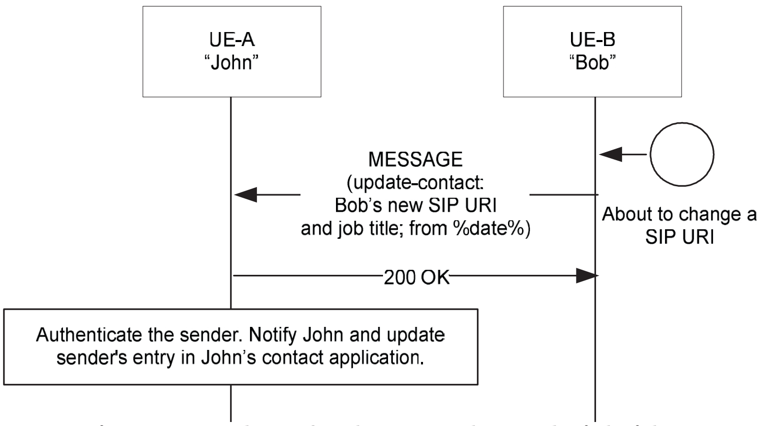


**Figure 14. Update of a phone number and a job title**

*Updating contact's profile (pull operation)*

John has the contacts of many friends listed in his phonebook. Some friends' profiles include only basic contact information like a phone number or a SIP URI. John decides to send an email to Bob asking him about a recently released video clip of his favorite rock band Foo. He opens the contacts manager application, selects Bob's contact, and chooses the 'send email' action. Unfortunately, the application informs John that Bob's email address is missing. It also asks John if he wants to automatically update Bob's profile. John approves the suggested operation. John's device sends a contact information request to Bob's device. The contact information request is realized using a SIP SUBSCRIBE request for the resource event package [8]. The SIP SUBSCRIBE request implements a fetch operation by setting the Expires header field to zero. The SIP SUBSCRIBE request contains a filter document to limit the contact information that John wants to receive. The syntax of the filter document is presented in [23]. Upon reception of the SIP SUBSCRIBE request, Bob's device sends a 200 OK response to John's device.

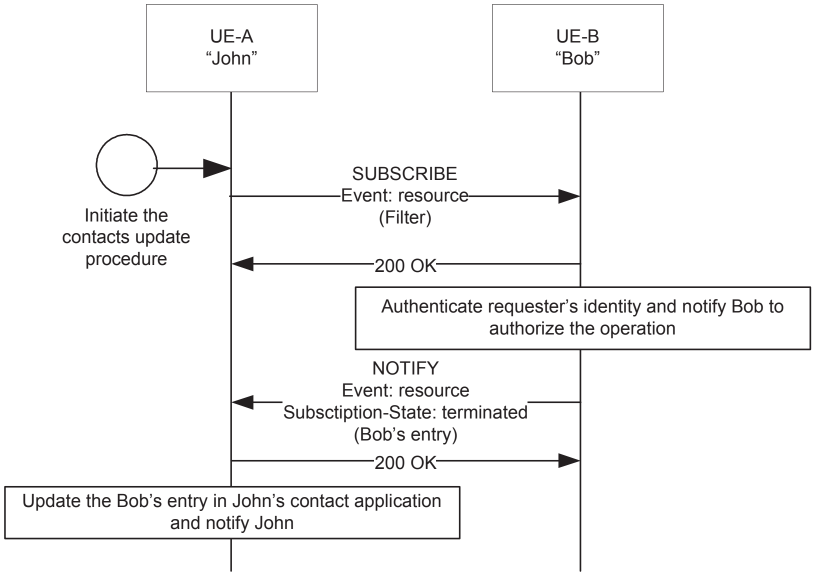


**Figure 15. The contacts update procedure**

Bob's device then authenticates John based on his identity. After that, Bob agrees to accept the request from John, and Bob's device sends his full or partial business card to John's device in a SIP NOTIFY request. The SIP NOTIFY request contains an XML document formatted according to the resource event package data format [23]. The Subscription-State header field in the SIP NOTIFY request is set to "terminated" indicating that no subscription is installed. Upon reception of the SIP NOTIFY request, John's device sends a 200 OK response and updates Bob's entry in the contact application. The message flow is presented in Figure 15.

Other data required for creating and maintaining the social network can be extracted using an automatic environmental data collection mechanism, for example, using Bluetooth technology, scanning user correspondence, etc.

## 5.3 Social DHT architecture

Even though it is possible to search for resources in the social network, the problem of mapping social ties to physical connections between nodes still persists. The Social DHT architecture introduced in [P4] separates the social layer from the P2P overlay layer. Figure 16 presents an example social layer and underlying P2P overlay layer as well as the mapping between these layers.

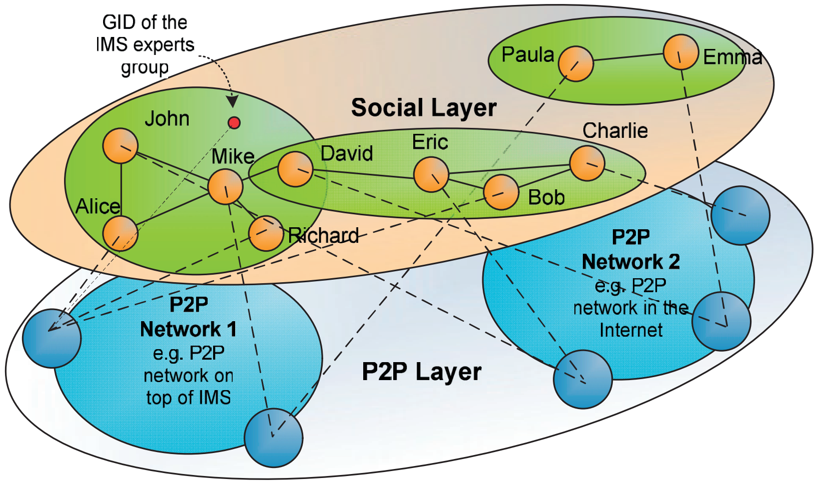


**Figure 16. An example social network and mapping to the P2P overlay network**

The P2P overlay layer takes care of locating a particular user in the P2P overlay and stores information about the social network and users' knowledge, whereas the social network layer takes advantage of the social context of user communication. Even though the communication is realized on the P2P overlay layer, it follows the user's social ties represented in the social layer.

## 5.4 Security

Privacy is an important aspect of communication in socially enhanced P2P networks. People do not want to make their private information such as photos, videos or contact information publicly available [P1]. They want to make them available to and accessible only by certain people. They also like to communicate securely. Since storage and information forwarding in P2P networks is distributed, without security enhancements it is very difficult to assure that data stored in P2P networks or communication between two peers is secured, for example do not pass through a malicious peer that modifies the forwarded messages or listens to the conversation. Well known P2P content sharing applications such as Gnutella and Kazaa assume that content shared in the P2P network can be accessed by every participant of the P2P network. This feature is adequate when a user shares public domain content, however, is not acceptable in socially enhanced mobile P2P networks, where users are taking pictures, shooting video clips, and creating content that they want to share with a limited group of users as has been shown in Section 2.2.

In the conventional mobile networks, servers are operated by service providers. If a user trusts a service provider, he also trusts servers the service provider maintains. In P2P overlay networks, peers can be operated by users themselves. In large P2P networks it is very difficult to trust all users, especially if we do not know them personally.

As it will be shown in Section 6.1 a single message can be forwarded by many peers before it reaches a destination. In such a setup a badly behaving peer may:

- listen to communication between peers
- modify incoming messages,
- discard incoming messages (the peer can discard requests and responses it is supposed to forward),
- generate incorrect responses to requests that are directed to some other nodes.

The second attack allows the peer to cause the overlay to store unauthorized or outdated information or return corrupted data. A peer may change the data record in the overlay by changing incoming publish requests (sometimes called PUT requests) or modify the result of a lookup operation by modifying incoming lookup responses. With this type of attack, the integrity of the P2P system can become compromised.

The third attack allows a malicious peer to prevent access to data stored in the overlay. This type of attack can degrade the performance of the P2P system making it useless from the end-user perspective. This is of high importance in P2PSIP overlays that store user's reachability data which is much more time-critical than content stored in file sharing networks.

The attack described in the last point above may lead to a requestor receiving corrupted data, for instance, connectivity information that points to some other node. This may happen if a malicious peer can respond to incoming requests that are directed to another peer.

A malicious peer may also launch other attacks towards the P2P overlay. If peers in the DHT based P2P overlay can freely choose peer identifiers or/and easily modify previously selected peer identifiers, an attacker may use join-leave attacks to place a malicious peer intentionally at any location in the overlay. Placing the peer at any location allows the attacker to obtain control of the location in the overlay where the attacked user or resource is stored. A malicious peer may discard, modify the data it is supposed to store and may discard lookup requests or reply with incorrect entries to the incoming requests.

An attacker may also try to publish a large number of resources to the P2P overlay increasing the processing load on peers that are responsible for storing the resources and limiting the overall capacity of the P2P overlay network. It may try to register all popular names preventing the name holders from registering their preferred URIs.

The security threats described above and other security attacks that are presented in [19] indicate that security is an important issue in the P2P overlay networks.

As we have shown in [7] many of the mentioned security attacks can be mitigated by applying security mechanisms such as an enrolment process, certificate based authentication and integrity protection as well as secure transport protocols. P2PP protocol, presented in [7], provides a hop-by-hop security model. The nodes in P2PP establish an unreliable or reliable secure channel with other nodes using DTLS or TLS protocols. P2PP uses an Enrollment and Authentication (E&A) server to enroll new users in the overlay and to authenticate existing users. Once a user is authenticated, E&A server issues a certificate that binds the user-identifier that uniquely identifies the user in the overlay with the public key of the user. Further, it also generates a peer or a client identifier and adds it to the certificate. On the other hand, P2PP allows nodes to operate without a central enrollment and authentication server. In such a scenario, nodes use self-signed certificates and generate their own identifiers.

In [20] we presented a P2P security mechanism that uses ZRTP protocol for distributed authentication of peers and secure session establishment. ZRTP is a solution for real-time peer-to-peer communication security problems, relying on Diffie-Hellman key exchange and Short Authentication Strings (SAS). This protocol does not require trusted third parties to be involved in the key exchange. Protection against Person-in-the-Middle attacks is done with short authentication strings and key continuity. SAS is a user-readable representation of a shared session key. Secure Real-time Transport Protocol (SRTP) keys are derived from that session key. Once ZRTP negotiation is complete, short authentication strings are generated from the session key and shown to the users. These strings are compared with each other by reading them out loud through the open voice connection. After the strings are verified, session secrets are stored and used as a part of the following sessions. The verbal verification of strings must be done once for every user's client that supports ZRTP.

Moreover, the architecture proposed in this chapter allows for mitigating the security threats by organizing data stored in the overlay and communication between nodes around trusted communities. Communication in this model happens between users who trust each other.

## 5.5 Flexibility in overlay network formation

The architecture proposed in Section 5.3 allows interconnecting many overlay networks using information on the social layer. It allows each community to run its own overlay and freely decide on the used P2P algorithms. For example, P2P network 1 in Figure 14 can be an unstructured P2P network, whereas P2P network 2 can be based on a DHT algorithm. The architecture also allows for splitting a single P2P overlay network, such as Mainline Bittorent 1 million nodes DHT, into a number of smaller overlay networks maintained by respective communities or allow smaller overlay networks to interconnect with large overlay networks. This allows mobile phones to participate in P2P overlays as peers because the overlay traffic can be limited by limiting the size of the overlay network. Moreover, the implementation of security mechanisms can be less difficult in smaller community maintained overlays since each community governs its own data.

As it has been presented in [P2][P3][P4], the social layer and its supporting mechanisms can be implemented on top of P2P overlay networks in the Internet or in the operator controlled environment such as IMS.

## 5.6 Chapter Summary

This chapter discusses the social layer of the community services provisioning framework. It presents the Social DHT architecture, a two-layer architecture that provides flexibility in resource representation and resource discovery. The combination of a structured overlay network and the social layer allows overcoming the limitation of the DHT algorithms such as the requirement of an exact naming and at the same time, enjoying the efficient location of resources in DHT networks. We also discussed mechanisms for maintaining social network connectivity that are critical for consistent routing on the social layer. Those mechanisms allow for automatic updates of contact information and distribution of contact updates between mobile phones taking part in the overlay. We also discussed security and privacy aspects of community service provisioning in mobile P2P networks. We listed potential security threats and mentioned security mechanisms that may mitigate many of the possible threats.

# 6. Community services layer

The previous chapters presented P2P overlay networks as well as architectures and methods that allow for formation of mobile communities. In this chapter we will present three example services that can be provided on top of mobile P2P overlay networks that support mobile communities.

## 6.1 Content sharing inside communities

In [P2], [P3], and [P4] example content sharing services are presented that allow mobile phone users to share content with selected people inside communities. All of the services were designed to fulfill user needs discovered in [P1] and shortly discussed in Section 2.2.

In the content sharing inside community service presented in [P4], the service logic is executed on peers residing in the fixed network. These nodes store data published by the users such as contact information, content metadata, access rights, and the list of groups. The P2P layer consists of one or many P2P overlay networks formed by peers.

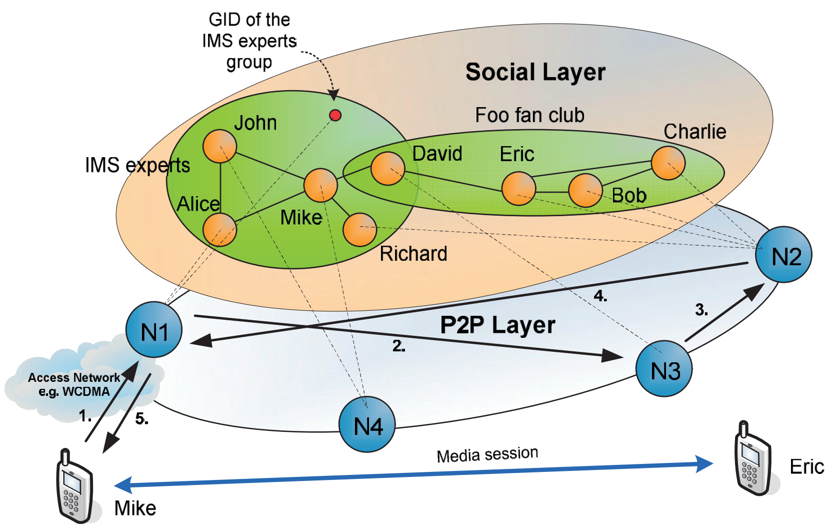


**Figure 17. Content sharing inside communities: the search process**

Figure 17 presents an example content search process in an overlay network with four peers N1, N2, N3, and N4. The figure shows the representation of two social groups: the 'IMS experts' group and the 'Foo fan club'. In principle, these groups do not share many commonalities. However, there is a key participant, David, who is a member of both groups. The 'IMS experts' group is a formal group whereas the 'Foo fan club' group is a virtual one. The main difference between these two types of groups is that a formal group requires specific enrolment of the user into the group. A formal group is identified by a group ID that is registered in the P2P overlay and keeps track of its members.

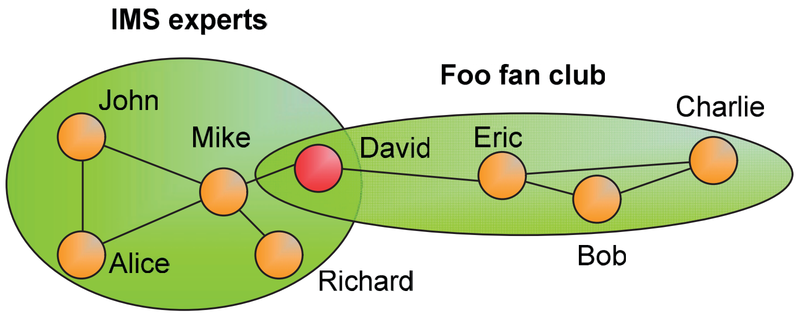


**Figure 18. Two social networks**

From David's point of view, his 'Foo fan club' group is just composed of Eric and himself. From Eric's point of view, his 'Foo fan club' includes David, Bob, and Charlie. This information is stored in Eric's and David's phonebooks. Each phonebook entry has a new field: 'Belongs to', which lists the groups Eric or David belong to. The value in this 'Belongs to' field can either indicate that a group is either a formal or virtual group.

Let us assume now that Mike wants to obtain some unique resource, such as a video clip from the Foo concert. Let us assume that the content Mike is looking for is stored in Eric's device. Mike and Eric are not known to each other; therefore, Mike is not aware of Eric's content. Furthermore, Mike is not part of the virtual 'Foo fan club'.

Mike has social ties with David. Since David is a member of both 'IMS experts' and 'Foo fan club' groups, David is in the best position to know more Foo fans than Mike knows. Therefore, Mike sends a search request addressed to David to the Peer 1 (N1) that acts as a front-end agent towards all of the messages originating from Mike's device (1). N1 consults its P2P routing table and forwards the message to Peer 3 (N3) that is responsible for storing David's social information in the form of business cards and information about resources he wants to share with others (2).

If David does not possess the resource that Mike is trying to find or Mike requests all the resources fitting his resource description, N3 may forward the request to Eric. Actually N3 forwards the request to Peer 2 (N2) that is responsible for storing Eric's records (3). Since Eric knows David, and David knows Mike, the chances that Eric allows Mike to download the video clip can be much higher. N2 can forward the request to the other members of the group that have social ties with Eric, namely, Bob and Charlie. If any of them have the searched resource, they can authorize Mike to download it. N2 sends the search response to N1 (4) that forwards the response to Mike's device (5). Alternatively N2 may send the search response directly to Mike's device.

Let us consider a second example (which is not visualized in Figure 17), in which Eric is searching for an IMS white paper, whose filename he has learnt through some web page. Eric forwards the request to David, knowing that he is an IMS expert and he might have better chances to get the content. David receives the request. Since the 'IMS experts' group is a formal group, David can forward the request to Peer 1 addressing the request to the Group IDentification (GID) of the IMS experts group. As a result, Eric will be provided with the list of users who store the searched report. After this, Eric can contact directly any of these users, providing that they are online at that time, and download the report.

The implementations of P2P content sharing systems prove the feasibility of the proposed P2P content sharing inside communities service. The presented mechanisms allow users to share content with only certain people and social groups. A user is able to join different communities and define what content is available to everyone and what content require social filtering.

## 6.2 Knowledge sharing inside communities

Even though the advantages of knowledge sharing are clear, getting employees to change their ways and start sharing knowledge with others is not an easy task. As we have shown in [P5], knowledge sharing is not only about the system. There are also many challenges related to human behavior. The system must be designed to support employees' work and their social behavior. The system has to adapt to employees' social interactions. What constitutes a good system for one profession or employee does not necessarily have to be suitable for another one.

Understanding the barriers to knowledge sharing and the effectiveness of the incentives models, we have designed and implemented a P2P knowledge sharing service that is presented in [P5]. The P2P knowledge service is dedicated to mobile platforms such as mobile phones. The service is inspired by the manner in which people interact in everyday life. It is based on the concept of social networking, allowing for the formation of communities that are based on trust.

The topology of the P2P overlay in the P2P knowledge system is based on the social links between system participants. The social layer of a P2P system utilizes information stored in business cards and relational references in the distributed knowledge database to route the messages in the system. This makes the knowledge search process simple and efficient.

Knowledge of P2P network users is represented in a knowledge database. The knowledge database is arranged into a hierarchical structure based on user-defined relations between objects and relational references to the databases of other peers. The objects in the knowledge database represent particular knowledge a user has. In the implemented system the user's knowledge database is represented as a knowledge tree.

The root of the knowledge tree is represented by:

- a personal business card that contains Uniform Resource Identifiers (URIs) that are used as network identifiers,
- a list of business cards (list of contacts), where each contact business card is a valid (can be used for identification in the network) subset of the business card structure, extended with a general trust rating and a cooperativeness score of the peer. The cooperativeness score is a dynamic value, which is automatically updated over the time based on the results of the previous search requests.

An example of the distributed knowledge database located on three peers representing three 'Foo fan club' community members from Figure 18, with a number of relational references (doted arrows) is illustrated in Figure 19.

Altogether, the knowledge trees of the community peers altogether form the distributed knowledge database of the community. Mapping of the individual knowledge trees is performed using a common thesaurus. The thesaurus provides the definition and a number of possible names (e.g. football and soccer) for each commonly used object. The thesaurus can be limited to a given group. Every group may have a different thesaurus.

The relational references are used as a direct guidance in the search process (as they can be interpreted as a recommendation to ask for more information from a given peer). They also help in maintaining common understanding (or mapping) of the objects in the P2P community. It is important to note that we talk about common object interpretation as a result of the hierarchical data organization, when a person can easily and quickly evaluate if the proposed object interpretation is matching to his/her understanding. Moreover, here we use the fact that people who know each other and share common interests, most likely use the relational references for the objects belonging to the areas of common interest.

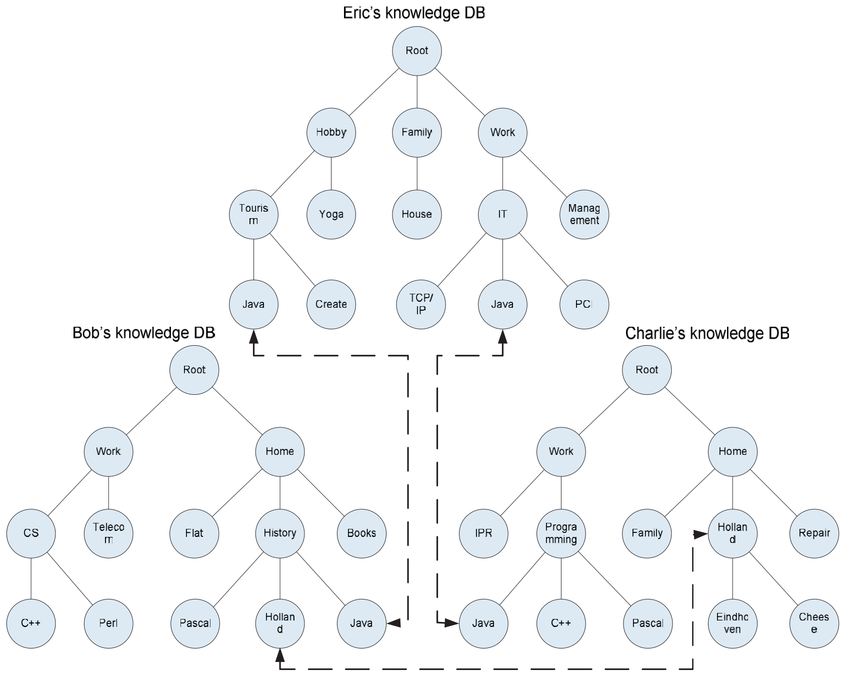


**Figure 19. Example of the knowledge databases with hierarchical structure and relational references defined for three community members**

In the presented example in Eric's knowledge tree there is an object called Java in two tree branches. One branch indicates that Java is an island in Indonesia and the other that Java is a programming language. The relational references indicate that more information about the island Java should be asked from Bob and that Charlie is a person that may know more about the programming language.

We have also designed a user friendly User Interface that allows users of the P2P knowledge system to easily navigate through the distributed knowledge network. The system renders all of the information so that it fits nicely into the small screens of mobile phones. The users may access resources from the distributed knowledge database and the recommended information from the world wide web. They may also contact other P2P network participants by establishing phone calls and by using messaging services. The system allows users to make calls using a circuit switched technology, IMS or SIP voice call as well as the P2P real-time communication service that will be discussed in Section 6.3.

Using the implemented system, mobile phone users are able to search for knowledge in their trusted communities. The information exchange is realized between people who trust each other and at the same time share background knowledge.

The test users of the prototype system indicated that the main issue in the implementation presented in [P5] is manual data input. Typically people are busy and do not want to input all of the information to their knowledge databases. They expect that software will do most of this work for them. We also realized that in reality users of our system would not create very complex knowledge trees.

Understanding these drawbacks we have been working on automatic data collection. [22] shows a supporting service that is used to keep contact information up-to-date.

## 6.3 Real-time communication inside communities

In [P7] we present a mobile P2PSIP VoIP service. The presented service and its implementation allow mobile phone users to set up real-time multimedia sessions, in particular VoIP calls, using resources of the P2P overlay with no or minimal assistance of centralized servers. The system distributes the SIP servers using the P2P technology. The paper also presents a method for locating relays that are used for NAT traversal. The service complements content sharing and especially knowledge sharing services that use the P2PSIP real-time communication service for establishing IM sessions or voice calls between community members involved in knowledge or content sharing. Let us here take a look at the following three operations.

*Relay location procedure*

Relays publish their contact information in the P2P overlay, for instance, a relay in Finland publishes information that it is located in Finland and is reachable under a certain IP address and port. A mobile endpoint behind a NAT that requires media or/and signaling relaying first searches the P2P overlay in order to retrieve an address of the relay. When the mobile endpoint retrieves the address of the relay, it connects to the relay and requests a public IP address and port pair. The relay reserves a port for the node and responds with its own IP address and the reserved port to the mobile endpoint. Consequently, the relay forwards all messages sent to this IP and port to the mobile endpoint. The relay may reserve separate IP addresses and ports for signaling and media traffic.

*Publication of own connectivity information towards other network nodes*

The mobile endpoints publish their connectivity information in the overlay instead of registering in a SIP Server in conventional SIP networks. The connectivity information of a mobile endpoint is stored in a user record that belongs to the user of the endpoint. The user records may include, among other information, business cards of social network participants and information about content a user wants to share with others. The user records are indexed using URIs belonging to the user. The publication guarantees that other nodes are able to find the user's current contact information in the overlay and contact the user even if his device changes its IP address.

*Session establishment*

When setting up a session, a mobile phone retrieves the contact information of the callee from the overlay. The contact information of the callee can be retrieved from user or group records stored in the overlay. If the contact information points to an IP address and a port of a relay, the caller sends SIP requests to the relay. The relay forwards all messages to the callee. Otherwise, the SIP request is sent directly to the callee's device.

In the implementation presented in [P7], peers and clients publish, retrieve and maintain the overlay network using P2P protocols (a client and a peer protocol) while multimedia sessions are established between endpoints using the Session Initiation Protocol that is widely implemented in mobile devices.

## 6.4 Chapter Summary

In this Chapter we presented three example services that can be provided using the socially enhanced mobile P2P infrastructure presented in Chapters 4 and 5. The first of the presented services was the content sharing in mobile communities service. The presented service allows mobile phone users to share content such as pictures, videos, or text files with other social group members. It mitigates the free riding issues discovered in our business studies by allowing for flexible group formation and group management. The service allows for efficient location of any resource in the P2P overlay network.

The second presented service was the knowledge sharing service that allows mobile phone users to efficiently search for knowledge in their social networks. The service represents a novel method for communicating between mobile phone users and exchanging knowledge. It allows a mobile user to contact (voice call, IM or SMS) and share information with a person that has certain knowledge about a particular subject rather than calling a contact from the contacts application contained in a mobile phone. It is an alternative communication scenario to the one that is widely used nowadays. It is based on the desire to solve a certain issue and the observation that people listed in the phonebook of our mobile phones cannot necessarily solve the problem in hand.

We also presented a real-time communication service that allows for establishing multimedia calls between mobile phone users. We showed how to traverse NATs in the distributed mobile environment and distribute connectivity related information required for multimedia communication. The service is a cost-efficient alternative to the conventional client server based real-time multimedia services supported by IMS and conventional SIP networks.

# 7. Conclusions

At the time of starting my research journey in 2003 there were only a few publications about P2P technologies in mobile networks. The published papers concentrated on concepts rather than on applications and systems that can be deployed in cellular networks on wide scale for millions of users. My research goal was to design practical innovations. To be practical, innovations need to go thru scientific research including theory development, user studies and verification prototyping. After this and in addition to this, also, a lot of engineering type of work is needed, before new innovations are adopted.

My research work was carried out on two levels: a technical and business dimension. I started my research work analyzing technical and business constraints in order to better understand barriers of deploying mobile P2P technologies. These results were crucial for the next steps of my research work namely: scenario planning, technology development and application deployment.

*Technical and business constraints*

On the business dimension I analyzed mobile P2P ecosystem from the viewpoint of consumers, business users and mobile operators. In order to analyze user needs for P2P services in the mobile environment, I conducted a user survey and a literature study.

The user survey conducted among potential consumers of mobile P2P content sharing systems revealed that mobile content produced by family members and close friends such as pictures, video clips and text files, is as popular as, or even more popular than, content created by the content industry. On the other hand, there is a strong free riding attitude towards sharing content with unknown people. Concerning willingness to share personal content, respondents preferred to share their personal content with their family and close friends rather than with other social groups. There was a clear interest in sharing both private and professionally created content on mobile phones but at the same time there was a strong need for group management features in mobile P2P content sharing applications. On average the results showed that groups of family and closest friends seem to be quite small in size. This suggested that P2P content sharing services have to support quite a fine granularity and flexibility in group formation.

Similarly, the study of business users showed that people are more likely to share their information with members of previously established social communities. When people get to know each other they are able to predict what to expect and how the other party will behave in certain situations. When trust is strong, people are more willing to openly communicate and exchange knowledge without fear of "loosing face" or that someone will intentionally ridicule them in public. The study showed that employees should not be forced to use knowledge sharing systems. The information sharing systems must be designed to support their work and their social behavior. What constitutes a good system for one profession or employee does not necessarily have to be suitable to another one. This had implications both for knowledge sharing as well as usage of knowledge management systems as a source of information. We presented possible incentive models that can be applied in the corporate environment in order to stimulate knowledge sharing. We showed that financial incentives achieve faster short-term results and may be useful to get a knowledge management project started, whereas soft incentives and community based knowledge sharing that is based on trust have more of a long term positive effect.

From the mobile operators position the future of P2P did not look so bright. On the one hand many experts expected flat rate data plans to be a dominant charging model in the future. Because of the lack of differentiation in the traditional mobile services market and flat rate charging models, mobile operators were increasingly looking for cost efficient solutions. This might have suggested that there was room for P2P technologies because they may improve scalability, self-configuration and automatic reaction to changes reducing capital and operational expenditures (CAPEX and OPEX) for service and network providers. On the other hand, mobile operators were concerned that P2P may significantly increase traffic in their networks forcing them to invest in the upgrade of their infrastructure especially expensive wireless infrastructure.

From the technology point of view I faced many constraints such as power consumption, NAT and firewall traversal, traffic overhead, usability, intermittent connectivity, and security. All of those constraints had to be addressed in order to make the deployment of wide scale mobile P2P systems feasible.

In order to better understand power consumption constraints I conducted research on battery consumption and traffic overhead of P2P applications from the mobile phone perspective. At that time, battery consumption was an important but not very widely studied aspect of P2P applications especially in mobile networks. We developed an analytical model of Kademlia DHT performance and used it to analyze a number of different strategies to use peer-to-peer with mobile devices especially for real-time applications. Our analytical model and the result of battery measurements in the Mainline BitTorrent DHT network have allowed us to investigate different possible deployment scenarios for P2PSIP. These analyses included studying a model where all mobile devices participate in the DHT and contrasting that with two other models: where only a subset of mobile devices forms the DHT and a model where the DHT is formed by powerful servers.

Our analysis showed that if we look at the energy consumption, it turns out that we have an interesting trade-off. To minimize the long-term cumulative energy consumption, we would like to have as little traffic as possible within the DHT. This can be achieved if the minimum number of nodes participates in the DHT. However, the less nodes participate in the DHT, the higher the traffic and energy per node becomes, which results in exceeded capacity in large scale mobile P2P networks.

Therefore according to our energy consumption and traffic analysis the optimal configuration for large scale mobile P2P deployments was the semi-centralized architecture with powerful servers running peer software and mobile phones acting as clients.

*Scenario planning*

In order to define the most probable deployment scenarios for mobile P2P services we designed a scenario planning methodology that proposes the Schoemaker's variant of scenario planning as a suitable method for evaluating emerging mobile services. Using the proposed methodology and knowledge about possible business and technical barriers such as NAT and firewall traversal, intermittent connectivity, security, battery consumption as well as user needs we developed consistent and coherent learning scenarios, which depict a wide range of possible scenarios for emerging mobile P2P services.

According to our analysis, the small-scale ad-hoc and private environments without interconnectivity to the Internet may benefit from P2PSIP technology. Revenue collection in those settings is possible by bundling devices and services in ad-hoc environments, and by charging installation and support fees in private environments.

The operators looking for global service provisioning should consider a public global semi-centralized scenario. In this scenario the service operator would gain its competitive advantage either by cost advantage over established network operators or by offering new features to users. The revenue model in this scenario could be transactional income from value-added transactions such as interconnections to other networks, or advertising. A subscription based revenue model is unlikely as consumers are accustomed to VoIP, IM and presence services provided for free.

The results of the business research formed an input to the research on the technical level. In my studies I concentrated on the public global semi-centralized scenario but I was keeping in mind that P2P mobile technologies can be used for the deployment of small-scale P2P systems.

*Technology research & development*

On the technical level I have researched P2P system architectures, protocols, and algorithms that enable the provisioning of community services in the mobile environment. The different technical aspects of community service provisioning were presented using a community services provisioning framework. The framework divides the provisioning of community services into four layers, namely: Network Infrastructure, P2P Overlay, Social, and Community Services. All of the layers were discussed separately. All of the systems and technologies were designed to overcome the constraints of mobile phone networks and devices such as: intermittent connectivity, firewall, NAT and battery constraints.

Having results of my analysis of business and technical constraints I developed various P2P overlay network architectures and protocols that are suitable for the mobile environment. I started the presentation of the developed architectures with two implementation scenarios that relay on 3-rd Generation (3G) mobile infrastructure. The first implementation scenario called the Distributed IMS architecture required only changes in the network infrastructure but did not require any modifications in mobile terminals themselves. The Distributed IMS architecture replaces the IMS core network with one or a collection of self-organizing DHT overlay networks. The idea was based on the fact that DHTs are known to scale very well to a very large number of nodes and allow for creating self-configuring networks with automatic replication capabilities. This work focused on distributing the IMS functional elements, such as S-CSCF and I-CSCF, as well as HSS, and up to some degree, presence servers. The work proposed three different configurations of distributed IMS: distributed HSS (HSS DHT), distributed S- and I-CSCFs (I/S-CSCF DHT), and combination of HSS, I/S-CSCF and presence server elements in a single P2P overlay (IMS DHT).

The architecture of the second implementation scenario allows operating a P2P overlay on top of IMS. The goal of this work was to design a P2P overlay that works on top of the conventional IMS and provides a minimum impact on underlying IMS networks. This was achieved by providing the SIP P2P Application Servers (SIP P2P ASes) as the sole network entities that have to be brought into operation for creating the overlay network. I have also been working on an alternative solution to the one presented above that allows for publishing and searching for resources using the presence service. In my supporting publications [28] and [29] I described how to use the presence service to locate resources stored in the overlay.

Finally I presented two implementation scenarios that does not relay on the mobile infrastructure. They assume that mobile operators provide access to the Internet and Internet service providers provide P2P services for all Internet users. The first of the proposed architectures allows running a P2P overlay network on top of the standard SIP infrastructure. In this architecture the SIP servers acted as peers that can be maintained by a service provider or by users themselves. A client is an application running on a handset. Every client is connected to one or many peers. Peers interact with other peers and clients using SIP signaling. In order to allow usage of SIP as a client protocol, we defined a new SIP 'resource' event package together with the syntax of an XML document that describes the generic resources available in the overlay. This event package also allows SIP endpoints to subscribe to changes in the availability of resources, or perform searches for the availability and location of a given resource. The event package and XML syntax are presented in my supporting publications [23][8]. Based on the comments from the IETF community we also defined a new SIP event package and data format for describing files that are presented in my supporting publications [66][9][30]. This event package and the syntax of an XML document are used for publication and searches for content files rather than more generic resources defined by the SIP ‘resource ’ event package.

I also designed a multipurpose P2P network architecture. In contrast to the previous architectures in this architecture, the P2P clients communicate with the peers using a client protocol that supports generic methods for storing, updating, deleting, and retrieving data from the overlay. Clients connect to the P2P network using XML RPC or P2PP protocols, whereas unmodified SIP devices use the standard SIP protocol to connect to peers that implement a SIP server.

In the proposed architecture the client protocol can also be a subset of the protocol running between peers, such as the P2PP protocol. This is an approach that we proposed in the supporting publication [7]. The advantage of designing a client protocol as a subset of the peer protocol is that a device that implements a single P2P protocol stack can act either as a client or as a peer. Depending on the situation, a device in particular a mobile phone may change its role from a peer to a client and vice versa. I designed a mechanism that allows for a smooth transition between peer and client roles. The mechanism is presented in the patent application [12]. We also designed mechanisms for mobility support and monitoring in mobile P2P networks that are presented in the supporting publications [26][27].

The multipurpose P2P overlay does not require peers to implement the SIP stack. Nodes store and retrieve their contact information from the overlay. There is no need for SIP servers. SIP is used for multimedia session establishment directly between mobile terminals without involvement of any peer node. The multipurpose P2P overlay may also store information about servers that assist clients and peers in NAT traversal such as Simple Traversal of UDP (User Datagram Protocol) through NATs (STUN) or Traversal Using Relay NAT (TURN) servers.

Working on this architecture, we were also researching and developing P2P protocols. I worked on a Peer-to-Peer Session Initiation Protocol (P2PSIP) security requirements draft [19], that describes security requirements in P2P overlay mechanisms standardized in IETF[15]. I also co-authored a P2PSIP Protocol Framework and Requirements draft [10], that specifies a P2P protocol framework, presenting design options and recommending specific P2P protocol functionalities. Finally, together with Salman Baset and Henning Schulzrinne from Columbia University we designed a multipurpose P2P Protocol (P2PP) that is suitable both for real-time communication and content sharing. The protocol is presented in [7].

I also researched solutions for P2P security and NAT traversal. We developed a P2P security mechanism for VOIP services utilizing the ZRTP protocol. The mechanism is presented in the supporting paper [20]. I also designed a mechanism for secure data update in DHT based networks. The mechanism is presented in my patent application [21]. I also researched and developed NAT traversal mechanisms for P2P services that were published in the patent applications [24][25].

The above presented architectures, protocols and mechanisms were designed to overcome mobile phone constraints and to make deployment of mobile P2P applications feasible. However, as our user studies revealed it was not enough to implement a mobile P2P application that runs on a mobile phone without draining its battery in a few hours. Mobile P2P overlays had to support community features.

In order to fullfil these requirements I was researching mechnisms and architectures supporting community management in mobile P2P networks. I introduced the Social Layer that enhances the P2P overlay with mechanisms of social networking such as group formation, social search, and social communication. The social layer uses information about social connections between users and maps them to the P2P overlay layer. Social connections are modeled using the social network theory. The theory represents individual actors and their social relations using a social network

[15] Internet Engineering Task Force (IETF), online: www.ietf.org, last time accessed on 12.05,2013.

abstraction. In the social networks we have nodes and ties. Nodes are individual actors, in our case mobile phone users or organizations, and ties reflect the relationships between the actors.

We proposed a P2P “requestor mediated” model. In this model each peer has a set of pointers to other peers, which collectively define a potential P2P network topology. The pointers “virtually” connecting peers are defined by the list of “business cards” contained in contact applications in mobile terminals. This method of topology creation possesses two main advantages. A person can be uniquely identified by some of the information contained in his or her business card. This information forms a universal address that is independent of the addressing mechanism of the underlying communication technologies. It also allows using a set of communication mechanisms to reach a particular person.

The effectiveness of the routing mechanism on the social layer depended on the social network connectivity. An accurate map of social relations is crucial for achieving high efficiency of the search process. In order to mitigate this problem, I proposed a user friendly mechanism for publishing, requesting, and updating information stored in business cards, such as URIs, names, addresses, photos of their contacts, across different User Agents such as mobile and fixed terminals. The mechanism can easily extend the functionality of Personal Information Management (PIM) software in cellular and fixed UEs.

I also proposed the SDHT architecture that allows interconnecting many overlay networks using information on the social layer. The architecture allows each community to run its own overlay and freely decide on the used P2P algorithms. Using the architecture we may split a single P2P overlay network, such as Mainline Bittorent 1 million nodes DHT, into a number of smaller overlay networks maintained by respective communities or allow smaller overlay networks to interconnect with large overlay networks. Mobile phones can participate in P2P overlays as peers because the overlay traffic can be limited by limiting the size of the overlay network overcoming the battery constraints. Moreover, the implementation of security mechanisms can be less difficult in community maintained overlays since each community governs its own data.

*Mobile P2P applications and services*

In order to test all developed technologies and verify my assumptions we implemented various mobile P2P services. In particular, we have designed and implemented the world’s first resource sharing system that works on top of SIP networks. The system enables mobile phone users to share resources with each other and does not require changes in the basic SIP infrastructure. We have also designed a Social DHT architecture that allows for efficient formation of mobile communities. We have proven that a P2P infrastructure can become a feasible cost efficient replacement for a mobile infrastructure by presenting a Distributed IP Multimedia Subsystem. We deployed the world first mobile P2PSIP system for real-time communication services. We also developed a P2P system that allows mobile phone users to search for knowledge in their trusted social communities overcoming the problems identified in the business study of the article dissertation. All systems were tested and were shown to work well in the mobile environment. The results of measurements and the conducted trials prove the technical feasibility of community service provisioning using P2P technology.

*Future work*

We have shown that P2P technologies can be used to provide mobile community services as well as to distribute mobile service infrastructure. The business and technical studies demonstrate the potential and technical feasibility of the proposed technologies and services. In order to fully verify the market potential of the innovations we need to carry out large scale user acceptance tests. This includes large-scale usability tests, user reviews, and trials. This work has yet to be undertaken, but will be included in future work. This future work will also include research on detailed quantitative economic models that will be developed based on the results from the large scale tests and further interviews with stakeholders. With the advance of mobile network technologies and changes in the mobile ecosystem, we see the possibility of the deployment of pure P2P overlay networks in the mobile environment in the future. In order to consider the implementation of pure P2P overlay networks in the mobile environment, there is a need for optimization of DHT algorithms in terms of battery consumption. Our battery consumption experiments in DHT networks were concentrated on the Kademlia algorithm due to its proven scalability in large scale deployments. Other DHT algorithms were not widely implemented at the time of our studies. Therefore, it is advisable to also study other DHT algorithms when they are implemented in the large scale DHT networks and compare the results.

# References


[1] K. Cho, K. Fukuda, H. Esaki, and A. Kato, "The impact and implications of the growth in residential user-to-user traffic", SIGCOMM Computer Communication Review, Vol. 36, No. 4, pp.207-218, October 2006.

[2] H. Tschofenig and M. Matuszewski, "Dealing with P2P Traffic in an Operator Network: State-of-the-Art", P2P Infrastructure Workshop, May 2008.

[3] J. Peterson and A. Cooper, "Report from the IETF Workshop on Peer-to-Peer (P2P) Infrastructure", RFC5594, IETF, May 2008.

[4] H. Haverinen, J. Siren and P. Eronen, "Energy Consumption of Always-On Applications in WCDMA Networks", IEEE 65th Vehicular Technology Conference, April 2007.

[5] A. Arjona and A. Ylä-Jääski, "VoIP Call Signaling Performance and Always-On Battery Consumption in HSDPA, WCDMA and WiFi", IEEE International Conference on Wireless Communications, Networking and Mobile Computing WICOM'07, September 2007.

[6] V. Pascual, M. Matuszewski, E. Shim, H. Zheng, and Y. Song "P2PSIP Clients", Internet-Draft, draft-pascual-p2psip-clients-01, IETF, February 2008 (expired).

[7] S. Baset, H. Schulzrinne and M. Matuszewski, "Peer-to-Peer Protocol (P2PP)", Internet-Draft, draft-baset-p2psip-p2pp-01, IETF, November 2007 (expired).

[8] M. A. García-Martín, M. Matuszewski, N. Beijar, and J. Lehtinen, "A Framework for Sharing Resources with the Session Initiation Protocol (SIP)", Internet-Draft, draft-garcia-sipping-resource-sharing-framework-01, IETF, December 2006 (expired).

[9] M. A. Garcia-Martin and M. Matuszewski, "Sharing Files with the Session Initiation Protocol (SIP)", Internet-Draft, draft-garcia-sipping-file-sharing-framework-01, IETF, November 2007 (expired).

[10] D. Brian, S. Baset, M. Matuszewski and H. Sinnreich, "P2PSIP Protocol Framework and Requirements", Internet-Draft, draft-bryan-p2psip-requirements-00, IETF, Jun 2007 (expired).

[11] M. Matuszewski and M. A. Garcia-Martin, "Method, System, and Devices for Network Sharing or Searching Of Resources", Patent Application, US20100095009 and WO 2008/041173, April 2010.

[12] M. Matuszewski, "Method, apparatus, and computer program product for providing a smooth transition between peer-to-peer node types", Patent Application, WO 2009/071971, December 2009.

[13] M. A. Garcia-Martin and M. Matuszewski, "Node", Patent Application, US20070237139 and WO 2007/116258, October 2007.

[14] E. Kokkonen, S. Baset and M. Matuszewski, "Demonstration of Peer-to-Peer Session Initiation Protocol (P2PSIP) in the Mobile Environment", IEEE Consumer Communications and Networking Conference, January 2008.

[15] M. Matuszewski, N. Beijar, J. Lehtinen and T. Soinio (previously T. Hyyryläinen), "Mobile Peer-to-Peer Content Sharing Application", IEEE Consumer Communications and Networking Conference 2006, January 2006.

[16] M. Matuszewski and M. A. Garcia-Martin, "Distributed IP Multimedia Subsystem", IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, June 2007.

[17] M. Matuszewski and M. A. Garcia-Martin, "Method for the routing of multimedia communication related signaling in a communication system", Granted Patent, US7796990 B2 and WO 2008/031924, September 2010.

[18] I. Kelényi, J. K. Nurminen and M. Matuszewski, "DHT Performance for Peer-to-Peer SIP – a Mobile Phone Perspective", IEEE Consumer Communications and Networking Conference, January 2010.

[19] M. Matuszewski, J-E. Ekberg, and P. Laitinen, "Security requirements for P2PSIP", Internet-Draft, draft-matuszewski-p2psip-security-requirements-02, November 2007 (expired).

[20] E. Kokkonen and M. Matuszewski, "Peer-to-Peer Security for Mobile Real-Time Communications with ZRTP", IEEE Consumer Communications and Networking Conference, January 2008.

[21] M. Matuszewski, "Method, Apparatus and Computer Program Product for Providing Data Management in a P2P Network", Patent Application, US20090125637 and WO/2009/060417, May 2009.

[22] M. Matuszewski, O. Rantapuska, and M. A. Garcia-Martin, "Contacter: an enhanced contact application for easy update and recovery of contacts using the Session Initiation Protocol (SIP)", IEEE International Conference on Portable Information Devices, March 2007.

[23] M. A. Garcia-Martin and M. Matuszewski, "A Session Initiation Protocol (SIP) Event Package and Data Format for Publication and Searching Generic Resources", Internet-Draft,

draft-garcia-sipping-resource-eventpackage-01, IETF, December 2006 (expired).

[24] M. Matuszewski, “Methods, apparatuses, and computer program products for facilitating establishing a communications session”, Patent Application, US20090287829, August 2012.

[25] J. Noyranen, O. Rantapuska, H. Kokkinen, M. Matuszewski, E. Kokkonen, J. Porio., “Method for the establishing of peer-to-peer multimedia sessions in a communication system”, Patent Application, US20100312901 and WO 2008/139033, December 2010.

[26] M. Matuszewski and M. A. Garcia-Martin, “Node and method to provide and keep real-time up-to-date data in a distributed hash table”, Patent Application, US20080288654 and WO 2008/142511, November 2008.

[27] M. Matuszewski, M. A. Garcia-Martin, and C. Schmidt, “Method, apparatus, and computer program product for providing device monitoring in a network”, Patent Application, WO 2009/093215, July 2009.

[28] M. A. Garcia-Martin and M. Matuszewski, “Resource Descriptions Extension to the Presence Information Data Format (PIDF)”, Internet-Draft, draft-garcia-sipping-resource-desc-pidf-00, IETF, December 2006 (expired).

[29] M. A. Garcia-Martin and M. Matuszewski, “File Descriptions Extension to the Presence Information Data Format (PIDF)”, Internet-Draft, draft-garcia-sipping-file-desc-pidf-01, IETF, November 2007 (expired).

[30] M. A. Garcia-Martin and M. Matuszewski, “An Extensible Data Format (XML) for Describing Files”, Internet-Draft, draft-garcia-app-area-file-data-format-00, IETF, November 2007 (expired).

[31] M. Matuszewski and R. Kantola, “Aligning mobile P2P content sharing services with a customer process”, Third IASTED International Conference on Communications and Computer Networks, October 2005.

[32] M. Matuszewski, “Peer-to-Peer content distribution”, 12th International Conference on Telecommunication Systems, Modeling and Analysis, July 2004.

[33] M. Giesler and M. Pohlmann, “The Anthropology of File Sharing: Consuming Napster as a Gift”, Advances in Consumer Research, Vol. 30, pp. 273-279, 2003.

[34] R. Mannak, H. Ridder, and D.V. Keyson, “The human side of sharing in peer-to-peer networks”, 2nd European Union Symposium on Ambient Intelligence, November 2004.

[35] C. Pauli and M. Shepperd, “An Empirical Investigation into P2P File-Sharing User Behaviour”, Eleventh Americas Conference on Information Systems, August 2005.

[36] U. Lee, M. Choiz, J. Choy, M. Y. Sanadidiy, and M. Gerlay, “Understanding Pollution Dynamics in P2P File Sharing”, 5th International Workshop on Peer-to-Peer Systems, February 2006.

[37] R. Cheng and J. Vassileva, “User Motivation and Persuasion Strategy for Peer-to-Peer Communities”, 38th Annual Hawaii International Conference on System Sciences, January 2005.

[38] GE Annual Report 1996, online: http://www.ge.com/annual96/share.htm, last time accessed on 3.03.2006.

[39] D.J. Teece, G. Pisano and A. Shuen, “Dynamic Capabilities and Strategic Management,” Strategic Management Journal, Vol. 18, No.7, pp. 509 – 533, August 1997.

[40] M. Huysman and D. de Wit, “Practices of Managing Knowledge Sharing: Towards a Second Wave of Knowledge Management”, Knowledge and Process Management, Vol. 11, No. 2, pp. 81-92, April/June 2004.

[41] J.P. Bansler and E.C. Havn, “Sharing best practices: an empirical study of it-support for knowledge sharing”, 9th European Conference on Information Systems, July 2001.

[42] C. O’Dell and C. Jackson Grayson, “If only we knew what we know: identification and transfer of internal best practice”, California Management Review, Vol. 40, No. 3, pp. 154-174, Spring 1998.

[43] M. T. Hansen, N. Nohria, and T. Tierney, “What’s your strategy for managing knowledge?”, Harvard Business Review, Mar-Apr, Vol. 77, No. 2, pp. 106-116, March-April 1999.

[44] J. Ewing, “Sharing the Wealth”, Business Week, March 19 2001.

[45] T.H. Davenport, “Ten principles of knowledge management and four case studies”, Knowledge and Process Management Vol. 4, No. 3, pp. 187-208, September 1997.

[46] T. Ghosh, “Creating Incentives for Knowledge Sharing”, MIT Sloan School of Management, May 2004.

[47] A. Ardichvili, V. Page, and T. Wentling, “Motivation and barriers to participation in virtual knowledge-sharing communities of practice”, Journal of Knowledge Management, Vol. 7, No. 1, pp. 64–77, 2003.

[48] E.J.H. Andriessen, “To share or not to share, that is the question. Conditions for the willingness to share knowledge”, Delft

Innovation System, Delft University of Technology, February 2006.

[49] R. Lindgren and O. Henfridsson, "Using Competence Systems: Adoption Barriers and Design Suggestions", Journal of Information and Knowledge Management, Vol. 1, No. 1, pp. 65-77, March 2002.

[50] L. Stevens, "Incentives for Sharing", destinationKM.com, September, last time accessed on July 20, 2005.

[51] R.G. Smith and A. Farquhar, "The Road Ahead for Knowledge Management", AI Magazine, Vol. 21 Nr. 4, Spring 2000.

[52] H. Hall, "Social exchange for knowledge exchange. Managing knowledge: conversations and critiques", University of Leicester Management Center, April 2001.

[53] J.S. Brown and P. Duguid, "Organizing knowledge", California management Review, Vol. 40, No. 3, pp. 90-111, Spring 1998.

[54] Y. Merali, "Self-organizing communities," In S. Rock (Ed.), Liberating knowledge, London: IBM/CBI, pp. 80-87, 2000.

[55] M. Sharrat and A. Usoro, "Understanding Knowledge-Sharing in Online Communities of Practice," Electronic Journal on Knowledge Management, Vol. 1, No. 2, pp. 187-196, 2003.

[56] W.P. Powell, "Learning from collaboration: knowledge and networks in the biotechnology and pharmaceutical industries," California Management Review, Vol. 40, No. 3, pp. 228-240, Spring 1998.

[57] D. DeLong and L. Fahrey, "Diagnosing cultural barriers to knowledge management", Academy of Management Executive, Vol. 14, No. 4, pp. 113-127, November 2000.

[58] J. Kugel and C. Schostek, "Rewards for knowledge sharing?" 2004, online: http://www.gurteen.com/gurteen/gurteen.nsf/id/rewards-k-sharing, last time accessed: 20/10/2005

[59] A.B. Hargadon, "Firms as knowledge brokers: Lessons in Pursuing Continuous Innovation", California Management Review, Vol. 40, No. 3, pp. 209-227, Spring 1998.

[60] P.J.H. Schoemaker, "Multiple scenario development: its conceptual and behavioral foundation", Strategic Management Journal, Vol. 14, No. 3, pp.193–213, March 1993.

[61] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, IETF, June 2002.

[62] G. Camarillo and M. A. García-Martín, "The 3G IP Multimedia Subsystem (IMS)", John Wiley & Sons, ISBN 0470 87156 3, August 2004.

[63] R. Price, C. Bormann, J. Christoffersson, H. Hannu, Z. Liu and J. Rosenberg, "Signaling Compression (SigComp)", RFC 3320, IETF, January 2003.

[64] S. Kent and R. Atkinson, “Security Architecture for the Internet Protocol”, RFC 2410, IETF, November 1998.

[65] P. Calhoun, J. Loughney, E. Guttman, G. Zorn and J. Arkko, “Diameter Base Protocol”, RFC3588, IETF, September 2003

[66] M. A. Garcia-Martin and M. Matuszewski, “A Session Initiation Protocol (SIP) Event Package and Data Format for Describing Files”, Internet-Draft, draft-garcia-sipping-file-event-package-01, IETF, November 2007 (expired).

[67] P. Maymounkov and D. Mazières, “Kademlia: A Peer-to-peer Information System Based on the XOR Metric”, First International Workshop on Peer-to-peer Systems, March 2002.

[68] R. Stoica, D. Morris, M. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications”, ACM SIGCOMM 2001, August 2001.

[69] N. Leibowitz, M. Ripeanu, and A. Wierzbicki, “Deconstructing the Kazaa Network," 3rd IEEE Workshop on Internet Applications, June 2003.

[70] N. Beijar, “Distributed Resource Discovery: Architectures and Applications in Mobile Networks”, Doctoral Dissertation, Aalto University, 2010.

[71] B. Bakos, G. Csucs, L. Farkas, and J. K. Nurminen, “Peer-to-Peer Protocol Evaluation in Topologies Resembling Wireless Networks. An Experiment with Gnutella Query Engine”, International Conference on Networks, September 2003.

[72] I. Keleny and J.K. Nurminen, “Energy Aspects of Peer Cooperation - Measurements with a Mobile DHT System”, IEEE CoCoNet Workshop 2008 Cognitive and Cooperative Wireless Networks, May 2008.

[73] Skype, online: http://www.skype.com, last time accessed on 12.05.2013.

[74] T. Kato, N. Ishikawa, H. Sumino, J. Hjelm, Y. Yu, and S. Murakami, “A platform and applications for mobile peer-to-peer communications.”, NTT DoCoMo & Ericsson Research Document 2003, online: http://www2.research.att.com/~rjana/Takeshi_Kato.pdf, last time accessed on 12.05.2013.

[75] T. Hoßfeld, A. Mäder, K. Tutschku, F-U. Andersen, C. Kappler, H. de Meer, I. Dedinski, and J. Oberender., “An Architecture Concept for Mobile P2P File Sharing Services”, Wurzburg University, Technical Report No. 344, November 2004.

[76] T. Hoßfeld, K. Tutschku, F-U. Andersen, H. de Meer, J. Oberender, “Simulative Performance Evaluation of a Mobile Peer-to-Peer File-Sharing System”, Next Generation Internet Networks, April 2005.

[77] A. Arora, C. Haywood, K. S. Pabla, “JXTA for J2ME 2 Extending the Reach of Wireless with JXTA Technology”, JavaOne Conference, February 2002.

[78] N. Maibaum, T. Mundt, “JXTA: a technology facilitating mobile peer-to-peer networks”, International Mobility and Wireless Access Workshop, October 2002

[79] J. Rosenberg, "A Session Initiation Protocol (SIP) Event Package for Registrations, RFC 3680, IETF, March 2004.

[80] N. Beijar, M. Matuszewski, J. Lehtinen and T. Soinio (previously T. Hyyryläinen), “Mobile peer-to-peer content sharing services in IMS”, International Conference on Telecommunication Systems, Modeling and Analysis 2005, November 2005.

[81] M. Catalan, C. Gomez, D. Viamonte, J. Paradells, A. Calveras and F. Barcelo, “TCP/IP analysis and optimization over a pre-commercial live UMTS network”, IEEE Wireless Communications and Networking Conference, March 2005.

[82] M. Gaynor , “Network Services Investment Guide: Maximizing ROI in Uncertain Times’, Wiley Publishing Inc., ISBN: 0471214752, January 2003.

[83] S. Guha, N. Daswani, and R. Jain, "An experimental study of the skype peer-to-peer voip system.", 5th International Workshop on Peer-to-Peer Systems, February 2006.

[84] M. A. Garcia-Martin and M. Matuszewski, Orthogonal subscription, Patent Application, US20080133644 and WO 2008/065242, June 2008.

[85] E. Harjula, J. Hautakorpi, N. Beijar and M. Ylianttila, “Peer-to-Peer SIP for Mobile Computing: Challenges and Solutions”, Mobile Peer-to-Peer Computing for Next Generation Distributed Environments: Advancing Conceptual and Algorithmic Applications, Hershey, IGI Publishing, ISBN: 1605667153 May 2009.

[86] D. Howie, M. Ylianttila, E. Harjula and J. Sauvola, “State-of-the-art SIP for mobile application supernetworking”, Nordic Radio Symposium, August 2004.

[87] I. Kelényi, P. Ekler, Z. Pszota, “Symtorrent”, Budapest University of Technology and Economics, http://amorg.aut.bme.hu/projects/symtorrent, last time accessed 12.05.2013.

[88] I. Kelényi, B. Forstner and B. Molnár, “Symella”, Budapest University of Technology and Economics,

http://amorg.aut.bme.hu/projects/symella,last time accessed 12.05.2013.

[89] J. A. Lehtinen, “Mobile Peer-to-Peer over Session Initiation Protocol”, Licentiate thesis, Helsinki University of Technology, 2008.

[90] S. Milgram, “The small world problem”, Psychology Today, pp. 60-67, May 1967.

[91] J. K. Nurminen and J. Nöyränen, “Energy-Consumption in Mobile Peer-to-Peer – Quantitative Results from File Sharing”, Fifth IEEE Consumer Communications & Networking Conference, January 2008.

[92] R. Schollmeier and G. Schollmeier, “Why peer2to2peer (P2P) does scale: an analysis of P2P traffic patterns”, 2nd International Conference on Peer/to/Peer Computing, September 2002.

[93] K. Singh and H. Schulzrinne, “Peer-to-peer internet telephony using SIP”, International Workshop on Network and Operating Systems Support for Digital Audio and Video, June 2005.

[94] H. Verkasalo, “A Cross-Country Comparison of Mobile Service and Handset Usage”, Licentiate thesis, Helsinki University of Technology, 2006.

Aalto-DD 136/2013

ISBN 978-952-60-5314-1
ISBN 978-952-60-5315-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

**Aalto University**
**School of Electrical Engineering**
**Department of Communications and Networking**
**www.aalto.fi**

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

**DOCTORAL**
**DISSERTATIONS**