

ACTA

UNIVERSITATIS OULUENSIS

Erkki Harjula

ENERGY-EFFICIENT PEER- TO-PEER NETWORKING FOR CONSTRAINED-CAPACITY MOBILE ENVIRONMENTS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING;
CENTRE FOR WIRELESS COMMUNICATIONS;
INFOTECH OULU



ACTA UNIVERSITATIS OULUENSIS
C Technica 573

ERKKI HARJULA

**ENERGY-EFFICIENT PEER-TO-PEER
NETWORKING FOR CONSTRAINED-
CAPACITY MOBILE ENVIRONMENTS**

Academic dissertation to be presented with the assent of
the Doctoral Training Committee of Technology and
Natural Sciences of the University of Oulu for public
defence in the OP auditorium (L10), Linnanmaa,
on 16 June 2016, at 12 noon

UNIVERSITY OF OULU, OULU 2016

Copyright © 2016
Acta Univ. Oul. C 573, 2016

Supervised by
Professor Mika Ylianttila
Professor Timo Ojala

Reviewed by
Professor Simin Nadjm-Tehrani
Docent Jukka K. Nurminen

Opponents
Professor Henning Schulzrinne
Docent Jukka K. Nurminen

ISBN 978-952-62-1248-7 (Paperback)
ISBN 978-952-62-1249-4 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2016

Harjula, Erkki, Energy-efficient peer-to-peer networking for constrained-capacity mobile environments.

University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering; Centre for Wireless Communications; Infotech Oulu

Acta Univ. Oul. C 573, 2016

University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Energy efficiency is a powerful measure for promoting sustainability in technological evolution and ensuring feasible battery life of end-user devices in mobile computing. Peer-to-peer technology provides decentralized and self-organizing architecture for distributing content between devices in networks that scale up almost infinitely. However, peer-to-peer networking may require lots of resources from peer nodes, which in turn may lead to increased energy consumption on mobile devices. For this reason, peer-to-peer networking has so far been considered unfeasible for mobile environment.

This thesis makes several contributions towards enabling energy-aware peer-to-peer networking in mobile environments. First, an empirical study is conducted to understand the energy consumption characteristics of radio interfaces and typical composition of traffic in structured peer-to-peer networks. This is done in order to identify the most essential obstacles for utilizing peer-to-peer technology in mobile environments. Second, the e-Aware model for estimating the energy consumption of a mobile device is developed and empirically verified to achieve 3-21% error in comparison to real-life measurements. Third, the e-Mon model for the energy-aware load monitoring of peer nodes is developed and demonstrated to improve the battery life of mobile peer nodes up to 470%. Fourth, the ADHT concept of mobile agent based virtual peers is proposed for sharing the peer responsibilities between peer nodes in a subnet so that they can participate in a peer-to-peer overlay without compromising their battery life.

The results give valuable insight into implementing energy-efficient peer-to-peer systems in mobile environments. The e-Aware energy consumption model accelerates the development of energy-efficient networking solutions by reducing the need for time-consuming iterations between system development and evaluations with real-life networks and devices. The e-Mon load monitoring model facilitates the participation of battery-powered devices in peer-to-peer and other distributed networks by enabling energy-aware load balancing where energy-critical mobile nodes carry less load than other nodes. The ADHT facilitates the participation of constrained-capacity wireless devices, such as machine-to-machine nodes, in a peer-to-peer network by allowing them to sleep for most of their time.

Keywords: battery life, constrained-capacity devices, energy-efficiency, green computing, internet of things, machine-to-machine networks, mobile networks, peer-to-peer networks

Harjula, Erkki, Energiatehokkaat vertaisverkkoratkaisut rajoitetun kapasiteetin mobiiliympäristöissä.

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähkötekniikan tiedekunta; Centre for Wireless Communications; Infotech Oulu

Acta Univ. Oul. C 573, 2016

Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Energiatehokkuus on kustannustehokas tapa vähentää päätelaitteiden käytön aiheuttamia kasvi-huonepäästöjä sekä parantaa niiden akunkestoa. Vertaisverkkoteknologia tarjoaa hajautetun, itse-organisoituvan, sekä lähes rajattomasti skaalautuvan verkkoarkkitehtuurin päätelaitteiden väliseen tallennustilan, mediasisältöjen ja tietoliikennekapasiteetin suorajakamiseen. Vertaisverkkojen suurin heikkous mobiilikäytön näkökulmasta on niiden päätelaitteille aiheuttama ylimääräinen kuormitus, mikä näkyy lisääntyneenä energiankulutuksena.

Tässä väitöskirjassa on tutkittu mekanismeja vertaisverkon päätelaitteiden energiatehokkuuden parantamiseksi, tavoitteena riittävä akunkesto mobiilikäytössä. Aluksi työssä tutkittiin empiirisesti langattomien verkkojen mobiilipäätelaitteille aiheuttamaa energiankulutusta sekä rakenteellisten vertaisverkkojen vertaispäätelaitteille aiheuttamaa liikennekuormitusta. Tavoitteena oli muodostaa käsitys suurimmista haasteista mobiililaitteiden käytölle vertaisverkoissa. Seuraavaksi mobiiliverkkojen energiankulutusta koskevasta havaintoaineistosta muodostettiin energiankulutusmalli, e-Aware, jolla voitiin arvioida mobiilipäätelaitteen energiankulutusta erilaisilla verkon liikenneprofileilla. Mallilla saavutettiin parhaimmillaan kolmen prosentin keskimääräinen virhe. Kolmannessa vaiheessa kehitettiin energiatietoinen kuormanseurantamalli, e-Mon, jota käyttäen saavutettiin jopa 470 % lisäys mobiilin vertaispäätelaitteen akunkestoon. Viimeisessä vaiheessa kehitettiin ADHT-konsepti, joka on uudentyyppinen tapa jakaa vertaispäätelaitteiden kuormaa usean saman verkkoklusterin päätelaitteen kesken käyttäen laitteesta toiseen kiertävää mobiiliagenttia.

Väitöskirjan tulokset osoittavat että mobiililaitteiden energiatehokkuutta vertaisverkoissa pystytään olennaisesti parantamaan energiatietoisia kuormanjakomekanismeja käyttäen. Työssä kehitetty e-Aware nopeuttaa energiatehokkaiden hajautettujen järjestelmien kehitystyötä tarjoamalla tehokkaan työkalun järjestelmän energiankulutuksen arvioimiseen jo kehitysvaiheessa. e-Mon mahdollistaa energiatietoisien kuormanjaon vertaisverkoissa tarjoamalla tarvittavan kuormanseurantamallin. ADHT puolestaan tarjoaa uudenlaisen tavan vähentää vertaisverkkojen aiheuttamaa päätelaitekuormitusta hyödyntäen maksimaalisesti rajoitetun kapasiteetin laitteiden utilojen käyttöön perustuvaa energiankulutusoptimointia.

Asiasanat: akunkesto, energiatehokkuus, esineiden internet, laitteidenväliset verkot, mobiiliverkot, rajoitetun kapasiteetin laitteet, vertaisverkot, vihreä tietotekniikka

Acknowledgements

This thesis has been carried out at the MediaTeam Oulu, CIE and CWC research groups, University of Oulu, Finland. The work was conducted as a part of DECICOM, Future Internet, MAMMoTH and CONVINCe research projects.

First of all, I would like to thank my supervisor Prof. Mika Ylianttila for his guidance and mentoring throughout my studies. He has patiently understood the special circumstances for me being the project manager for all the above mentioned projects while encouraging me to progress with my studies. I am also grateful for my second supervisor Prof. Timo Ojala for his high quality reviewing work and several fruitful discussions along the synopsis writing process. I would like to thank Adj. Prof. Jukka K. Nurminen and Prof. Simin Nadjm-Tehrani for their professional and constructive comments as pre-examiners. I also kindly thank Prof. Henning Schulzrinne and Adj. Prof. Nurminen for serving as the opponents in the doctoral defense. I also thank Dr. Pertti Väyrynen for proofreading the manuscript.

Dr. Timo Koskela, Dr. Otso Kassinen, Dr. Zhonghong Ou, Adj. Prof. Andrei Gurtov, Mr. Teemu Leppänen and Mr. Jari Korhonen deserve my deep appreciation for their contributions to the original articles in this thesis. I also would like to thank Mr. Arto Heikkinen, Mrs. Pawani Porambage, Dr. Meirong Liu and Mr. Jussi Ala-Kurikka, among many other present and past colleagues at the university, who I have had the privilege to collaborate with. I also thank my colleagues at Ericsson and Aalto University for long-term collaboration in the field of study of this thesis.

For financial support, I am grateful to Infotech Oulu Graduate School and the foundations of Walter Ahlström, Tauno Tönning, Seppo Säynäjäkangas, Riitta and Jorma J. Takanen, Ulla Tuominen, TeliaSonera, Nokia, KAUTE, TES, HPY, TAES and EIS. For project funding, I am grateful for the Finnish Funding Agency for Innovation (Tekes), and the industrial partners in the corresponding projects.

I would like to express my deep gratitude to my father Allan who has made a tremendous work raising me and my two brothers on his own. I also would like to thank my mother Anneli, my siblings Esa, Ari, Tiina, Anne and Marko, and many other relatives and friends for being present in my life. Special thanks go to my uncle Eero for early inspiration to choose my field of study. Finally, my wife Päivi deserves my deepest respect for her understanding and support for finishing the thesis during the extremely busy years of building our home and the births of our lovely daughters Ella and Malla. Without you, this would not have been possible.

Oulu, April 2016

Erkki Harjula

Abbreviations

2G	Second-generation wireless telephone technology
3G	Third-generation wireless telephone technology
4G	Fourth-generation wireless telephone technology
AC	Alternating Current
AP	Access Point
BLE	Bluetooth Low Energy
CAM	Constantly Awake Mode
CDN	Content Delivery Network
CO ₂	Carbon Dioxide
CoAP	Constrained Application Protocol
CPU	Central Processing Unit
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
DCH	Dedicated Channel
DHT	Distributed Hash Table
DRX	Discontinuous Reception
FACH	Forward Access Channel
GAP	Generic Access Profile
GATT	Generic Attribute Protocol
HTTP	Hypertext Transport Protocol
I/O	Input/Output
ICT	Information and Communication Technology
ID	Identifier
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
L2CAP	Logical Link Control and Adaptation Protocol
LL	Link Layer
LTE	Long Term Evolution for 3G (also known as 4G)
M2M	Machine-to-Machine
MAC	Medium Access Control
MTU	Maximum Transfer Unit
NAT	Network Address Translation

P2P	Peer-to-Peer
PC	Personal Computer
PCH	Paging Channel
PDA	Personal Digital Assistant
PHY	Physical Layer
PO2C	Power of Two Choices
PSM	Power Save Mode
PS-Poll	Power Save Poll
QoS	Quality of Service
RFC	Request For Comments
RFID	Radio Frequency Identification
RLC	Radio Link Control
RNC	Radio Network Control
RRC	Radio Resource Control
SHA	Secure Hash Algorithm
TCP	Transmission Control Protocol
TTL	Time-To-Live
UDP	User Datagram Protocol
UI	User Interface
VS	Virtual Server
WLAN	Wireless Local Area Network
WMSN	Wireless Multimedia Sensor Network
WSAN	Wireless Sensor and Actuator Network
WSN	Wireless Sensor Network

List of original publications

This thesis is based on the following original publications, which are referred to in the text by Roman numerals (I–VI):

- I Kassinen O, Harjula E, Korhonen J & Ylianttila M (2009) Battery Life of Mobile Peers with UMTS and WLAN in a Kademlia-based P2P Overlay. Proc. 20th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. Tokyo, Japan: 662–665.
- II Ou Z, Harjula E, Kassinen O & Ylianttila M (2010) Performance Evaluation of a Kademlia-based Communication-oriented P2P System under Churn. Computer Networks 54 (5): 689–705.
- III Harjula E, Koskela T & Ylianttila M (2011) Comparing the Performance and Efficiency of Two Popular DHTs in Interpersonal Communication. Proc. IEEE Wireless Communications and Networking Conference. Cancun, Mexico: 2173–2178.
- IV Harjula E, Kassinen O & Ylianttila M (2012) Energy Consumption Model for Mobile Devices in 3G and WLAN Networks. Proc. 9th IEEE Int. Conference on Consumer Communications and Networking. Las Vegas, NV, USA: 532–537.
- V Harjula E, Gurtov A, Koskela T, Ojala T & Ylianttila M (2016) Energy-Aware Load Monitoring for Improving Battery Life of Mobile Peer-to-Peer Nodes. *Manuscript*.
- VI Harjula E, Leppänen T, Ojala T & Ylianttila M (2014) ADHT: Agent-based DHT Architecture for Constrained Devices. Proc. 2014 IEEE Global Communications Conference. Austin, TX, USA: 2763–2769.

Papers I–III provide the foundation for the research work in this thesis by studying the characteristics of structured P2P networking from the viewpoint of load inflicted on device, performance and energy efficiency. Papers IV–VI propose solutions for improving energy-efficiency in peer-to-peer networking of mobile and constrained-capacity nodes.

Table of contents

Abstract	
Tiivistelmä	
Acknowledgements	7
Abbreviations	9
List of original publications	11
Table of contents	13
1 Introduction	15
1.1 Background	15
1.2 Motivation and research problems	18
1.3 Scope and methodology	20
1.4 Contributions of the thesis	21
1.5 Organization of the thesis.....	24
2 Literature review	25
2.1 P2P networking	25
2.1.1 Classification of peer-to-peer systems.....	25
2.1.2 Challenges for peer-to-peer networking in mobile environment.....	28
2.2 Load balancing in P2P systems	30
2.2.1 Basic load balancing concepts.....	31
2.2.2 Special requirements for load balancing in P2P systems.....	32
2.2.3 Sources of load imbalance in P2P systems.....	33
2.2.4 Load balancing process	35
2.2.5 Load monitoring	36
2.2.6 Load information exchange	37
2.2.7 Load migration	39
2.3 Energy consumption in mobile networking	44
2.3.1 Decomposition of energy consumption in mobile devices	45
2.3.2 Wireless communication technologies	47
2.3.3 Optimization methods above radio interface	53
2.3.4 Energy consumption in mobile peer-to-peer networking	56
2.4 Summary	58
3 Research contributions	61
3.1 Empirical feasibility analysis of mobile P2P networking	61
3.2 e-Aware - energy consumption model for mobile networking.....	64

3.3	e-Mon - energy-aware load monitoring model to enable load-balanced P2P networking	67
3.4	ADHT –mobile agent-based virtual peer concept for constrained-capacity P2P nodes.....	70
4	Discussion	75
4.1	Summary of results	75
4.2	Limitations and generalizability.....	77
4.3	Significance of the results and future work.....	77
5	Conclusions	81
	References	83
	Original publications	93

1 Introduction

1.1 Background

“Obviously, the highest type of efficiency is that which can utilize existing material to the best advantage.” J. Nehru

Energy efficiency is a powerful measure for promoting sustainability in technological evolution. Improvements in energy efficiency also bring cost-savings, reduce energy infrastructure investments and increase business competitiveness. The role of information and communications technology (ICT) in this context is twofold. On the one hand, ICT is one of the main tools for improving the energy-efficiency of the infrastructures around us (Kramers *et al.* 2014), in the form of vehicle, home, traffic and factory automation, just to mention a few. On the other hand, ICT’s intrinsic energy demand is rising rapidly (Schlommann *et al.* 2015, Celtic-Plus 2012).

A growing share of Internet nodes are mobile (KPCB 2015, CISCO 2014). The rapidly growing and continuously evolving mobile ecosystems have significant strategic and economic implications in all market segments of the ICT (Basole & Karla 2011). The increasing use of mobile and fixed end-user terminals, together with rapid growth of Internet of Things (IoT) including its subtype machine-to-machine (M2M) communication between networked everyday objects, consume more and more energy every day. Along with the growth of the number of consumer devices, the supporting infrastructure has to scale up at the same pace to support the growing capacity demand. Whereas in 2012, ICT sector generated around 2% of global CO₂ emissions, by the year 2020, its share is expected to double (Celtic-Plus 2012). As a consequence, green computing referring to the design, manufacturing, use and disposal of ICT hardware and software with minimal impact on the environment is growing its importance in meeting the requirements of sustainable development.

Along with the requirements for sustainability and overall energy-efficiency, maintaining feasible battery life sets strict requirements for the energy consumption of mobile (battery-powered) end-user devices. These devices include mobile phones, tablets and other handheld devices, as well as smart objects, such as wireless sensor and actuator nodes. During the past decade, mobile technology has developed even faster than was expected during the ICT boom at the turn of the new millennium (CISCO 2014). The requirements concerning the physical size and capacity of batteries have become stricter due to the size miniaturization and growing

performance requirements, while at the same, time more and more advanced hardware and software require more energy (Ravi *et al.* 2008). Since the battery technology has not developed at the same pace, the average battery life of a typical high-end smartphone, for instance, has significantly reduced during the past fifteen years. However, as battery life is one of the centric factors of long-term user experience of mobile devices (Kujala *et al.* 2011, Ickin *et al.* 2012, Heikkinen & Nurminen 2010), sufficient battery life is one of the core requirements of future mobile systems.

Due to the physical limitations and manufacturing costs of battery technology, improving energy-efficiency is a cost-efficient way to increase battery life of mobile devices. Traditionally, energy savings in mobile computing have mostly been obtained from hardware and firmware level optimizations in radio interface and computing architectures. These optimization methods include, e.g. *adaptive transmission power management* on radio interfaces and *duty cycling*, allowing radio interfaces and processors to have long idle periods between active periods (Bontu & Illidge 2009, Rault *et al.* 2014, Xing *et al.* 2007, Jurdak *et al.* 2010). In order to obtain the maximal benefit of these technologies, higher-layer solutions must, among other measures, exploit these features as efficiently as possible. However, the optimization of application-level energy-efficiency has still great potential in further improving the overall energy-efficiency of mobile computing. Approaches such as *optimizing traffic patterns* (Hoque *et al.* 2014, 2015, Qian *et al.* 2012, Ehsan & Hamdaoui 2012), *offloading computation* to more capable nodes (Kumar *et al.* 2013, Kosta *et al.* 2012, Saarinen *et al.* 2012, Kumar & Lu 2010), *content compression* (Ma *et al.* 2013, Kimura & Latifi 2005) and *reducing overhead* of communication protocols and applications (Bormann *et al.* 2012, Ciraci *et al.* 2009) can significantly improve the battery life of a mobile device.

An increasing share of user-created mobile content is stored outside the user device, in order to either share the content with others, to save local storage space or to back up the content. According to this, the virtualization of content and service management is one of the current megatrends in information technology (CISCO 2014, KPCB 2015, Analysys Mason 2013, CISCO Infographic 2013). The content and computation in these services is usually distributed in clouds that may consist of thousands of devices, ranging from dedicated servers to end-user devices. Cloud computing provides a scalable infrastructure for application service providers, relieving them from expensive and time-consuming investments to service infrastructure. This is especially beneficial for small companies that can now focus their limited resources on developing their services instead of building the needed service

infrastructure. Due to the inherent on-demand provision of computing resources, cloud computing helps reducing the overhead energy consumed by servers running idle, which makes cloud computing also an important approach to improve energy-efficiency on a global scale.

Peer-to-Peer (P2P) technology provides a scalable architecture for distributing content between devices in networks that scale up almost infinitely (Androutsellis-Theotokis & Spinellis 2004). Today, many cloud-based services use P2P for optimizing content delivery (Korzun & Gurtov 2013). P2P networks are known for their superior scalability: since computational resources are provided by participating end-user devices, resources inherently scale up with network size. Whereas with traditional client/server-based architectures, the requirements for service infrastructure grow along with the number of users and the amount of stored and transferred data, P2P-assisted cloud computing (Babaoglu *et al.* 2012, Trajkovska *et al.* 2010, Zhelev & Georgiev 2011) alleviates server load by taking into use the spare computational and storage resources of participating peer nodes. These peer nodes can include end-user devices in addition to dedicated server machines. For instance Skype, the popular cloud-based internet communication tool, streams media content directly in P2P manner between end-users instead of relying on centralized streaming servers. Spotify, the popular cloud-based social music service, reduces server load by using end-user devices as a cache. PeerCDN uses the same principle for Web sites.

Current P2P systems have also weaknesses from the viewpoint of mobile computing. Distributed content management and network maintenance require frequent messaging and data transfers among peer nodes in the background, even when devices are not actively used. This is problematic for mobile devices and networks since frequent signalling keeps the mobile radio (particularly 3G, but also 4G) interface active almost continuously, and thus significantly increases energy consumption (Vergara & Tehrani 2013, Vergara *et al.* 2014, Ou *et al.* 2010a, Eronen 2008, Nurminen & Nöyränen 2008). This issue will be emphasized in the future as a growing share of Internet nodes is battery-powered due, e.g. to the emergence of higher-performance wireless communication technologies and device miniaturization that enables higher mobility and new types of mobile services (CISCO 2014, KPCB 2015). Thus, so far, P2P computing has been feasible only if all participating nodes have a high-performance Internet connection, hardware capacity comparable to a personal computer and a fixed power connection. The lack of mechanisms relieving the burden on mobile devices was one of the reasons that most likely caused, e.g. Skype, to abandon its previous fully decentralized P2P model.

Different load balancing mechanisms (Felber *et al.* 2014, Hautakorpi & Mäenpää 2010) and hierarchical solutions (Korzun & Gurtov 2011, 2013, Ou 2010, Artigas *et al.* 2005) have been proposed to balance load between nodes in P2P systems. Faik (2005), Putrycz (2003) and Quin (2008), in turn, have proposed P2P load balancing models equipped with load monitoring functionalities that enable resource-aware load balancing. These models, however, do not include parameters that are important to mobile and other battery-powered devices, such as remaining battery capacity. In order to facilitate the participation of battery-powered devices, heterogeneity in hardware resources and energy characteristics has to be better taken into account.

In addition to increased burden on end-user devices, the participation of mobile nodes in P2P networks also presents challenges to the functionality of P2P networks (Kellerer *et al.* 2006). The mobility of devices makes them less stable in terms of their availability, connection quality and performance. Since mobile devices are an integral part of P2P infrastructure, their mobility affects negatively the QoS (Quality of Service) properties and reliability of P2P networks. Furthermore, due to the limited battery capacity, mobile device users have a strong incentive to disconnect their devices from P2P networks while they are not actively using their devices. The resulting frequent leaving and joining of nodes causes changes in overlay topology, leading to additional maintenance signalling and data transfer load in P2P networks (Stutzbach & Rejaie 2006, Surana *et al.* 2006, Bienkowski *et al.* 2005), which in turn increases energy consumption at system level.

1.2 Motivation and research problems

Since a growing share of Internet nodes are mobile, the requirement for energy-efficiency will be emphasized in the future. In order to obtain the maximal benefit of the advanced energy-saving functions at hardware and firmware levels of processing architectures and radio interfaces, further application-layer optimizations are needed, particularly on P2P networking.

As discussed above, in mobile P2P networking a majority of energy consumption is caused by intense maintenance signalling despite the small amount of data. This is explained by the relatively long period before a mobile radio enters the power saving mode after receiving or transmitting data, which with frequent signalling causes the radio interface to remain active almost continuously. Therefore, optimization of messaging patterns and reducing protocol overheads are empha-

sized in P2P systems. Furthermore, computational offloading and content compression are used to reduce energy consumption in P2P systems, but they are effective only with certain application types, such as distributed computing or multimedia communication.

Load balancing and hierarchical overlays are able to distribute load between P2P nodes based on their hardware capacity and load. Load balancing can, for example, adjust the distribution of stored objects or incoming bindings between peer nodes. Hierarchical solutions, in turn, can place mobile nodes on a lower layer in the node hierarchy to reduce maintenance and third-party load inflicted on them. In order to enable energy-aware distribution of load, these mechanisms require up-to-date load information from participating nodes, including their battery and energy consumption status.

In P2P networks, challenges related to mobile node participation are twofold. On the one hand, the use of P2P technologies can significantly increase the energy consumption of mobile devices. On the other hand, while operating in P2P networks, the special characteristics of mobile devices cause problems related to routing integrity, QoS properties and energy-efficiency at the system level. Power saving functions, mobility and limitations in network coverage makes the online presence of mobile devices transient in its nature. The resulting frequent leaving and joining of nodes increases the need for network maintenance signalling. This is a major issue in mobile P2P networking, where maintenance signalling is a significant source of energy consumption.

Based on the rationale above, this thesis focuses on addressing the following research questions in the context of mobile P2P networking:

- RQ1: How different traffic patterns and network topologies affect the energy consumption of a mobile device?
- RQ2: What load monitoring methods can be used to exploit energy-awareness in load distribution among peer nodes and how much energy can be saved in mobile peer nodes using these methods?
- RQ3: What load balancing methods can be used to reduce constrained-capacity mobile peer node resource usage and how much energy can be saved using these methods?

In Fig. 1, the original publications and their relationships are mapped relative to the three research questions. Papers I–III contribute the preliminary empirical study, where existing technology has been examined from the particular viewpoint

of mobile P2P networking. These papers provide the empirical foundation for papers IV–VI that focus on finding solutions to the research questions.

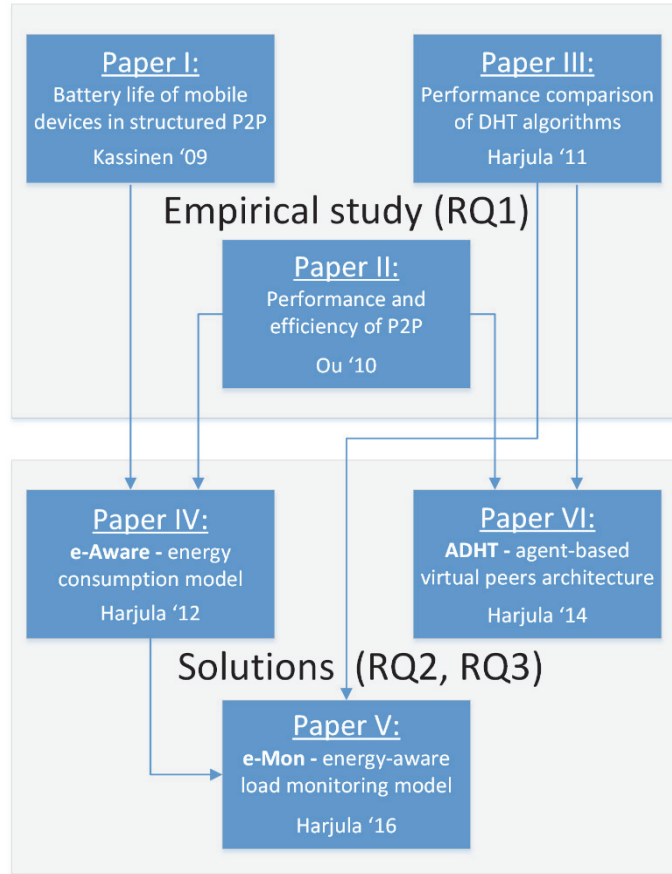


Fig. 1. The mapping of the original publications to the research questions.

1.3 Scope and methodology

This thesis explores mechanisms for enabling energy-aware P2P networking, where load is distributed fairly among nodes with different energy characteristics. The main goal is to enable the accommodation of mobile devices as peer nodes in a P2P overlay, by neither compromising their battery life nor the integrity and performance of the P2P overlay. The key mechanisms for achieving these goals include

load monitoring that takes into account the battery and energy consumption status of peer nodes and load balancing that enables dynamic load redistribution when needed. Following the principles of constructive research, the developed solutions are evaluated and validated by using simulations and prototype implementations as follows.

First, an empirical study is conducted to find the most essential obstacles for utilizing P2P in mobile environments. Papers I–II study a messaging scenario of frequently sent/received User Datagram Protocol (UDP) packets with different packet transmission intervals and packet sizes, to understand the energy consumption characteristics of the radio interface and other critical components of a communicating mobile device. Furthermore, papers II–III contribute experimental evaluations of the performance and overhead of different P2P overlay solutions, as well as the load they inflict on mobile and constrained peer nodes, specifically energy consumption.

Second, a twofold contribution is made towards improving energy efficiency of mobile P2P nodes. First, paper IV develops a model for estimating the energy consumption of a mobile device in particular P2P application scenarios. The goal is to find out how accurately the energy consumption of a mobile device can be estimated in comparison to real-life measurements. Then, paper V studies how energy-aware load monitoring can improve the battery life of a mobile device in these application scenarios. The goal is to find out how much the battery life of a mobile device can be improved by using energy-aware load balancing. This is done by simulating each application scenario with different load balancing and monitoring parameters.

Third, a concept for sharing the peer responsibilities among participating nodes within the same subnet using mobile agent-based virtual peers is introduced in paper VI. The goal is to introduce a disruptive load balancing approach for making P2P networking feasible among constrained-capacity nodes in M2M scenarios where the use of P2P would be beneficial, but dedicated super-peer nodes that would work as gateway nodes for other nodes are neither available nor feasible to use. The performance and efficiency of the architecture are analysed through a comparative analysis.

1.4 Contributions of the thesis

The main contributions of the original publications are summarized below and they are elaborated in more detail in Chapter 3.

1. Empirical study for understanding the energy consumption characteristics of radio interfaces and typical composition of traffic in structured P2P networks, in order to identify the most essential obstacles for utilizing P2P technology in mobile environments (Papers I–III).
2. The e-Aware energy consumption model for estimating the energy consumption of a mobile device in particular application scenarios (Paper IV).
3. The e-Mon load monitoring model for enabling energy-aware load balancing using existing load balancing models (Paper V).
4. The ADHT concept of mobile agent-based virtual peer load balancing for facilitating the participation of constrained-capacity nodes in a P2P overlay (Paper VI).

Paper I evaluates the battery life of mobile devices in a structured P2P overlay network scenario and in an artificial messaging scenario where the mobile node is sending and receiving UDP packets with different parameters. The main goal of the paper was to find out how different parameter settings affect the battery life of a mobile device connected to wireless local area network (WLAN) and third generation (3G) mobile networks. The paper provides valuable information on the energy consumption characteristics of a networked smartphone. The author was responsible for the analysis of the UDP messaging evaluations as well as for defining the scope of the work as the project manager. Dr. Otso Kassinen was the main author of the paper, MSc. Jari Korhonen participated in prototyping and Professor Mika Ylianttila supervised the work.

Paper II evaluates the performance and efficiency of communication-oriented structured P2P systems under churn from the viewpoint of computational and messaging load, as well as the energy consumption of mobile peer nodes. From the viewpoint of this thesis, the most important result is the analysis of energy consumption, which provided further information on the energy consumption characteristics of a networked smartphone in a real-world scenario, with a richer set of parameters compared to Paper I. The author's main contribution was the definition and analysis of the energy consumption measurements. The author also made a significant contribution to the writing of the paper. Dr. Zhonghong Ou was the main contributor of the work and of the paper, Dr. Otso Kassinen participated in prototyping and Professor Mika Ylianttila was the supervisor.

Paper III compares the performance of two popular Distributed Hash Table (DHT) algorithms, Kademlia and Chord, in P2P-based communication. The focus of the evaluation was on messaging overhead and routing performance in overlay

networks of different sizes. The results complement the findings of papers I and II by providing information about the differences between the algorithms in terms of lookup path length and signalling overhead. The author was the main contributor of the work and the paper. MSc. Jari Korhonen contributed to the evaluation, Dr. Timo Koskela participated in finalizing the paper, and Professor Mika Ylianttila was the supervisor.

Paper IV introduces a model called e-Aware for estimating how application layer properties affect the energy consumption of mobile devices operating in mobile networks. The model uses energy consumption profiles that are based on the empirical results of Papers I–II. It facilitates the development of energy-efficient networking solutions by reducing the need for time-consuming iterations between system development and evaluations with real-life networks and devices. The model was implemented and simulated using Matlab, and correct functionality was verified by comparing the results with real-life measurement results. The author was the main contributor of the work and the paper. Dr. Otso Kassinen participated in finalizing the paper, and Professor Mika Ylianttila was the supervisor.

Paper V introduces a novel energy-aware load monitoring model called e-Mon which aims to improve battery life of mobile terminals in distributed systems. The proposed model senses the battery status of a peer node together with other load factors, and enables moving overlay load from energy-critical nodes to less energy-critical nodes. The performance of e-Mon was evaluated in a distributed P2P overlay by applying it with three existing load balancing technologies. The results demonstrate that using energy-aware load monitoring in load balancing has significant potential to increase the battery life of mobile peer nodes. The results also reveal the overall potential of energy-awareness on the application level and exemplify how this potential could be exploited in a distributed network of end-user terminals. The author was the main contributor of the work and the paper. Dr. Andrei Gurtov collaborated in writing the paper and Dr. Timo Koskela participated in finalizing the paper. Professors Timo Ojala and Mika Ylianttila were the supervisors.

Paper VI proposes a novel architecture called ADHT to enable the participation of wireless and constrained-capacity devices in a P2P overlay without compromising their battery life. The architecture is a disruptive load balancing approach for making P2P networking feasible among constrained-capacity nodes in scenarios where the use of P2P would be beneficial, but dedicated super-peer nodes are neither available nor feasible to use. The effect on the power consumption of mobile

devices is analysed by comparing the model with alternative architectures in a comparative analysis. The author was the main contributor of the work and the paper. The underlying agent architecture (Leppänen *et al.* 2013) was contributed by MSc. Teemu Leppänen who also contributed to the writing of the paper. Professors Mika Ylianttila and Timo Ojala were the supervisors.

In addition to the included original publications, the author has published several other original papers providing background for the topic of this thesis. The most significant of them is Harjula *et al.* (2004), proposing a P2P middleware for mobile devices. With its 85 citations so far, it has achieved a landmark paper status in the area of mobile P2P networking. The author has also co-authored a survey article on P2P group management systems (Koskela *et al.* 2013) in a high impact journal, providing valuable information on P2P systems in general, also relevant from the viewpoint of this thesis. In Ou *et al.* (2010b), the author has participated in defining a pyramid-based peer-to-peer architecture over structured DHT networks, and in Harjula *et al.* (2006, 2009), the author has studied the Peer-to-Peer Session Initiation Protocol (P2PSIP) architecture.

1.5 Organization of the thesis

The rest of the thesis is organized as follows. Chapter 2 frames the thesis in form of a literature review of related work. The related work introduces the concept of P2P networking, its current use in mobile environment and existing load balancing mechanisms, as well as energy consumption characteristics of different mobile device types and networks, and the most widely used energy saving mechanisms in them. Chapter 3 presents the main research contributions of the thesis: the empirical feasibility analysis of mobile P2P networking, the e-Aware energy consumption model for mobile networking, the e-Mon energy-aware load monitoring model for load-balanced mobile P2P networking, and the ADHT mobile agent-based virtual peer load balancing concept. Chapter 4 summarizes the contributions, discusses their limitations, generalizability and significance, and explores avenues for future work. Chapter 5 concludes the thesis.

2 Literature review

This chapter provides a review of related work in P2P networking (Section 2.1), P2P load balancing (Section 2.2) and energy consumption in mobile networking (Section 2.3). Section 2.4 summarizes the literature review.

2.1 P2P networking

Peer-to-peer (P2P) systems are distributed systems that operate without centralized organization or control. The concept of P2P has been defined in multiple different ways in the literature, for example by Androutsellis-Theotokis & Spinellis (2004), Xu *et al.* (2003), Halepovic & Deters (2003) and Sakaryan & Unger (2004). In Koskela *et al.* (2013), we have provided the following definition, which is based on the common properties of the different definitions found in the literature:

P2P systems are distributed systems that are populated by autonomous and heterogeneous nodes being capable of self-organising into network topologies without centralized control, for the purpose of sharing resources such as content, CPU cycles, storage and bandwidth.

In P2P systems, the participating devices form an overlay network of logically linked peer nodes that provide the needed routing, computing and storage resources for the system. The topology of a P2P overlay network is independent of the topology of the lower layers, and several overlays can be linked to form, e.g. hierarchical overlay structures (Ou 2010). Since the needed resources are provided by the participating nodes, the volume of system resources inherently scales up with the network size (Stoica *et al.* 2001, Ratnasamy *et al.* 2001, Rowstron & Druschel 2001). P2P systems are used for creating highly scalable content discovery and distribution systems that minimize the need for large and expensive server infrastructures. In addition to practically infinite scalability, P2P systems also provide good failure tolerance, low investment cost, and minimal need for maintenance (Androutsellis-Theotokis & Spinellis 2004).

2.1.1 Classification of peer-to-peer systems

According to Androutsellis-Theotokis & Spinellis (2004) and Koskela *et al.* (2013), P2P systems can be classified based on the *type of the index* and the *structuredness* of the overlay.

The type of the index may be either *centralized*, *distributed* or *hybrid*. The index is the data structure used for locating nodes or shared resources in a P2P network. In centralized indexing, the index is stored in a single location on a server or a server pool (Brands & Karagiannis 2009). In distributed indexing, the index is collectively maintained by the peer nodes (Brands & Karagiannis 2009). In hybrid indexing, the index is distributed to a subset of peer nodes called *super-nodes* (or *super-peer* nodes that will be used from now on) (Yang & Garcia-Molina 2003). The index is collectively maintained by the super-peer nodes on behalf of the ordinary nodes connected to them. The super-peer nodes are usually selected based on the hardware capacity and stability of peer nodes. The existing P2P systems combine different levels of structuredness and indexing strategies. For example, all DHT-based P2P networks are structured systems with distributed indexing, Gnutella (v0.6) is an unstructured system with distributed indexing, BitTorrent uses, depending on the version, either unstructured or a combination of structured and unstructured networks while its index is centralized (Koskela *et al.* 2013).

In *unstructured P2P systems*, the network emerges arbitrarily in the course of time and nodes are arranged in an unorganized manner. The message routing is usually based either on *flooding*, *random walk* or *expanding-ring* search (Lua *et al.* 2005, Yang & Fei 2009, Ciraci *et al.* 2009, Androutsellis-Theotokis & Spinellis 2004). In flooding, the messages are broadcast to all known neighbour nodes of the originating node that again broadcast the message to all of their neighbouring nodes. In random walk, the nodes forward the incoming request messages to a limited set of neighbour nodes, selected randomly. With both techniques, the message propagation continues until the message reaches its destination or until the time-to-live (TTL) value decreases to zero. The TTL is a numerical value that is subtracted by one at each routing hop. It is used to limit the scope of the message propagation. The expanding ring search is a subtype of flooding, where the TTL is initially set to a small value. If the lookup does not succeed, the TTL value is increased and a new request is sent. This continues until the request succeeds or the maximum TTL value is reached. The original Gnutella (v0.4) and FastTrack are the most well-known examples of unstructured P2P systems.

In *structured P2P systems*, the overlay structure determines the type of routing algorithm (Brands & Karagiannis 2009). The data structure is collectively maintained by participating peer nodes, and is usually based on DHT algorithms (Androutsellis-Theotokis & Spinellis 2004). DHT maps both the nodes and their shared resources into a numeric address space. Each peer node and resource has a unique numeric identifier specifying their location in the address space.

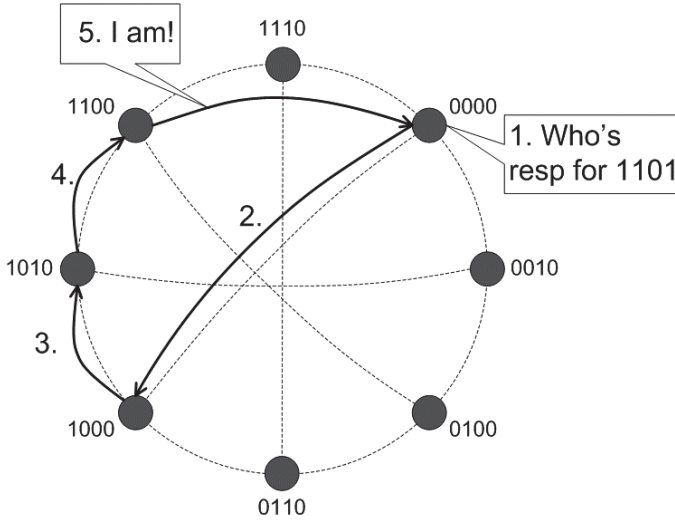


Fig. 2. Query routing in Distributed Hash Tables (DHT).

Peer nodes in structured P2P networks obtain their node identifiers, e.g. from an enrolment server, and join the overlay through bootstrap process, which is defined by the used P2P protocol (Brands & Karagiannis 2009, Androutsellis-Theotokis & Spinellis 2004). During the join process, a peer node takes over a part of the address space according to its node identifier. The resource lookup is illustrated in Fig. 2. When a resource or another peer node is looked up, its hash code is first resolved using a hash algorithm (Karger *et al.* 1997), such as Secure Hash Algorithm (SHA) (NIST 2012), and the lookup query is then sent to the overlay. The query is then routed to the neighbouring node whose node identifier is numerically closest to the looked up resource or node. This procedure is repeated by the next nodes along the path until the lookup query reaches the target node responsible for the hash code. The routing algorithm and structure is defined by the used DHT algorithm (Brands & Karagiannis 2009). Further messaging and data transfers between the nodes are typically routed directly between the nodes using their IP addresses. When a hash-code of a specific node or a resource is known, successful routing between any two overlay-connected nodes can be guaranteed in an undivided P2P overlay (Koskela *et al.* 2013). The most popular DHT-algorithms are Kademlia (Maymounkov & Mazieres 2002), Chord (Stoica *et al.* 2001), CAN (Ratnasamy *et al.* 2001), Pastry (Rowstron & Druschel 2001) and Tapestry (Zhao *et al.* 2004). Examples of structured P2P systems and protocols include BTDig, a P2P version of BitTorrent,

FAROO, a P2P Web search engine, CJDNS, a routing engine for mesh-based networks, and P2PSIP, a P2P version of Session Initiation Protocol (Jennings *et al.* 2013). Also non-DHT based structured P2P systems exist, such as Mercury (Bharambe *et al.* 2004). In Mercury, the data structure is maintained by peer nodes as with DHTs, but DHT algorithm is not used.

P2P overlays can also be structured hierarchically, in order to achieve better performance, efficiency, maintenance cost or load balance. *Hierarchical overlay structures* can be categorized into traditional *two-level P2P systems* with super-peer and regular peer nodes, *single-overlay P2P systems* with internal node hierarchies, or *multi-overlay P2P systems* with hierarchically-linked separate overlays (Korzun&Gurtov 2011, 2013, Koskela *et al.* 2013). Two-level P2P systems typically include a P2P overlay network consisting of super-peer and regular peer nodes that connect the overlay through super-peers. Regular peer nodes neither participate in the overlay management nor provide resources for the overlay. Single-overlay systems include structured P2P networks that have internal hierarchical routing schemes (Korzun&Gurtov 2011, 2013). Multi-overlay P2P systems can be further divided into *structured*, *hybrid* and *unstructured* hierarchical P2P systems (Ou 2010). Structured hierarchical P2P systems utilize structured overlay networks in their topologies, hybrid hierarchical P2P systems utilize both unstructured and structured overlays, and unstructured hierarchical P2P systems utilize only unstructured overlays in their topologies (Ou 2010).

2.1.2 Challenges for peer-to-peer networking in mobile environment

In many aspects, mobile network environment is more constrained and controlled than fixed network environment. Limitations concern the hardware resources of end-user devices, the available communication technologies and their features, as well as different network operator restriction policies (Kim & Park 2006, Kellerer *et al.* 2006, Mäkinen & Nurminen 2008). In addition to cellular technologies, mobile computing today covers a wide variety of other technologies, such as wireless nodes in IoT and its subtype M2M computing. These new technologies bring the Internet everywhere around us, as part of the infrastructure we use in our everyday life (Gubbi *et al.* 2013). It is predicted that M2M device connections worldwide will grow by a factor of 10-20, from around 2 billion devices in 2013 to around 20-50 billion devices in 2023 (Analysys Mason 2013, CISCO Infographic 2013). IoT/M2M computing environment is even more constrained than traditional mobile

computing due to extremely low computational and networking capacity of many of the networked nodes.

A predominant challenge for P2P networking in mobile environment is related to the *hardware capacity of nodes* (Hoßfeld *et al.* 2005, Kim & Park 2006, Kellerer *et al.* 2006). Wireless radio technologies and mobile computing architectures are in many aspects more restricted than their fixed counterparts. In addition to lower computing performance and network throughput, there are also differences in network latency and communication range. Also limited battery life affects the above mentioned features through strict energy saving requirements (Vergara & Tehrani 2013, Vergara *et al.* 2014, Hoque *et al.* 2015). The use of sleep modes, for example, may significantly increase network latencies (Jurdak *et al.* 2010, Gomez *et al.* 2012, Ayoub *et al.* 2011). However, major differences in the hardware capacity between peer nodes pose even a more substantial challenge for the P2P network functionality than the limited hardware capacity itself would inflict (Rao *et al.* 2003, Kim & Park 2006). The reason is that – with the assumption that the maintenance overhead is not too high – in a typical P2P application scenario, the third party load can easily be handled by a node that has around the same hardware capacity as the node generating the load. In contrast, if a lower-capacity node (with no load balancing capability) would participate in an overlay where the other nodes are higher-capacity nodes, it would have to carry an equal amount of load as the higher-capacity nodes.

There are also other challenges for P2P networking in mobile environment. The *transient nature of mobile device online presence*, that is caused by, e.g. power saving functions, the user interventions and limitations in the network coverage, affects negatively the operation of P2P overlays (Kellerer *et al.* 2006). This is especially harmful for structured P2P networks, where the overlay has to adapt to frequent leaving and joining of nodes, which in turn increases the needed network maintenance signalling (Stutzbach & Rejaie 2006, Surana *et al.* 2006, Bienkowski *et al.* 2005). This phenomenon of frequent joining and leaving in P2P is called *churn*. In unstructured P2P networks, churn does not inflict as high extra overhead, but the overhead can otherwise be very high due to their inefficient query routing.

Yet another challenge for P2P networking in mobile environment is the *restricting policies and technologies* used by the network operators. Most of today's cellular devices have private IP addresses and they are behind firewalls that hinder their direct connectivity with nodes outside the operator network and may significantly limit the number of allowed protocols and their performance (Mäkinen & Nurminen 2008). Network Address Translation (NAT) is a mechanism to create a

private IP address space within a network. NAT remaps one address space into another by modifying the Internet Protocol's (IP) network address information (Sriuresh & Egevang 2001). The deployment of the new IP version 6 (IPv6) with larger address-space has removed the original motivation to extend the limited address-space of IPv4, but it seems that the use of NAT will be continued with IPv6 for other reasons. NAT, for instance, effectively hides the number of hosts behind the NAT box and allows administrators to better control what goes out or comes into the network. The use of NAT is problematic for P2P networking, since it is vital that participating peer nodes are able to communicate with each other without intermediaries (Mäkinen & Nurminen 2008). In P2P networks, the network management and computational resources are provided by the nodes on the edge of the network, instead of centralized servers. This requires peer nodes to establish and maintain connections between each other, which is problematic if both are behind NATs. Several mechanisms for traversing NATs have been proposed, but the problem with them is the increased signalling overhead. The frequent signalling keeps the mobile radio interface active almost continuously due to the fundamentals of radio interface properties, and thus significantly increases the energy consumption of mobile nodes (Stutzbach & Rejaie 2006, Surana *et al.* 2006, Bienkowski *et al.* 2005).

Overall, mobile networks constitute a challenging environment for the use of P2P technologies. The limited battery life is already a widely known problem with smartphones, and the utilization of P2P technologies makes the situation even more problematic due to increased maintenance messaging load, also during the idle periods. This maintenance messaging may drain the battery of a mobile device in only a few hours. Thus, optimization mechanisms that would reduce the burden on mobile peer nodes in P2P overlays are needed.

2.2 Load balancing in P2P systems

Load balancing is a widely used method for distributing computational tasks between nodes in a distributed network so that there are no major differences in load experienced by them. Load balancing is also a valid method in P2P systems to avoid situations where some peer nodes or links would experience much heavier load than others (Felber *et al.* 2014).

Load balancing in P2P systems has special aspects and requirements that distinguish it from load balancing in distributed server systems. Whereas in most client/server systems it is enough to prevent the load within a node to grow in such an

extent that the node would fail in performing its tasks, in P2P systems the requirements for the fairness of load distribution are higher. This is due to the fact that P2P nodes are not usually dedicated to certain tasks. Instead, P2P software usually runs on the background while the user of the device is doing other tasks with the device. If background tasks require too much computing resources, the user experience suffers from it (Ickin *et al.* 2012, Doherty & Sorenson 2015). In addition, background tasks increase power consumption of mobile devices, which may significantly reduce their battery life. Furthermore, the performance of the P2P system is negatively affected if some of the nodes are overloaded.

During the past years, load balancing has been among the popular research topics in P2P networking. This is exemplified by the broad number of P2P load balancing systems that have been proposed in the literature for various purposes. In this section, the basic concepts of P2P load balancing and sources of load imbalance are analysed and then a survey of load balancing technologies is presented.

2.2.1 Basic load balancing concepts

A generic definition for *load* in a computer system is given by Saravanakumar & Prathima (2010) and Cooper (1983). They define load as a *measure of the amount of work that a computer system performs*. The work performed by a computer system originates from different tasks, such as processing, communication or data storage. Load can be further divided into *computational load*, *network load*, *memory load* and *input/output (I/O) load*, based on the hardware component inflicted. In addition, energy consumption can be seen as an important load factor in today's networks that include mobile devices. Energy consumption is affected by all of the mentioned load components and the impact depends on the device characteristics.

Load can be *measured* by two fundamentally different methods. The first method is to measure the *hardware load*, i.e. the actual utilization of a node's hardware resources, such as network interface, processor, memory or battery (Ferrari & Zhou 1987, Song & Yang 2006). The second method is to measure the *service load*, including, e.g. the number of service requests, routed messages, the size of stored items, maintenance signalling or the total volume of the traffic (Shen & Xu 2008, Bianchi *et al.* 2006). The relation between the load types can be described as follows; service load generates hardware load and the impact depends on the hardware capabilities of the device in question.

Load can be considered to be *balanced* when the load of different nodes in a distributed computer system is equal or nearly equal. Depending on how load is

measured, this means either balanced service load or balanced hardware load. A node is usually considered *overloaded* (or heavy), if it is burdened by load exceeding the overload threshold value set by the system, and *underloaded* (or light), if its load is below the underload threshold value (Rieche *et al.* 2005, Zhu & Hu 2004, Rao *et al.* 2003).

The primary goal for load balancing in any distributed computer network is to prevent overload in any of the nodes in the network. This is done in order to maximize *performance* (Ferrari & Zhou 1987) and *efficiency* (Zhu & Hu 2004, 2005) of a distributed computer system by harnessing all available resources in the system, and to eliminate traffic bottlenecks (Shen & Xu 2008). The need for load balancing grows along with the overall load of a distributed computer system, since the probability of nodes getting overloaded grows. The performance of an overloaded node may be only a fraction of a properly loaded node, which also affects the system performance in addition to the node itself. Accordingly, when the system is lightly loaded, the probability of nodes getting overloaded is lower.

2.2.2 Special requirements for load balancing in P2P systems

In P2P systems, an additional requirement for load balancing is to ensure that the users have an incentive to provide the hardware capacity of their device for the use of the P2P overlay (Veciana & Yang 2003, Anceaume *et al.* 2005). According to this, load in P2P systems must be *fairly* balanced to keep the user experience on an acceptable level (Felber *et al.* 2014). The load originating from P2P service provisioning or maintenance should not, for example, overutilize the central processing unit (CPU) or the network interface while the user is actively utilizing the device for CPU-intensive networking actions. The same logic applies to other components as well.

P2P traffic patterns are significantly different from client/server based systems. P2P systems usually provide two fundamental functions: 1) Search and discovery of nodes, content, or services (also called *signalling*), and 2) Transferring content or information between the nodes in either real-time or non-real-time manner (also called *media transfers*). The most notable difference to client/server systems is the higher proportion of signalling traffic; in addition to extra signalling originating from node and resource discovery, the overlay maintenance generates frequent signalling traffic (Ou *et al.* 2010a, Vergara & Tehrani 2013, Vergara *et al.* 2014, Eronen 2008, Nurminen & Nöyränen 2008). Load balancing in P2P systems has to take

into account both signalling and media transfers in order to successfully balance the load experienced by a node.

In addition, the decentralized content storage model of P2P systems poses a major challenge for load balancing. Load balancing system must not generate too high a burden on the system. Particularly, the load movement cost originating from moving data objects between nodes should be minimized (Godfrey & Stoica 2005, Surana *et al.* 2006, Song & Yang 2006).

2.2.3 Sources of load imbalance in P2P systems

There are several possible sources for load imbalance in P2P systems. This Section introduces the most important of those.

Uneven distribution of objects among peer nodes causes uneven load for the peer nodes hosting the shared objects. In structured P2P systems, arriving nodes are placed into the overlay and leaving nodes are removed from the overlay in an organized manner (Androutsellis-Theotokis & Spinellis 2004, Brands & Karagiannis 2009). Uneven distribution of objects in structured P2P systems is a consequence of the properties of the used hash algorithm (Byers *et al.* 2002, Rao *et al.* 2003). In unstructured P2P systems, the nodes that own an object are also the primary hosts for them. The object is usually replicated to a set of neighbour nodes, based on some replication strategy (Cohen & Schenker 2002). The average number of the replicated objects (owned by other nodes) the node has to host is proportional with the number of links to the other nodes.

Churn is one of the sources for uneven distribution of objects (Zhu & Hu 2005, Ledlie & Seltzer 2005). In structured P2P systems, a leaving node generates imbalance in the object distribution by giving up its hosted objects and resources, as well as address space, to one of its neighbour nodes that may now get overloaded (Felber *et al.* 2014). In unstructured systems, the topology of the network changes arbitrarily due to churn. Due to this, the links between the nodes are also generated and removed in an uncoordinated manner when nodes join and leave the system. This generates load imbalance between the nodes (Lua *et al.* 2005, Yang & Fei 2009, Ciraci *et al.* 2009). The effect is not as severe in structured systems that inherently suppress this phenomenon, since the probability of a joining peer node to split the address-space (and load) of a node responsible of a larger proportion of the finite address space is higher than a node with a smaller proportion.

Uneven distribution of routing load among the peer nodes causes uneven load between them. It is a consequence of the properties of the used algorithms and protocols, as well as churn. In structured P2P systems, nodes that are either topologically located in congested network segments or those that have long uptime have more links with their neighbour nodes than the other nodes in the overlay (Rao *et al.* 2003, Surana *et al.* 2006, Bienkowski *et al.* 2005). Thus, load imbalance caused by uneven distribution of message routing responsibilities is present in structured P2P systems. With unstructured P2P systems, the arbitrary and uncoordinated nature of building the node topology leads to the situation where some nodes have more links with their neighbour nodes than other nodes. The nodes with a larger number of links are usually more heavily burdened by message routing responsibilities (Yang & Fei 2009, Ciraci *et al.* 2009). Unstructured P2P networks with flooding-based query routing are especially problematic in this sense, since nodes flood the incoming queries to all linked nodes.

Also the *differences in object properties*, such as popularity, size and type generate load imbalance in P2P systems (Soltani *et al.* 2012). This type of load imbalance is independent of whether the used P2P system is structured or unstructured. The actual request distribution in P2P applications roughly follows Zipf-like distribution, in which relatively few popular objects account for most of the requests (Hefeeda & Saleh 2008, Serbu *et al.* 2007). Consequently, popular data objects inflict higher routing load for the nodes near the object's location (Zhou & Koyanagi 2013). According to the object size, each object has a *movement cost*, which means the load caused by moving the object data from a device to another. This cost can be considered as being proportional to the object's storage size (Surana *et al.* 2006). However, also the object type affects the load inflicted on the involved nodes; continuous content, such as live media stream causes a different type of load to the involved nodes when compared to moving discrete data blocks.

In structured P2P systems, the *size of the address space interval a node hosts* varies, which causes load imbalance in the system. Most of the structured P2P systems form a finite address-space using consistent hashing scheme (Karger *et al.* 1997). Consistent hashing is used to hash both the object and node identifiers (ID) in a numeric form, after which they are placed in the address-space. The address-space is partitioned into address intervals that are taken care by the participating peer nodes, according to the DHT algorithm in use. If a node is removed, its address interval is taken over by another node. In the worst case, consistent hashing may result in $O(\log N)$ imbalance of object distribution between the peer nodes (Rao *et al.* 2003, Godfrey & Stoica 2005, Serbu *et al.* 2007). Since a node is responsible

for all requests targeted at identifiers belonging to its address space, the average number of incoming requests messages depends (along with node popularity) on the address space size (Rao *et al.* 2003). Larger address space also increases the message forwarding responsibilities, since the node is responsible for a larger part of the overlay.

Node heterogeneity is another major source of load imbalance in P2P systems, particularly in systems including mobile nodes. Even in networks with equally balanced service load some nodes may still be overloaded while others are very lightly loaded, since less capable nodes have less hardware resources to carry the network load (Kim & Park 2006, Rao *et al.* 2003, Surana *et al.* 2006, Zhu & Hu 2005). For instance, a desktop PC may have tens of times higher available network bandwidth, CPU processing capacity or storage capacity than a smartphone. However, both of these devices can still participate in the same P2P network.

2.2.4 Load balancing process

Load balancing can be described as a process consisting of three distinguishable components presented in Fig. 3. *Load monitoring* defines the load of a peer node and it is a mandatory component in any load balancing system, regardless of the used load type (service load or hardware load). The existing load monitoring methods are discussed in Section 2.2.5.

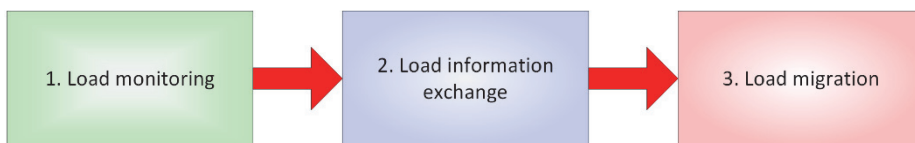


Fig. 3. Load balancing process.

When the load of each node is known, the load needs to be compared between nodes in order to trigger load balancing actions. For enabling this, load information has to be *exchanged* between the nodes subject to load balancing and the entity making the load balancing decisions. This entity may be a third-party node (server), an individual peer node in the overlay, or a group of collaborating peer nodes. The existing load information exchange methods are discussed in Section 2.2.6.

When the load balancing entity has sufficient information on the load of all or a part of the nodes, the entity is able to make load balancing decisions. The deci-

sions are made based on a load balancing algorithm, and realized using a *load migration* mechanism. Load migration is usually tightly coupled with the used load balancing algorithm. Thus, in this literature review, load balancing algorithm and migration are considered to be located on a single component. The current load migration methods are discussed in Section 2.2.7.

2.2.5 Load monitoring

To enable load balancing, the load of each participating node has to be measured and presented in a comparable format. A traditional method for defining the load of a node in a distributed system is to measure how much service load it carries. The common service load metrics include *service query rate*, *service time*, *number of active sessions*, or *total amount of transmitted bytes* (Cardellini *et al.* 1999, Bryhni *et al.* 2000). These metrics can be used alone or together with other metrics. Mitzenmacher (2001) and Breitgand *et al.* (2010), for instance, have presented load monitoring models based on a simple supermarket model that uses service query rate and service time as the load metric. Using the service load metrics, however, is feasible only when the load-balanced nodes are equal with regard to their hardware capacity. With heterogeneous-capacity nodes, equal service load inflicts different hardware load for devices with different hardware capacities. Thus, measuring hardware load is more feasible in most P2P networks, where nodes are expected to be heterogeneous with regard to their hardware capacity.

Measuring *CPU activity* is traditional and the most straightforward way to measure the hardware load in real-time, since most operating systems provide CPU load information for the applications. Different CPU load monitoring mechanisms have been evaluated in (Ferrari & Zhou 1987). CPU is one of the most significant physical components of a computing device affecting the system responsiveness, since basically all computing actions require CPU processing. High CPU load can, for example, reduce responsiveness to UI actions and cause general slowness in accomplishing different computing-intensive tasks. Responsiveness, in turn, is a centric factor of user experience, referring to a person's emotions and attitudes about using a particular product, system or service (Ickin *et al.* 2012, Doherty & Sorenson 2015).

Most applications put strain on a device also in the terms of its *network interface*, *memory* and *I/O capacity*. Since these metrics also have strong impact on the user experience (Ickin *et al.* 2012), they are as well good candidates to be used as load metrics. These metrics are provided by most operating systems in both mobile

and fixed domains. Many load balancing models use advanced load monitoring mechanisms that combine different load metrics obtained from the operating system. For example, a resource monitoring system utilizing measured information of a node's CPU and network interface load is proposed by Faik (2005). Another model, taking memory load also into account, is presented by Putrycz (2003). It also considers the number of running threads on a node in addition to CPU load. Furthermore, Quin (2008) has proposed a load monitoring mechanism taking into account I/O, CPU and memory load.

Another way to monitor a node's hardware load is to measure its runtime performance. Bryhni *et al.* (2000) have proposed a model where the round-trip time of a test request is used to measure a node's load. The round-trip mechanism is a useful way to detect overload of a node. However, before the round-trip time begins to significantly grow, the load of a device has usually grown to such an extent that the node's performance is already significantly suffering from it. Thus, the method is more useful for overload prevention instead of load balancing in P2P systems.

The presence of mobile devices generates additional requirements for load balancing. Since mobile devices are most of the time running on their batteries, battery life is one of the centric factors of long-term user experience of mobile devices (Ickin *et al.* 2012, Kujala *et al.* 2011, Heikkinen & Nurminen 2010). Load balancing can be used as an effective measure to increase the battery life of mobile devices participating in P2P networks. However, monitoring battery or energy consumption status as a load factor has not been investigated in the literature, although several studies concerning energy consumption profiles of different wireless device types exist. Paper III of this thesis (Harjula *et al.* 2012) introduces an advanced model, called e-Aware, for estimating how the application and protocol properties affect the energy consumption of mobile terminals operating in 3G and WLAN networks. Later, a fundamentally similar model called EnergyBox was proposed by Vergara & Tehrani (2013) and Vergara *et al.* (2014). In an earlier study, Mahmud *et al.* (2004) have proposed a prediction model for energy-consumption for a personal digital assistant (PDA) computer with 3G and WLAN network interfaces.

2.2.6 Load information exchange

To enable dynamic load balancing, nodes need to exchange their load information. Two main types of load information exchange can be identified, *uncoordinated* and *coordinated* load distribution that are illustrated in Fig. 4.

Uncoordinated load exchange mechanisms are based on comparisons of load between two or more nodes in a *point-to-point* manner (Fig. 4A). The comparison of load between nodes can be performed by either 1) arbitrarily during runtime, or 2) during the execution of some overlay operation, such as join or object lookup query. The first method is used in the *one-to-one* scheme in Rao *et al.* (2003). Here, each non-overloaded node periodically picks a random ID and performs a lookup operation to find the node responsible for the ID and review its load status. If the responding node is overloaded, load migration may take place between the two nodes. The second method is used by Byers *et al.* (2002), where a node storing an object calculates multiple hash codes (ID candidates) for the object using different hash functions, queries the load of the responsible nodes for these hash codes, and finally stores the object to the least loaded of the nodes. Boukhelef & Kitagawa (2009) use an overlay network's periodic routing table maintenance messages to piggyback load information between both neighbouring and distant nodes. The advantage of this method is that it does not require additional messaging dedicated to load balancing.

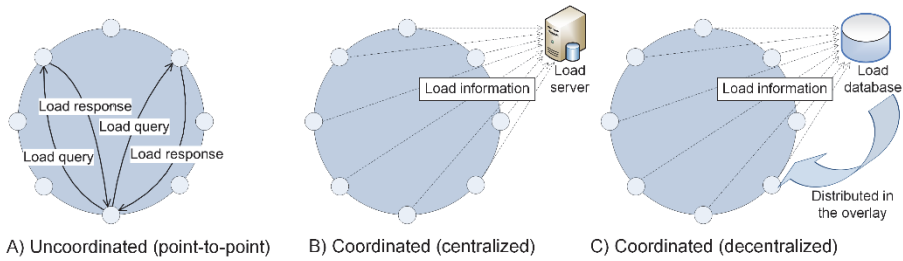


Fig. 4. Main types of load information exchange.

Coordinated load information exchange can be implemented either in a *centralized* (Fig. 4B) or a *decentralized* (Fig. 4C) manner. In centralized load information exchange, all nodes periodically report their status to a load server, which has full information about all nodes in a centralized load index and governs the needed load balancing actions. Rao *et al.* (2003) use a centralized load information exchange system as an option to implement a centralized load balancing scheme, called many-to-many. In decentralized load information exchange, the system maintains a distributed load index, in which the load status of each node is periodically stored. Furthermore, Rao *et al.* (2003) use a decentralized load exchange mechanism, called one-to-many, as an alternative to the centralized load exchange mechanism.

In this mechanism, an overloaded node uses a distributed load index to find a set of non-overloaded nodes and information about their load, and then migrates load from itself to one or more non-overloaded nodes. The load index is distributed using the overlay itself, and contains load information of a limited set of nodes in the overlay.

Furthermore, the *integrity* and *freshness* of load information must be ensured. Without guaranteed integrity and freshness, large scale network monitoring systems may provide misleading load information. An example of a system that helps ensuring the consistency and integrity of load information is introduced by Jain *et al.* (2008). The system provides a DHT-based consistency metric for query results in dynamic, large-scale monitoring systems. In the system, an overlay monitors its own state so that, in addition to other attribute values, the queries also return information on the number of nodes whose recent updates may be missing and whose inputs may be double counted due to overlay reconfigurations. When needed, the system provides notification that the query results should not be trusted.

2.2.7 Load migration

When a sufficient amount of load information is available, load can be dynamically balanced among the participating nodes. Dynamic load balancing consists of a decision algorithm and a load migration mechanism. As the decision algorithm is in most cases tightly bound with the used migration mechanism, they are discussed together in this section. There are several ways to migrate load in P2P systems. This section introduces the most common of them.

Object replication

Object replication means storing copies of the data object outside the primary hosting node. It is used in P2P systems in order to improve the availability, fault tolerance and performance through redundancy of data. Replication also improves the search performance, since the nodes hosting the replicas can also return the object to the requestor.

According to Androutsellis & Theotokis (2004), the following replication strategies are used in P2P systems. In *passive* replication, content is replicated by several nodes requesting and copying content from one to another. In *cache-based* replication, content is copied in intermediary nodes as it passes through nodes in the network. In *active* replication, different external methods are employed to improve

locality and availability of data, as well as performance. *Introspective* replica management techniques monitor traffic and requests in a P2P network, and create replicas of content objects to accommodate the resource demand. In *dynamic* replica management, client latency and server load are optimized by primarily using existing replicas that meet the client latency constraint without being overloaded, and, secondarily, placing new replicas.

Controlling the number and location of object replicas is also one of the most basic load migration mechanisms, due to its applicability to most P2P systems. Replication of an object among several peer nodes distributes the load that is generated by retrieving the object. Replication is, however, challenging for structured systems where the location of objects is directly bound to the object identifiers (Androutsellis & Theotokis 2004). In some cases, the use of aliases is employed to allow replication (Stoica *et al.* 2001). Another method is to use a set of random keys (Rhea 2005, Ratnasamy *et al.* 2001), for storing replicas at diverse locations in the address space. The general drawback of replication, in both structured and unstructured P2P systems, is the increased overlay data management overhead requiring more storage space and bandwidth from the peer nodes and the network.

Object fragmentation

In *object fragmentation*-based load balancing, a data object is typically divided into equal-sized blocks in order to balance the overlay load within a group of peer nodes that are simultaneously acquiring the same data object.

BitTorrent (Pouwelse *et al.* 2005) is a good example of a P2P system that uses object fragmentation. In BitTorrent the primary goal of object fragmentation is not to implement load balancing, but to enable fast and efficient distribution of large data objects by leveraging also the upload bandwidth of the downloading peer nodes (Bharambe *et al.* 2006). However, efficient file distribution also typically stands for efficient load balancing, since efficient use of network resources inherently takes into account the heterogeneous network capabilities of the peer nodes. Each shared data object in BitTorrent is advertised as a torrent. A torrent is a small file containing both the information about the blocks of data and the contact information of a tracker that is responsible for coordinating the distribution of the data object (Pouwelse *et al.* 2005). Depending on the version of BitTorrent, a tracker can be either reside in a centralized server or in a DHT-based P2P system. When a DHT-based P2P system is used, a torrent file contains an address in the DHT-based

P2P network, where the list of peer nodes belonging to a particular file sharing group is located.

Address space balancing

Address space balancing is one of the most basic load balancing schemes for structured P2P systems. It aims to divide the overlay address-space equally between the participating nodes. The *consistent hashing* scheme (Karger *et al.* 1997), on which most of the random hash functions used in DHTs are based, inherently provides a simple address space balancing mechanism. The schema hashes both objects and nodes using the same hash function. If a node is removed, its address interval is taken over by a node with an adjacent address interval. The key characteristic of the mechanism is that the IDs of the other nodes remain unchanged, reducing the overhead in the overlay. However, in the worst case, consistent hashing may result in $O(\log N)$ imbalance of object distribution between the peer nodes (Rao *et al.* 2003, Godfrey & Stoica 2005, Serbu *et al.* 2007).

A load balancing model aiming to improve the address space balance of P2P systems together with an object location controlling model has been proposed by Karger & Ruhl (2004). The main idea of the model is that each device has a fixed set of $O(\log N)$ possible locations, virtual nodes, on the overlay, calculated using different hash algorithms. In each device, only one of the virtual nodes is active at a time. The virtual node that spans the greatest address space between its own address and succeeding active virtual node, is selected as the active virtual node for the device. The selection is made occasionally. With high probability, each device will be responsible for $O(1/N)$ fraction of the address space.

Virtual servers

An advanced load balancing scheme for structured P2P systems, *virtual servers*, was originally introduced with Chord DHT algorithm (Stoica *et al.* 2001). In a virtual server model, a single physical node may host several virtual servers that are all participating independently in the same overlay. The virtual server scheme should not be mixed with the previously introduced virtual node scheme, which is based on a fundamentally different concept. For the overlay network, virtual servers appear as they were real devices, and their location on the overlay is calculated

using the primary overlay hash function. The basic concept of multiple virtual servers per physical node balances the load statistically without any active migration mechanisms.

The virtual server scheme is most effective when the number of active virtual nodes per node can be changed (Stoica *et al.* 2001). This model enables resource-aware migration of load into higher capacity devices, by allocating more virtual nodes for them than for lower-capacity nodes. The scheme can also be made *dynamic* by allowing nodes to dynamically exchange virtual servers between each other during runtime. Dynamic virtual server load balancing is proposed by Rao *et al.* (2003). In their model, after a node has defined itself as heavy and learned contact information for one or more light nodes, it defines the virtual server(s) that will be transferred to the light nodes with the condition that it will not make them heavy.

Many further improvements and optimizations have been proposed for the virtual server scheme, concerning, e.g. churn (Godfrey *et al.* 2005), peer proximity and locality (Zhu & Hu 2004, 2005), peer activity and object popularity (Bianchi *et al.* 2006). In these models, load information distribution and decision algorithms take into account the mentioned system characteristics. However, the load transfer is still made by moving virtual servers from heavily loaded nodes to lightly loaded ones, thus the fundamental load migration model remains intact.

Controlling node location

Controlling the node location on a structured P2P overlay can also be used for load balancing. *Node relocation*, proposed by Karger & Ruhl (2004), is a mechanism allowing non-overloaded nodes to relocate themselves to share the address-spaces the overloaded nodes are responsible for. In this model, a non-overloaded node queries the load of a randomly picked node and makes a pairwise load comparison between itself and the remote node. If the remote node is overloaded, the node relocates itself on the overlay to share the address space of the overloaded node.

Rieche *et al.* (2004) have presented another load balancing model that is fundamentally based on controlling node location in the address-space. In this model, the address space is divided into intervals containing multiple nodes. The nodes within the interval are collectively responsible for all of the objects stored in the interval. If the interval gets overloaded, it is either split to two or more intervals or the border between overloaded intervals is moved to balance the load between them.

In addition, several models optimizing the node location on the overlay based on, e.g. *peer proximity* and *locality* (Zhu & Hu 2004, 2005), or *peer activity* and

object popularity (Bianchi *et al.* 2006), have been proposed. These optimizations help minimizing latency and maximizing bandwidth by locating physically nearby nodes or nodes interacting actively between each other in logical proximity on the overlay.

Controlling object location

Yet another fundamental load balancing scheme for structured P2P systems is to *control the object location* on the overlay. *Power of Two Choices* (Byers *et al.* 2002), is a load balancing system based on the concept of using multiple (two or more) hash functions per object. When inserting an object to an overlay, all respective hash values (ID candidates) are computed and the load information of each corresponding node candidate (a node that would store the object with certain ID) is retrieved. Finally, the object is stored at the peer node with the lightest load. The objects are queried by sending separate lookups using each respective ID candidate.

Controlling overlay bindings

Structured P2P systems are based on fixed and temporary bindings between peer nodes. The fixed bindings, established by DHT algorithms, are used to constitute the overlay topology, whereas the temporary bindings are used in optimizing the routing efficiency. Temporary bindings are usually generated when nodes cache the recently contacted nodes that are not in their primary (DHT-established) routing table. Controlling the generation and termination of overlay bindings can also be used for load balancing purposes.

This scheme is used in *Advanced finger selection* (AFSA) load balancing model (Hautakorpi & Mäenpää 2010). The main idea of the model is that each out-finger (outgoing binding) is selected from a set of out-finger candidates in a manner that results to an even distribution of out-fingers in a structured P2P network. There are two different modes in AFSA: *implicit* and *explicit*. In the implicit mode, the probability of the node getting selected as an out-finger decreases as number of the in-fingers increases. In the explicit mode, the selection between out-finger candidates is done by accepting an out-finger candidate that has the lowest number of in-fingers. The simulations show that an uneven distribution of incoming bindings correlates with an uneven distribution of signalling load. In addition, as the number of incoming requests directly affects the media transfer load, controlling the generation and termination of overlay bindings also affects the media transfer load.

Hierarchical network architectures

Hierarchical architectures (Korzun&Gurtov 2011, 2013, Koskela *et al.* 2013, Ou 2010, Artigas *et al.* 2005) can also be used for load balancing in both structured and unstructured P2P systems. As described in Section 2.1.1, the basic principle is to select a subset of nodes as super-peers, based on their hardware capacity and stability. Super-peers are peer nodes with a larger set of functionalities, providing, e.g. a gateway functionality to other peer nodes lower on the node hierarchy and thus carrying more load on behalf of them. Super-peers can also act as super-peers for each other, and thus generate multi-layer architectures.

Super-peer architectures are particularly suitable for resource-aware load balancing among nodes in a same subnet. *mDHT* (Lee *et al.* 2009b) is an enhancement for structured P2P networks, where the nodes of a subnet appear in a DHT overlay as a single node. Queries are routed as usual until they reach the destination subnet, where they are resolved among the hosts using multicast. The model removes the single point of failure and balances the load between subnet nodes. However, the multicast-based routing between subnet nodes is resource-consuming and the model does not inherently provide dynamic load balancing between nodes, although the use of additional load balancing functionalities is possible. Another model using node hierarchy for load balancing is proposed by Bianzino *et al.* (2014). In this proposal, a proxy node is connected to the rest of the network, e.g. a P2P overlay, using a high-capacity and high-energy radio interface. The other nodes in the group are connected to the proxy node using a low-energy radio interface. The model is shown effective in reducing the energy consumption of the ordinary nodes by an empirical evaluation. The model is generic, i.e. not P2P-specific, but is well suitable for P2P context in balancing the load between nodes in a same subnet, making it also a suitable approach for e.g. M2M/IoT networks.

2.3 Energy consumption in mobile networking

Mobile devices are powered with batteries that are limited in their capacity due to the constraints on size and weight of the device. As a consequence, energy-efficiency is emphasized in mobile computing. A core requirement for designing energy-efficient mobile systems is to understand where and how the energy is consumed. This section presents a survey of the literature on energy consumption characteristics of different mobile device types (Section 2.3.1), wireless communication technologies (Section 2.3.2) and higher-layer optimization methods (Section 2.3.3).

Lastly, Section 2.3.4 reviews the existing work on analysing the effect of P2P networking traffic patterns on the energy consumption of different mobile devices.

2.3.1 Decomposition of energy consumption in mobile devices

The main components of mobile devices are CPU, communication and signal processors, memory and storage, display and graphics hardware, audio, user input devices, camera and various networking interfaces (Pathak *et al.* 2012, Carroll & Heiser 2010, Shye *et al.* 2009). These can be classified into four main categories: processing, storage, input/output and communication hardware (Fig. 5).

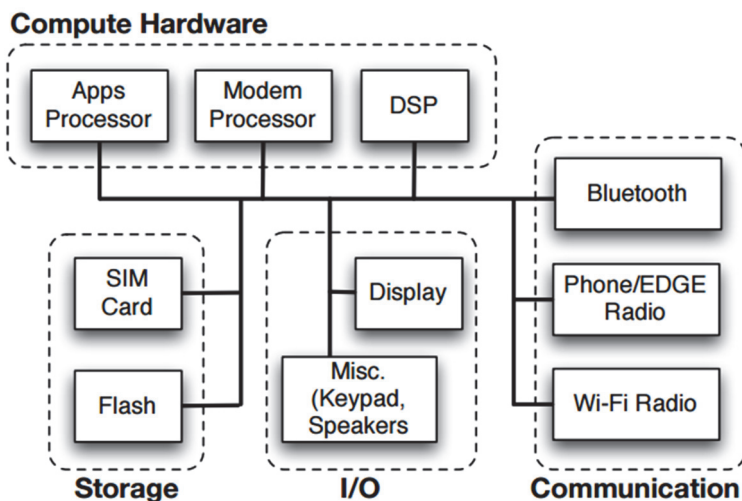


Fig. 5. Overview of a mobile computing architecture (Shye *et al.* 2009) © 2009 ACM, Reprinted by permission.

The distribution of energy consumption between different hardware components depends heavily on the device type and applications (Shye *et al.* 2009). For instance, in a typical use of a smartphone or a tablet computer, the display and the network interfaces are the most energy consuming components (Pathak *et al.* 2012, Shye *et al.* 2009). Fig. 6 illustrates the long-term power consumption breakdown for a set of personal smartphones used by real-life users. As can be seen, the idle consumption, which mostly originates from the network interfaces maintaining the connection, changing access points when needed, etc., dominates the power consumption

(around 50% on average). The active network usage (call, wifi and edge in the figure), display (screen_on, brightness) and computing (cpu, system) come next, consuming roughly similar amounts of energy, around 15-20% each.

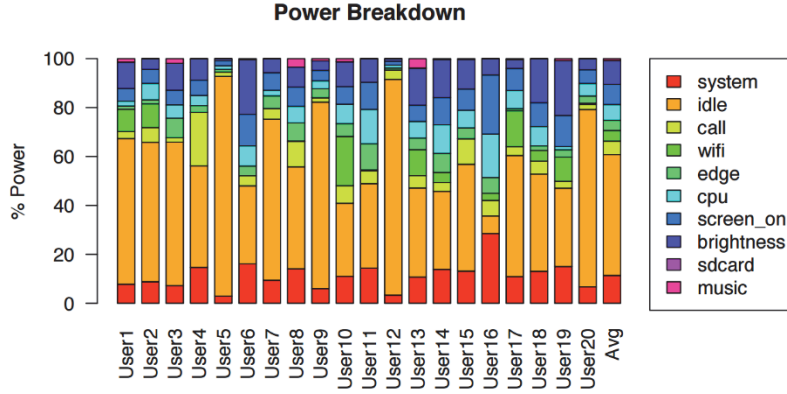


Fig. 6. Power consumption breakdown of a smartphone (Shye *et al.* 2009) © 2009 ACM, Reprinted by permission.

The distribution of energy consumption between the hardware components in wireless M2M nodes differs significantly from smartphones and other mobile devices. Since display is typically not present, the network interface dominates even more clearly the energy consumption (Jeong *et al.* 2011, Zhang *et al.* 2011, Lu *et al.* 2011, Xing *et al.* 2007). Fig. 7 presents the typical power consumption breakdown of a wireless sensor network (WSN) node. As can be seen, the radio interface consumes more than 80% of the total power consumption. In wireless multimedia sensor networks (WMSNs), the sensing and processing of multimedia data (sound, video) require higher-performance computing and communication hardware that in turn consumes more energy (Aziz & Pham 2013, Akyildiz *et al.* 2007). Depending on the application and the type of the multimedia data, the majority of the energy is consumed either on the processing architecture or the network interface.

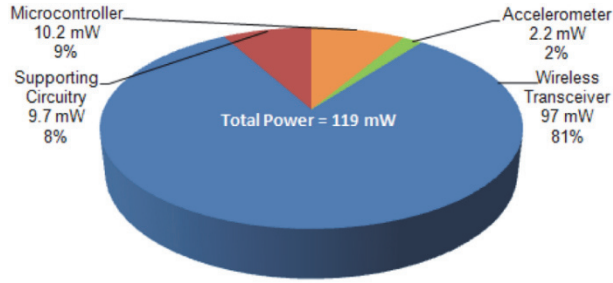


Fig. 7. Power consumption breakdown of a typical wireless sensor node (Zhang *et al.* 2011).

As shown above, most of the energy consumption in mobile devices originates from network interfaces. Furthermore, the majority of the burden from P2P networking falls on the network interface (Kelenyi & Nurminen 2008a, 2008b). Based on this, the emphasis of the following sections is on the energy consumption characteristics and optimization technologies for communication technologies.

2.3.2 Wireless communication technologies

Different wireless communication technologies have different power consumption characteristics. This section focuses on the energy consumption characteristics of 1) cellular networks and wireless local area networks that are typically used for mobile interpersonal communication, and 2) short-range networks that are typically used for providing low-power connectivity for M2M networking.

Cellular networks

In cellular networks, low latency, high throughput and support for high mobility are important design features, aiming to achieve high-quality user experience. The energy saving mechanisms have been optimized for minimizing the energy consumption during long idle periods and ensuring low latency and high throughput during intensive network usage.

The energy consumption of the 3G radio interface is mostly influenced by the radio resource management performed at the network operator side by the Radio Network Controller (RNC) (Vergara & Tehrani 2013, Vergara *et al.* 2014, Perälä *et al.* 2009, Haverinen *et al.* 2007). The RNC uses the Radio Resource Control (RRC) and the Radio Link Control (RLC) protocols. RRC includes three states: IDLE,

CELL_PCH (paging channel, alternative to IDLE), CELL_DCH (dedicated channel), and CELL_FACH (forward access channel). IDLE and CELL_PCH states provide low energy consumption for idle periods, but the latency is relatively high in this state. CELL_DCH state gives maximum throughput with minimum delay at the cost of high energy consumption. In CELL_FACH, the energy consumption is reduced at the cost of lower throughput, when compared to CELL_DCH. Changing from the idle state to either of the data transfer states requires a setup time. The additional energy consumed during the setup time is called *ramp energy* (Balasubramanian *et al.* 2009). After the completion of a data transfer, the radio link remains in a data transfer state for a while waiting for further data transfers, before moving to idle state. The additional energy consumed during this time is called *tail energy* (Balasubramanian *et al.* 2009). The transitions between states are controlled using inactivity timers and RLC data buffer thresholds set by mobile operators. Fig. 8 presents the power consumption in an example data transfer scenario and the state machine describing the different states and transitions between them in the 3G RRC protocol.

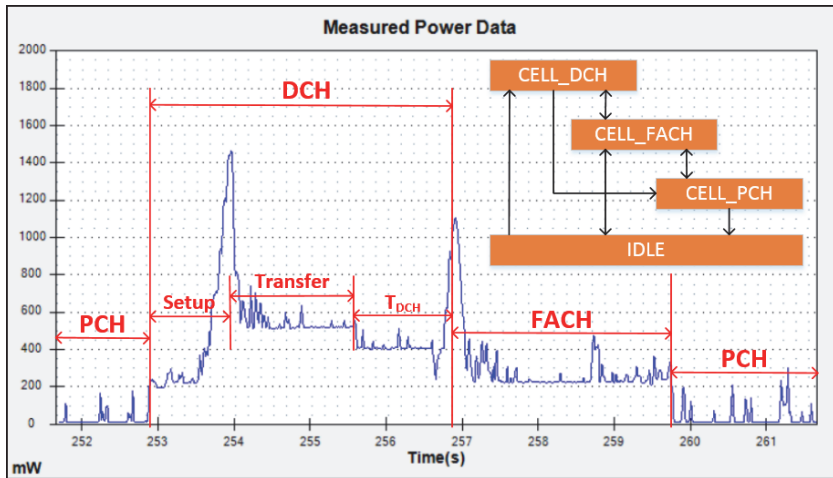


Fig. 8. Power consumption profile and state machine of 3G radio interface.

The 4G technology (also known as 3G LTE) is an evolution version of 3G technology. Although the basic characteristics of the 4G radio interface are rather similar to 3G, the power saving mechanism is more sophisticated (Bontu & Illidge 2009, Koskela & Vajtus-Anttila 2015, Hoque *et al.* 2015). In 4G, RRC provides three main states: RRC_IDLE, RRC_CONNECTED and RRC_DORMANT. In

RRC_IDLE state, the radio is in a low-power state, only listening to control traffic. RRC_CONNECTED (or ACTIVE) state is used to transmit data or to listen to incoming data. In this state, network resources are allocated for the device to give maximum throughput with minimum delay at the cost of high energy consumption. RRC_DORMANT state has the following sub-states: Short discontinuous reception (DRX) and Long DRX. In RRC_DORMANT state, no specific network resources are allocated for the device. Wireless radio is most of the time off to save energy and is periodically turned on to check for new incoming traffic. With short DRX, the period is shorter, providing faster response time and with Long DRX, the period is longer, providing lower energy consumption. The transition from RRC_CONNECTED state to RRC_IDLE state happens through Short DRX and Long DRX states. The DRX timer values can be adjusted according to the pattern of the incoming traffic. According to Siekkinen *et al.* (2013), the typical transition time from the RRC_CONNECTED state to Short DRX occurs in only a few seconds. This significantly improves the energy consumption of 4G as compared to 3G, especially in scattered signalling scenarios. In addition, 4G provides better energy per bit ratio despite the higher energy consumption in RRC_CONNECTED state (Huang *et al.* 2013). Fig. 9 presents the power consumption in an example data transfer scenario and the state machine describing the different states and transitions between them in the 4G RRC protocol.

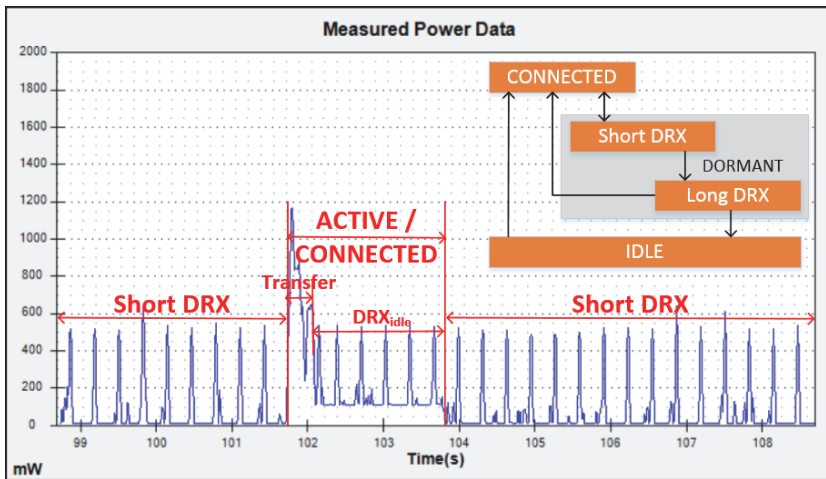


Fig. 9. Power consumption profile and state machine of 4G radio interface.

Wireless local area networks

In IEEE 802.11 WLAN networks, the energy saving mechanisms have been optimized for minimizing energy consumption in active network usage. The main principle is to avoid negative effects on performance, particularly the long response times that are present in older cellular network technologies.

The transmission energy consumption for WLAN is mostly influenced at the driver level in the WLAN client (Vergara & Tehrani 2013, Vergara *et al.* 2014). In *constantly awake mode* (CAM), the power-saving features are disabled to achieve the best performance. The *power save mode* (PSM) allows clients to switch to low power mode for predefined periods when not transferring any data, similarly to DRX states in 4G RRC protocol. In this mode, the access point buffers downlink frames for the clients and clients wake up periodically to receive buffered frames from the access point using Power Save Poll (PS-Poll) message.

Zhang & Shin (2012) have shown that idle listening is the dominant source of energy consumption in WLAN. The PSM mode reduces the energy consumed by idle listening by sleep scheduling. However, through a real-world traffic analysis, Zhang & Shin (2012) found that more than 60 percent of energy is consumed in idle listening, even with PSM enabled. Thus, further energy consumption optimization methods have been developed. *Adaptive PSM* is a common mechanism to overcome the overhead and latency drawback of using PS-Poll mechanism (Krashinsky & Balakrishnan 2002, Pyles *et al.* 2012). In adaptive PSM, the state transitions between CAM and PSM is made based on heuristics taking into account, e.g. the traffic profile or device display mode (on/off). Fig. 10 illustrates the power consumption profile and different states of a WLAN radio interface with adaptive PSM activated.

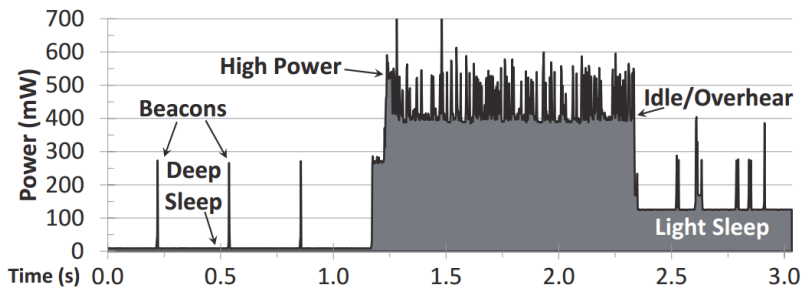


Fig. 10. Power consumption profile and different states of WLAN radio interface with adaptive PSM (Manweiler & Choudhury 2011) © 2011 ACM, Reprinted by permission.

In addition to activity within a single AP cell, the network contention among different APs can dramatically increase a client's energy consumption, as shown by Manweiler & Choudhury (2011).

Short-range networks

Due to extensive battery life requirements, low energy consumption is the dominant design principle in M2M networks that include wireless sensor networks (WSN), wireless sensor-actuator networks (WSAN) and wireless multimedia sensor networks (WMSN) (Xing *et al.* 2007, De Mil *et al.* 2008). Since network interface dominates the energy consumption of most M2M nodes, the industry has developed several low-power short-range wireless network protocols. The goal for these protocols is to provide ultra-low power connectivity for M2M devices that usually do not require high throughput or fast response time.

IEEE 802.15.4 is a largely adopted protocol specifying the physical (PHY) and medium access control (MAC) layers for low data rate wireless personal area networks. It is used as a basis for many low-power short-range communication protocols, such as ZigBee, WirelessHART and ISA100.11a. The standard introduces a beacon-enabled mode that allows energy to be saved by implementing duty cycling, so that all nodes can periodically go to sleep (Rault *et al.* 2014). It uses a slotted carrier sense multiple access/collision avoidance CSMA/CA carrier sensing (IEEE TG 15.4 2006) to access the shared channel (Gao *et al.* 2008, Rault *et al.* 2014).

6LoWPAN is an open standard defined in RFC 6282 by the IETF for IPv6 over M2M networking, including the required encapsulation and header compression mechanisms. It provides a networking technology and adaptation layer that allows IPv6 packets to be carried efficiently within small link layer frames, such as those defined by IEEE 802.15.4 (Olsson 2014).

ZigBee defines a communication protocol set based on the IEEE 802.15.4 standard (Siekkinen *et al.* 2012, Dementyev *et al.* 2013, Ayoub *et al.* 2011). It also uses the CSMA/CA carrier sensing to access the shared channel. The specification includes a packet-switched radio protocol for low-cost battery-operated devices. The protocol features include support for multiple networks, low duty cycle, low latency and reliable communication including collision avoidance, retries and acknowledgements. ZigBee supports both mesh and star network topologies and scales up to 65,000 nodes per network. The maximum range of a radio link varies between 10 and 100 meters. The over the air data rate is 250kbit/s. The original Zigbee specification has later been enhanced by, e.g. the ZigBee IP specification

that enables end-to-end IP-based networking by implementing 6LoWPAN specification. The relative energy consumption of Zigbee in comparison to its main competitors has been evaluated by Siekkinen *et al.* (2012) and Dementyev *et al.* (2013). The studies showed that the energy consumption of ZigBee in idle state ranges from 52 μ W (120s sleep interval) to 1.45mW (5s sleep interval), the active state consumption is around 31mW, and data transfer consumption ranges from 3.3 to 9 μ J per byte when throughput varies between 8 bit/s and 80 kbit/s.

Bluetooth Low Energy (BLE) is a low energy version of Bluetooth, specified in version 4.0 (Siekkinen *et al.* 2012, Bluetooth 2014, Gomez *et al.* 2012). The specification includes PHY layer, link layer (LL), logical link control and adaptation protocol (L2CAP), generic attribute protocol (GATT) and generic access profile (GAP) protocols. BLE uses adaptive frequency hopping spread spectrum to access the shared channel. In contrast to ZigBee, BLE supports only star topology. The number of simultaneous slaves per master node varies between 2 and 5,917, depending on, e.g. the used connection interval and data throughput (Gomez *et al.* 2012). BLE provides considerably reduced power consumption and cost but maintaining communication range similar to the original Bluetooth (Siekkinen *et al.* 2012). According to Siekkinen *et al.* (2012) and Dementyev *et al.* (2013), the energy consumption of BLE in idle state is around 33-627 μ W with 5-120s sleep interval, around 15mW in active state, and 1.9-3.5 μ J per byte in data transfers with varying throughput (48 bit/s – 800 kbit/s) and connection event size (1-4 frames).

Z-Wave, promoted by Z-Wave Alliance, is a proprietary wireless protocol for automation in residential and light commercial environments (Gomez & Paradells 2010). Z-Wave uses CSMA/CA carrier sensing to access the shared channel. The protocol supports both star and mesh network topologies and the maximum number of nodes per network is 232. The data transmission rates vary between 9.6 and 40 kbit/s and maximum radio link range varies from 30 to 100 meters. Similarly to ZigBee and BLE, Z-Wave provides low-power sleep modes and the protocol is very energy-efficient. The energy consumption figures are close to those of ZigBee.

ANT is a proprietary short-range radio protocol, primarily targeted to the sports and fitness performance monitoring (Khssibi *et al.* 2013). It is widely adopted by sports computer manufacturers. ANT communication is based on bi-directional channels, allowing a wide range of topologies to be built upon it. As with previously presented protocols, ANT provides low-power sleep modes and the protocol is very energy-efficient. According to Dementyev *et al.* (2013), its performance and energy consumption characteristics are close to ZigBee.

2.3.3 Optimization methods above radio interface

In mobile networking, several higher-layer energy consumption optimization methods have been developed in order to obtain maximal benefit of the energy-saving features of the underlying wireless communication technologies. These methods include 1) Network topology and routing optimization, 2) advanced sleep/wakeup schemes, 3) data reduction, 4) parallel data transfers and 5) computational offloading (Rault *et al.* 2014, Nurminen 2010, Nurminen & Nöyränen 2009, Kumar *et al.* 2013, Kosta *et al.* 2012, Saarinen *et al.* 2012, Kumar & Lu 2010). Fig. 11 provides a taxonomy of optimization methods above radio interfaces.

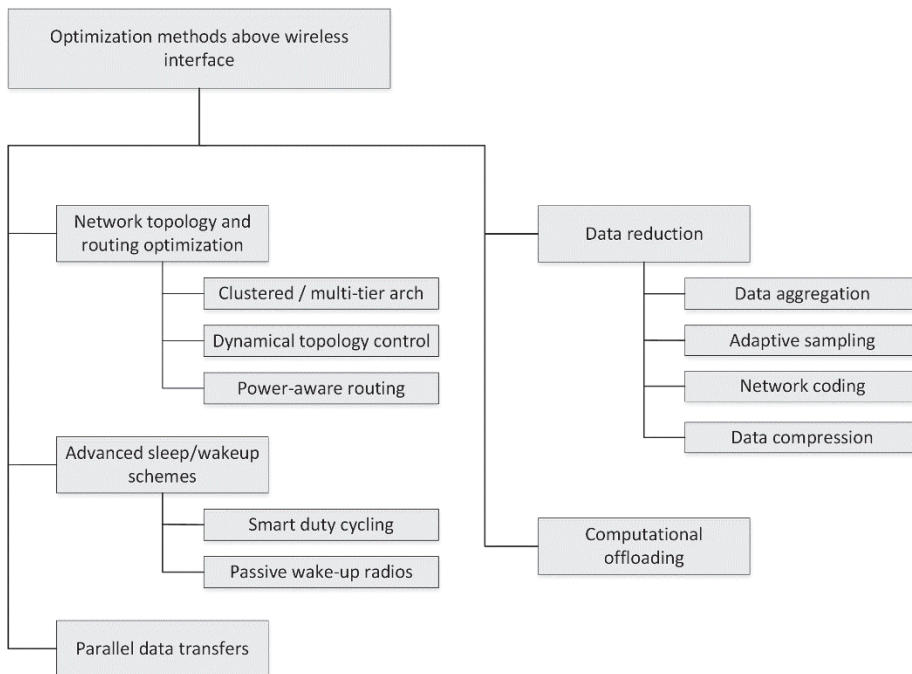


Fig. 11. Taxonomy of energy consumption optimization methods above radio interface.

Network topology and routing optimization

Network topology and routing optimization is a set of methods for optimizing the use of short-range radio networks in M2M networking. One of the methods aims to minimize the physical distance between communicating nodes, e.g. by *utilizing*

clustered and hierarchical architectures so that minimal transmission power can be used between the nodes (Rault *et al.* 2014, Xing *et al.* 2007, Akyildiz *et al.* 2007, Bianzino *et al.* 2014, Pering *et al.* 2006). Another method is to *exploit redundancy to dynamically adapt the network topology* based on the application's needs, in order to minimize the number of active nodes (Rault *et al.* 2014). Similarly, power-aware routing protocols allow choosing minimal appropriate transmission ranges and optimal routes for multi-hop packet transmission to reduce energy consumption (Xing *et al.* 2007, Lee *et al.* 2009a).

Advanced sleep/wakeup schemes

Advanced sleep/wakeup schemes above the radio interface aim to reduce energy consumption of a sensor or actuator node in M2M networking by allowing them to enter a sleep mode while not active. These schemes include *smart duty cycling* (activity scheduling) (Lu *et al.* 2011, Jurdak *et al.* 2010), and *passive wake-up radios* (Rault *et al.* 2014, Ba *et al.* 2013). In smart duty cycling, nodes periodically wake up to listen to messages from neighbour nodes and return to a sleep mode if no activity is detected. There are three types of duty cycling (Rault *et al.* 2014); 1) *on-demand*, where the node is waken up only when another node wants to communicate with it, 2) *scheduled rendezvous*, where nodes wake up periodically according to a wakeup schedule, such as staggered wakeup, agreed between the nodes in a cluster, and 3) *asynchronous*, where each node wakes up independently but its active period is adjusted to overlap with its neighbours. Passive wake-up radios, such as RFID tags (Ba *et al.* 2013), are used to trigger the wake up of a M2M node and its main radio interface only when needed. In this scenario, the primary radio interface, such as ZigBee, BLE, etc. can be switched off when the node is not active. This improves energy-efficiency by avoiding unnecessary radio interface activity that is present with duty cycling.

Data reduction

Data reduction schemes aim to minimize energy consumption of a mobile node by reducing the amount of data to be delivered. There are multiple ways to reduce the amount of data, including *data aggregation*, *adaptive sampling*, *network coding*, and *data compression* (Rault *et al.* 2014), as well as *lightweight communication protocols* (Bormann *et al.* 2012). In data aggregation, nodes reduce the amount of data between the source and sink nodes by performing data fusion (Rajagopalan &

Varshney 2006). Adaptive sampling techniques optimize the sampling rate of sensor nodes to meet the application requirements with minimum energy consumption (Anastasi *et al.* 2012). Network coding is a method to reduce the broadcast traffic by sending linearly combined packets instead of separately sending a copy for each packet (Wang *et al.* 2011). Data compression is a method to minimize the amount of traffic by encoding information in a manner that the number of bits needed to represent the initial message is reduced (Kimura & Latifi 2005). Lightweight application-layer protocols, such as Constrained Application Protocol (CoAP) (Bormann *et al.* 2012) aim to reduce the communication overhead of application-layer protocols, in order to reduce the amount of data to be delivered. CoAP is an open application-layer standard, considered as a lightweight version of HTTP for IoT communication.

Parallel data transfers

Parallel data transfers (Nurminen 2010, Nurminen & Nöyränen 2009) are a method for improving energy-efficiency of mobile networking by coordinating the communication of applications running concurrently in a same node in a manner that favours parallel data transfers. Parallel data transfers improve energy-efficiency of the radio interface by lengthening idle periods and reducing the effect of ramp and tail energy (Balasubramanian *et al.* 2009), described in Section 2.3.2. According to Nurminen (2010) and Nurminen & Nöyränen (2009), significant energy savings can be achieved by parallel data transfers in comparison to sequential transfers.

Computational offloading

Computational offloading is a method for saving energy in a mobile node by moving computationally intensive tasks to be executed on another node, such as a server or a cloud platform. Kumar & Lu (2010), Kumar *et al.* (2013), Kosta *et al.* (2012) and Saarinen *et al.* (2012), among others, have studied the potential energy saving that can be achieved by computational offloading. According to them, offloading can be considered beneficial from the viewpoint of energy consumption if large amounts of computation are needed with relatively small amounts of communication. If the need for communication increases, the energy-efficiency of computational offloading decreases. The break-even point depends on the energy consumption characteristic of the used communication technology.

2.3.4 Energy consumption in mobile peer-to-peer networking

In P2P networking, a major part of traffic consists of frequently delivered small messages, typically encapsulated in a single UDP datagram or Transmission Control Protocol (TCP) segment (Stutzbach & Rejaie 2006, Kelenyi & Nurminen 2008a, 2008b). This concerns both structured and unstructured P2P networks. In structured P2P networks, the signalling consists of overlay maintenance, request and response messages (Kelenyi & Nurminen 2008a, 2008b). Keepalive is the dominant maintenance message type (Hautakorpi & Mäenpää 2010, Ou *et al.* 2010a), used to keep the overlay's neighbour tables and network address translation bindings, among others, up to date. In unstructured systems, the amount of maintenance messaging is lower since there are no coordinated overlay topologies to maintain. However, due to inefficient request routing, the burden originating from request messages and responses to them is high (Yang & Fei 2009).

Signalling

The effect of frequent signalling to the battery life of a mobile device has been evaluated by Eronen (2008) and Nurminen & Nöyränen (2008). The results show that keepalive signalling consumes significant amounts of energy when operating in cellular (2G and 3G) networks, regardless of the relatively small total amount of transferred data. This phenomenon is a direct consequence of the energy consumption profiles of wireless communication technologies, described in Section 2.3.2. Since especially the cellular radio interfaces remain in connected states for some time after sending or receiving data, the tail energy effect (Balasubramanian *et al.* 2009) is considerable. The length of this period is defined by different radio interface-specific timeout intervals that are set by the network operator. With signalling intervals below the connected state (CONNECTED in 4G or DCH in 3G) timeout interval, the radio interface remains continuously in connected state. With signalling intervals below the dormant state (DRX states in 4G or FACH in 3G) timeout intervals, the radio interface switches frequently between connected and dormant states never entering to the sleep state. Since the state machines for 3G and 4G technologies are different (Figs. 8 and 9), there are also differences in their energy consumption characteristics. According to Siekkinen *et al.* (2013), the typical transition time from CONNECTED to DRX states in 4G is significantly shorter than the typical transition time from DCH to FACH in 3G. Respectively, the typical

transition time from DRX states to IDLE state is significantly longer than the typical transition time from FACH to PCH or IDLE mode. However, as the power consumption of both DRX states is radically lower than that of FACH, the overall tail energy effect is usually higher in 3G networks. In addition, 4G provides also better energy per bit ratio compared to 3G (Huang *et al.* 2013). In conclusion, the energy-efficiency of 4G seems to be higher than that of 3G in scattered signalling scenarios.

With WLAN, mobile devices typically use Adaptive PSM mode where the cost of maintaining the association with the AP is small. While associated to the AP, the ramp and tail overhead is low and the energy consumed is proportional to the size of the data transfer and the transmit power level (Balasubramanian *et al.* 2009). In WLAN, also 3rd party traffic, depending on, e.g. the Adaptive PSM settings, causes more or less extra overhead in the radio interface (Zhang & Shin 2012, Manweiler & Choudhury 2011). Thus, the typical traffic patterns of P2P networking do not cause as significant degradation for energy-efficiency in WLAN radio interfaces, as with cellular radio interfaces.

With short-range radios, the communication overhead is typically low but latencies may be higher, depending, e.g. on how strict energy saving settings are used or what is the used network topology (Gomez *et al.* 2012, Ayoub *et al.* 2011). It can be assumed that P2P networking is possible also on short-range networks, provided that the full TCP/IP-stack can be used, the message sizes are small enough to fit in the small Maximum Transfer Units (MTU) of short-range networks and the used P2P protocol tolerates high latencies.

Full-bandwidth data transfers

When larger amounts of data are moved, the energy consumption is more proportional to the amount of data (Balasubramanian *et al.* 2009, Hoque *et al.* 2015). The ramp and tail overhead are still present, but their proportion to the total consumed energy gets lower as the energy consumed per data transmission gets higher. This is the case especially when full capacity of the network interface is used, e.g. when a big file is transferred over the wireless link.

Since short-range radio technologies typically provide low data transfer performance and relatively high energy per bit ratio, moving larger amounts of data over them is generally unfeasible. BLE is the only short-range wireless communication technology among the presented ones that can be considered to be used for moving data units exceeding a megabyte with a tolerable transfer time.

Data streaming

With data streaming, the radio interface is used for data transfer for only a part of its time. The energy consumption between data transfers depends on the timeout intervals after which radio is switched to an idle mode (Siekkinen *et al.* 2013, Hoque *et al.* 2015). For instance, using large buffer size allows radio interface to enter the idle mode between subsequent data transfers and thus reduces the energy consumption (Hoque *et al.* 2015). This is a widely used method in on-demand streaming applications. The maximum buffer size is, however, strictly limited by QoS requirements in many live streaming applications requiring real-time playback, such as telepresence or remote gaming applications.

As with full-bandwidth data transfers, moving larger amounts of data over short-range wireless communication technologies is generally unfeasible. BLE is the only short-range wireless communication technology among the presented ones that can be considered feasible for data streaming.

2.4 Summary

This chapter provided a literature review of the centric research results and existing technologies in the area of P2P networking, P2P load balancing and energy-efficient mobile communication.

P2P networking provides a decentralized, self-organizing and scalable alternative for traditional client/server-based networking. The drawback of P2P networking is the higher resource consumption at peer nodes in comparison to that of client nodes in client/server-based networking. Load balancing is a widely used method for controlling the distribution of load between P2P nodes based on their hardware capacity. In the context of mobile computing, it would be beneficial to use load balancing for distributing load between nodes in an energy-aware manner. For this, advanced load monitoring and distribution technologies taking into account the energy profile and energy status of P2P nodes are needed. The literature review discovered a gap in the literature in this area; several studies concerning energy consumption of different wireless device types exist but related work on energy-aware load monitoring and balancing in P2P context were not found.

Energy-efficiency is emphasized in mobile computing. Among the different components of mobile and embedded wireless devices, most of the energy consumption in communication applications originates from the network interface. The energy consumption of a mobile device in communication applications comprises

of the radio interface and the software and protocols above it. In cellular networks, the energy saving mechanisms have been optimized for minimizing energy consumption during long idle periods and ensuring low latency and high throughput during intensive network usage. The current main cellular technologies include 3G and 4G (3G LTE) wireless communication standards. In WLAN networks, the energy saving mechanisms have been optimized for minimizing energy consumption in active network usage and the main principle is to avoid a negative effect on performance, particularly the long response times that are present in cellular network technologies. The current main WLAN technologies are based on IEEE 802.11 standard family (a/b/g/n). In M2M networks, low energy consumption is a dominant design principle. Since network interface dominates the energy consumption of most M2M nodes, the industry has developed several low-power short-range wireless network protocols. The goal for these protocols is to provide ultra-low power connectivity for M2M devices that usually do not require high throughput or real-time response time. The current main short-range wireless protocols include 6LoWPAN/802.15.4, ZigBee, Bluetooth Low Energy, Z-Wave and ANT. The higher-layer energy consumption optimization methods include 1) routing and network topology optimization, 2) advanced sleep/wakeup schemes, 3) data reduction, 4) parallel data transfers and 5) computational offloading. These methods have been developed in order to obtain maximum benefit of the energy-saving features of the underlying wireless protocols.

To summarize, the challenges related to the participation of mobile nodes in P2P networks are twofold. On the one hand, the use of P2P technologies significantly increase the energy consumption of mobile devices, since the majority of energy consumption is caused by intense maintenance signalling, regardless of the amount of data moved. This is a consequence of high communication overhead for which especially the cellular protocols are prone to. The high communication overhead originates from the typical traffic patterns of P2P communication, including intensive maintenance and 3rd party signalling. On the other hand, while operating in P2P networks, the special characteristics of mobile devices cause problems related to the routing integrity, QoS properties and energy-efficiency on the system level. The resulting frequent leaving and joining of nodes increases the need for network maintenance signalling, which further increases the overhead. Therefore, the optimization of communication patterns and reducing signalling overhead are emphasized in P2P systems. Resource-aware load balancing is among the most effective methods for addressing these challenges.

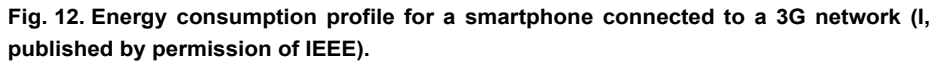
3 Research contributions

This chapter summarizes the contributions of the original publications with respect to the three research questions laid out in Section 1.2. The presentation is structured according to the four main contributions of the thesis as follows. Section 3.1 presents the empirical study for identifying the most essential obstacles for utilizing P2P technology in mobile environment. Section 3.2 presents the e-Aware energy consumption model developed for estimating how application layer properties affect the energy consumption of mobile devices. Section 3.3 presents the e-Mon load monitoring model for enabling energy consumption-aware load balancing utilizing existing load balancing models. Section 3.4 presents the ADHT, a mobile agent-based virtual peer concept that allows the participation of constrained-capacity devices in a P2P overlay without compromising their battery life.

3.1 Empirical feasibility analysis of mobile P2P networking

RQ1 deals with the effect of different traffic patterns and network topologies on the energy consumption of a mobile device. An empirical feasibility analysis of P2P networking in mobile and constrained environments was conducted. The goal was to understand the energy consumption characteristics of radio interfaces and typical composition of traffic in structured P2P networks, in order to identify the most essential obstacles for utilizing P2P networking in mobile environments.

Paper I evaluates the battery life of a smartphone acting as a peer node in a structured P2P overlay network. The objective was to identify the energy consumption characteristics of a networked mobile device. The evaluation involved two cases. In the first case a smartphone acted as a peer node in a Kademlia DHT-based P2P overlay network. The other peer nodes of the overlay were simulated on a server array while the energy consumption measurements were conducted on a real mobile device. The used overlay parameters were the number of peer nodes in the overlay, the resource lookup activity, and the level of churn. With 3G radio interface, the measured battery life varied around 3 hours and with the WLAN radio interface the battery life varied between 5 and 10 hours. The overlay parameters did not have a significant effect on the battery life of a device when 3G was used, whereas with WLAN the effect was more tangible. The second case focused on defining more detailed energy consumption profiles for mobile devices by using an artificial messaging scenario where the mobile node was sending and receiving UDP packets with predefined transmission intervals and packet sizes. The measurements were



smartphone sending and receiving UDP packets, but the focus is on a more detailed analysis on the combined effect of different sending intervals and packet sizes on the energy consumption of the mobile device. The evaluations using the 3G radio interface revealed a threshold in the packet size after which the energy consumption elevated to a higher level. The packet size threshold was observed to have a dependency with the transmission interval: the shorter the interval, the smaller the packet size threshold. In retrospective analysis, the threshold effect was an indication of the radio interface entering from the idle or PCH state directly to the DCH state to send or receive a packet instead of the FACH state that consumes less energy. With WLAN, not as clear a threshold effect was found. In addition, Paper II evaluates asymmetric messaging scenarios where the packet sizes and intervals for sending and receiving messages are different. The asymmetry between the packet sizes or transmission intervals of incoming and outgoing messages did not have noticeable effect on the energy consumption in most of the cases, except when the sending interval exceeded 100 ms while the receiving interval remained as 50 ms in WLAN environment. In conclusion, Papers I and II together provide an empirical basis for defining energy profiles for 3G and WLAN smartphone radio interfaces in messaging scenarios.

Paper III focuses on evaluating two popular DHT algorithms with respect to their routing performance that is an important factor for energy consumption in mobile P2P networking. The paper compares the lookup path length in terms of hop count and signalling overhead in terms of the amount of maintenance signalling. According to the results, Chord provides better scalability in terms of processed messages, although Kademlia's average message size is slightly smaller. When comparing the performance, Kademlia provides significantly shorter lookup path length, leading to lower lookup latency. As a conclusion, Kademlia seems more suitable for networks requiring high lookup performance. Chord, for one, is a more suitable option for large networks due to its better scalability in the terms of messaging overhead. Paper III provides valuable information on the composition of signalling traffic and common traffic patterns of different DHT-based structured P2P overlay networks. The paper also provides an empirical basis for defining traffic profiles for P2P overlay networks based on Kademlia and Chord.

In conclusion, Papers I–III jointly address RQ1 by:

- Defining the reference level for energy-efficiency optimizations by showing the average battery life of a smartphone acting as a peer node in a P2P overlay without any optimizations,

- Helping to understand the most essential properties of mainstream P2P networking and wireless communication technologies that affect the energy-efficiency of a mobile device, and
- Providing the empirical basis for defining the needed traffic and energy profiles for simulation-based energy consumption estimation.

3.2 e-Aware - energy consumption model for mobile networking

RQ2 deals with the load monitoring methods that can be used to exploit energy-awareness in load distribution among peer nodes and the amount of energy that can be saved using these methods in mobile peer nodes. Paper IV addresses RQ2 by proposing the e-Aware model for estimating how application layer properties affect the energy consumption of mobile devices already during the system design phase.

e-Aware estimates the energy consumption originating from network operations based on two energy consumption elements, signalling and media transfers. The distinction between signalling and media transfers is made due to their different energy consumption characteristics that were studied in Papers I–III. The model takes as parameters the average transmission interval and packet size for transferred data chunks smaller than 1500 bytes and the amount of moved data for data chunks greater than 1500 bytes. Power and energy consumption estimates are produced as output for a given scenario. The model calculates the total energy consumption as a sum of energy consumption originating from signalling (data chunks under 1500 bytes) and from media transfers (data chunks over 1500 bytes) (Eq. 1).

$$P(t) = \max[P_{sig}(t), P_{med}(t)] \quad (1)$$

For signalling, the model simulates the power consumption of a mobile device with different network interfaces using a set of equations based on the state machines of the wireless protocols and the empirical power consumption measurement data gathered in Papers I–II. In 3G networks, the network operator defines the conditions for transitions between different RRC states. Thus, in addition to the traffic parameters (packet size and transmission interval), the algorithm needs the following network parameters to work properly: the packet size threshold after which DCH is used instead of FACH, the timeout intervals for transitions from DCH to FACH and from FACH to idle, and the estimations for practical upload/download data transfer rates. As observed in Paper II, the packet size threshold seems to be a dynamical value, depending on the transmission interval, whereas the timeout intervals seem to be fixed. Fig. 14A shows an estimated power consumption curve in

an example case where the packet size is below the packet size threshold and Fig. 14B shows the curve in a case where the packet size exceeds the threshold value. τ_1 is the packet transmission time including the setup time for opening the data channel. τ_2 is the timeout interval for transition from DCH to FACH state and τ_3 is the timeout interval for transition from FACH to idle state.

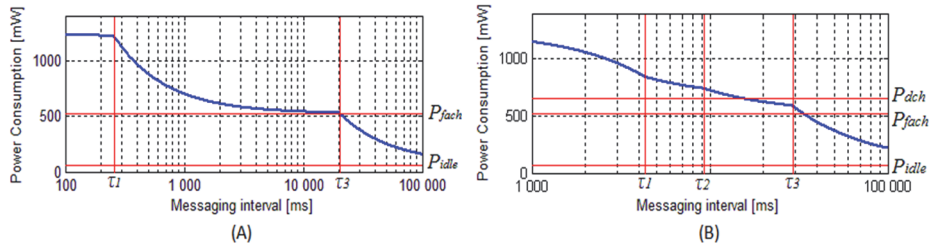


Fig. 14. Estimated power consumption curves for signalling in a 3G network (IV, published by permission of IEEE).

In WLAN networks, the conditions for transitions between different modes are mostly defined by the radio interface of the mobile device. In addition to the traffic parameters (packet size and transmission interval), the algorithm needs the following network parameters to work properly: the threshold after which the radio interface mode is changed from CAM mode to PSM mode and the estimations for practical upload/download data transfer rates. Fig. 15 shows an estimated power consumption curve in an example case. τ_1 is the packet transmission time including the setup time for opening the data channel. τ_2 is the timeout interval for transition from CAM to PSM mode.

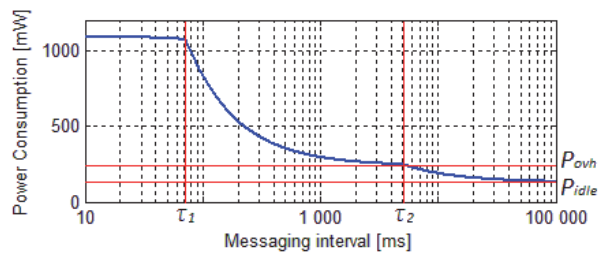


Fig. 15. Estimated power consumption curves for signalling in a WLAN network (IV, published by permission of IEEE).

For full-data-rate transfers, Paper IV presents a model that first calculates the data transfer time (t_{tra}), including the setup time, by using the size of the moved data object and the practical upload/download data transfer rates given as parameters. During t_{tra} , the power consumption is averagely P_{ul} for uploads or P_{dl} for downloads, after which the power consumption returns to P_{idle} through intermediate power states. Figs. 16 and 17 illustrate the estimated and measured power consumption in an example case for both 3G and WLAN.

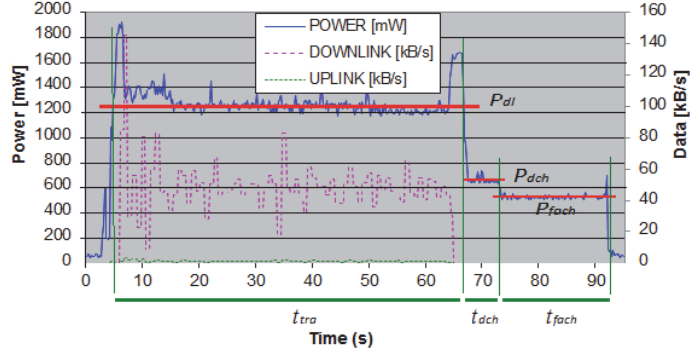


Fig. 16. Estimated and measured power consumption for a file download in a 3G-network (IV, published by permission of IEEE).

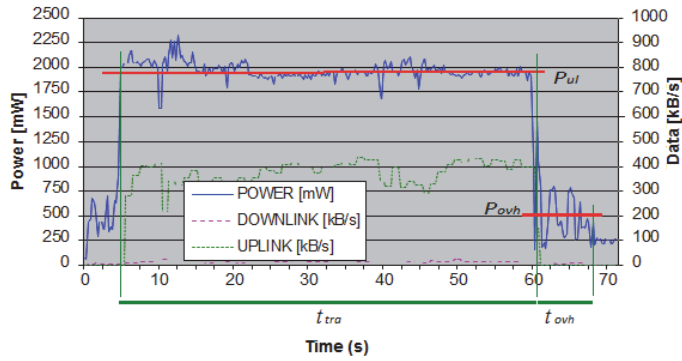


Fig. 17. Estimated and measured power consumption for file upload in a WLAN network (IV, published by permission of IEEE).

The model was implemented in Matlab, and evaluated by comparing the simulation results with the results of real-life measurements. The accuracy of the model was

evaluated for both signalling and media transfer parts. In addition, the applicability to a P2P scenario was evaluated with a real-world networking scenario to find out how the model copes with a scenario with highly varying packet sizes. According to the results, the model achieves high accuracy (3–6% estimation error) in signalling scenarios with fixed packet sizes and transmission intervals, as well as in file transfer scenarios (<1% error). In a signalling scenario with high variance in both packet sizes and transmission intervals, the estimation error was 14–21%.

3.3 e-Mon - energy-aware load monitoring model to enable load-balanced P2P networking

Paper V introduces the e-Mon energy-aware load monitoring model, which facilitates the participation of battery-powered devices in P2P networks by enabling energy-aware load balancing. e-Mon includes the energy status of a peer node as one of the load factors used in load balancing. e-Mon addresses RQ2 and RQ3 by making it possible to move overlay load from energy-critical battery-powered nodes to fixed nodes or other battery-powered nodes with higher remaining battery capacity. This way, the model helps saving the energy of mobile nodes, particularly in cases when the remaining battery capacity is low. Paper V provides a thorough energy-efficiency evaluation, where e-Mon is applied to three existing load balancing technologies. The results demonstrate that e-Mon can significantly improve the quality and fairness of load balancing between heterogeneous nodes with a proper selection of a load balancing model for the application scenario.

$$L_n = \sum_{i=1}^N \omega^i L_n^i, \quad (2)$$

$$L_n^{bat} = \mu + \frac{100 - \beta_n}{100 / (100 - \mu)}, \quad 0 \leq \mu < 100 \quad (3)$$

e-Mon calculates the load of a node as a weighed sum ω^i of different load factors L_n^i , as shown in Eq. 2. As a default, e-Mon uses communication load, computational load and battery load as load factors, but the model allows using other load factors as well. Computational and communication load values are obtained by using exponentially averaged CPU and communication channel utilization. Battery load is calculated as shown in Eq. 3. The model uses 0 as the battery load for all AC-plugged devices and the remaining battery percentage β_n for calculating the battery load L_n^{bat} of a devices running on their batteries. L_n^{bat} grows linearly from μ to 100, based on the remaining battery capacity. μ -parameter can be set between 0 and 100. It is used to adjust the balance of to what extent being battery powered

and remaining battery capacity affect the battery load. The effect of μ parameter is illustrated in Fig. 18. The weighing between load factors can be made either system-wide or node-specified, depending on the application scenario. As mentioned above, the load factors are not limited to the three mentioned. Instead, other load factors, such as memory allocation, storage utilization, and I/O operations, can be used as well. The model neither limits the number of used load factors.

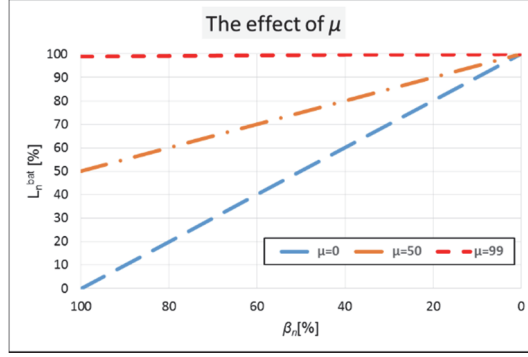


Fig. 18. Battery load as a function of batter percentage with different μ parameter values.

For evaluating the accuracy of e-Mon, an advanced evaluation setup was developed in Paper V. The evaluation setup uses real-world traffic and energy profiles to accurately simulate the energy consumption of a node in distributed networking scenarios. The simulation model consists of three parts: traffic generator, node simulator and energy consumption emulator. The traffic generator generates artificial overlay traffic for the node under evaluation. The node simulator simulates a peer node participating in the overlay network. It includes components for simulating both load monitoring and load balancing. e-Mon is used as the load monitoring component and existing load balancing models are used as load balancing components. The e-Aware energy consumption estimation tool, presented in Paper IV, is used as the energy consumption emulator.

e-Mon was evaluated by comparing its performance with three different load balancing models, including Virtual server (VS), Power of two choices (PO2C) and Advanced Finger Selection (AFSA), in three different setups: 1) a setup with load balancing equipped with e-Mon, 2) a setup without load balancing and 3) a setup with load balancing but without battery monitoring. The comparison reveals the total gain in battery life when e-Mon is in use and the contribution of battery load monitoring. The evaluation was conducted with two application scenarios having

different usage profiles: 1) video conferencing and 2) mobile cloud storage. Fig. 19 provides an example of the performance of e-Mon in the video conferencing scenario and Fig. 20 provides an example of the performance of e-Mon in the mobile cloud storage scenario.

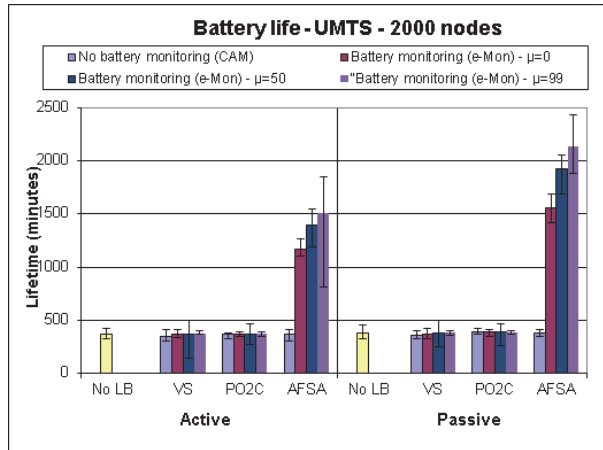


Fig. 19. Battery life in video conferencing scenario (N=2000, 3G).

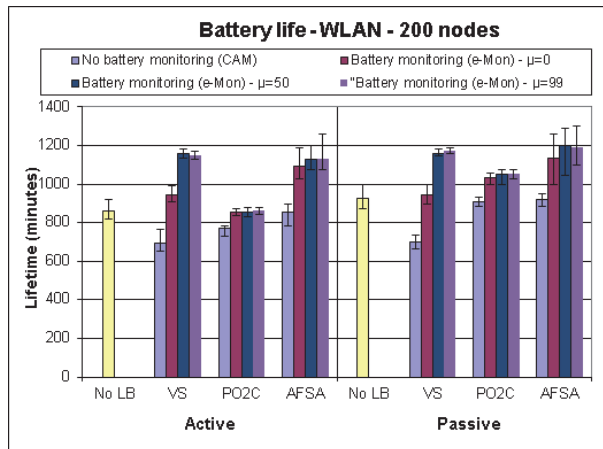


Fig. 20. Battery life in mobile cloud storage scenario (N=200, WLAN).

According to the results of the study, e-Mon achieved up to 470% battery life extension compared to the case with battery monitoring deactivated. The impact of e-

Mon was most visible with AFSA load balancing, where the battery life was extended by 17–470% depending on the application scenario and used parameters. With VS load balancing, the use of e-Mon improved the battery life by 1–68%. The effect of e-Mon on the overall performance of VS load balancing was weaker than with AFSA due to the massive overlay maintenance overhead inflicted by VS. This phenomenon is particularly visible in the performance of VS shown in Fig. 20. With PO2C load balancing, the use of e-Mon improved the battery life by 9–30% in the mobile cloud storage scenario. In the video conferencing scenario, the use of PO2C could not bring any energy savings. This is explained by the fact that PO2C cannot influence the load originating from the overlay maintenance signaling that is the dominant traffic type in the video conferencing scenario.

The results also revealed that using higher μ parameter values helps a battery-powered device to adapt faster to the changed energy status when it is plugged off the external power source. Despite the increased load-balancing overhead, the battery life typically improved by 10–20% in the evaluated scenarios when μ parameter was increased from minimum to maximum value. In conclusion, the results showed that with a proper selection of a load balancing model for different types of applications, e-Mon can significantly improve the battery life of mobile peer nodes.

3.4 ADHT –mobile agent-based virtual peer concept for constrained-capacity P2P nodes

Paper VI proposes the Agent-based DHT (ADHT) conceptual architecture to facilitate the participation of constrained-capacity wireless devices, such as M2M nodes, in P2P networks. This is achieved by using mobile agent-based virtual peer nodes to share the peer responsibilities among the nodes within a subnet. ADHT reduces peer node resource usage and energy consumption, thus addressing RQ3. According to the comparative analysis in Paper VI, ADHT helps reducing the overall energy consumption of a constrained-capacity node as compared to a flat P2P network and a previously proposed mDHT (Lee *et al.* 2009b). When compared to traditional super-peer architectures, ADHT removes the super-peer overload problem that is present in networks including only constrained-capacity nodes.

In the proposed ADHT concept, one node among the *sub-peer* nodes (i.e. nodes allocated for acting as a peer node when needed) in a subnet takes care of the peer responsibilities while the other nodes are either in sleep mode or act as ordinary

peer nodes. As illustrated in Fig. 21, the peer functionality is encapsulated in a mobile agent as a virtual peer node, called *peer agent*. The peer agent is rotated between the sub-peer nodes, based on a *peer allocation algorithm*. The mobile agent is composed of *code*, *resource* and *state* segments. The code segment contains the executable codes, including the DHT algorithm and the peer allocation algorithm. Peer allocation algorithm determines the most suitable sub-peer node as the next host for the agent (peer). The peer allocation algorithm can be a simple time share-based algorithm, where each sub-peer node hosts the peer at its turn for a pre-defined time period, or it can be a more sophisticated dynamic algorithm, that, e.g. takes into account the remaining battery capacity of the sub-peer nodes. While active in a sub-peer node, the peer agent updates its resource tables concerning that node, and keeps the data available in resource tables when migrating to other nodes. The ranking of the most suitable sub-peer nodes is represented in the host table, also located in the code segment. The resource segment contains the stored data, including the DHT routing table and the resource table of the resources stored in sub-peer nodes. The state segment contains the current status of computation.

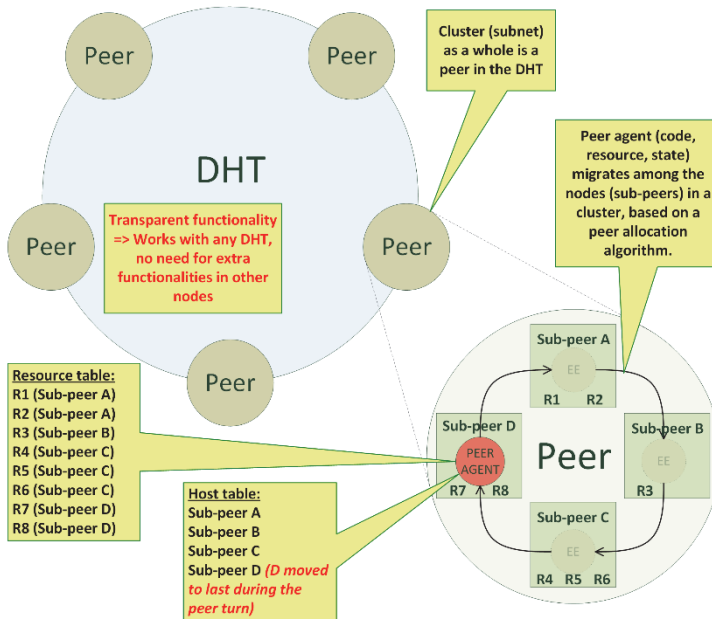


Fig. 21. Overview of the ADHT architecture (VI, published by permission of IEEE).

Since the data stored in a P2P system is located at the end nodes, the data management within the subnet has to be taken into account when the peer responsibility is shared. For this, ADHT provides three *data delivery modes* with different characteristics concerning the data freshness and hardware resource consumption. In *time-share* mode, the sub-peer data is refreshed in the peer agent's state segment the next time the peer migrates to the sub-peer node hosting the resource, allowing sub-peer nodes to sleep while not acting as a peer. When data request is received, the peer agent returns the data on behalf of the sub-peer node. In *publish-subscribe* mode, sub-peer nodes frequently wake up to update their data to the peer, in order to achieve higher freshness of the data. When a data request is received, the peer agent returns the data on behalf of the sub-peer node, as with time-share mode. In this mode, the data freshness is improved with the cost of higher energy consumption of sub-peer nodes. In *request-reply* mode, the peer forwards all incoming requests straight to the target sub-peer nodes in order to achieve high freshness of the data. As a result, sub-peer nodes are waken up more frequently than with other models. The energy-efficiency is lower but the data is always up to date.

The power consumption characteristics of the proposed architecture was analysed and compared with alternative architectures with a comparative analysis based on the literature and measurements conducted in Papers I–II. Table 1 summarizes the results that indicate the average total power consumption per subnet node in the ADHT architecture to be similar or lower than with the traditional super-peer architecture, depending on the used data delivery mode. The power consumption savings are significant compared to the flat P2P architecture. The most important difference is the even distribution of the load among the subnet nodes over time. The benefit is especially clear compared to the traditional super-peer architecture, where the (static) super-peer node consumes up to 450mW while the ordinary peer nodes consume only 30mW. It is obvious that if the super-peer node is one of the constrained-capacity nodes in a M2M cluster, its battery life would be unfeasibly short. Thus, ADHT has a great potential to solve the super-peer overload problem, as compared with the most feasible existing architecture.

Table 1. Comparison of the approximated power consumptions between ADHT and other P2P architectures.

Architectural model	Approx. average power / node
Flat	400 mW
Super-Peer	72 mW
mDHT	375 mW
ADHT	
Time-share	63.4 mW
Publish-Subscribe	71.5 mW
Request-reply	74 mW

4 Discussion

This chapter reflects on the thesis. Section 4.1 summarizes the results of this thesis and Section 4.2 deals with their limitations and generalizability. Section 4.3 discusses the significance of the results and explores avenues for future work.

4.1 Summary of results

Papers I and II provided an empirical study to identify the energy consumption characteristics of the most widely used mobile wireless communication technologies. The measurement results of Paper I revealed a non-linear correspondence between data rate and power consumption. The phenomenon was found to originate from the various states used by the radio interfaces in order to save energy. Paper II extended the study with a more detailed analysis on the combined effect of different sending intervals and packet sizes on the energy consumption of the mobile devices. The evaluations provide an insight into the energy consumption effects of different radio interface states for 3G and WLAN communication technologies.

Paper III continued the empirical study by providing a breakdown of the traffic for a P2P node running two popular DHT algorithms, Kademlia and Chord, comparing their performance and efficiency. As a result, Paper III provided valuable information on the composition of the signalling traffic and common traffic patterns of structured P2P networks. A centric observation was that the maintenance signalling dominates the traffic with both DHT algorithms. Furthermore, the average message transmission interval was found relatively short and the message sizes relatively small with both DHT algorithms. The study also provided valuable information on the effect of overlay parameters on the message transmission interval and lookup efficiency. Together, Papers I–III establish an empirical basis for defining energy consumption profiles for a mobile device connected to 3G and WLAN radio interfaces and traffic profiles for structured P2P overlay networks.

The models for energy consumption optimization proposed in Papers IV–V are based on the empirical basis established in Papers I–III. Paper IV introduced the e-Aware model for estimating how application layer properties affect the energy consumption of mobile devices operating in mobile networks. According to the results, the accuracy of the e-Aware model is high in application scenarios where the traffic consists of frequently sent packets with low variance in packet sizes (3-6% estimation error) and full-bandwidth data transfers ($< 1\%$ estimation error). When the packet sizes vary significantly, the accuracy was found lower, but still acceptable

(14–21% estimation error). The e-Aware model was found to have strong potential to facilitate the development of energy-efficient networking solutions by reducing the need for time-consuming iterations between system development and evaluations with real-life networks and devices. Paper V utilizes the energy consumption estimation model provided by Paper IV by proposing the e-Mon model for energy-aware load monitoring. The e-Mon model facilitates the participation of battery-powered devices in P2P and other distributed networks by enabling energy-aware load balancing. This is done by including the energy status of a peer node as one of the load factors used in defining the load of a peer node. The results demonstrate that the model can significantly improve the battery life of mobile nodes by improving the quality and fairness of load balance between heterogeneous nodes. e-Mon achieved up to 470% battery life extension compared to the case with battery monitoring deactivated.

Paper VI proposed the ADHT concept, including a novel concept to reduce the energy consumption of peer nodes in M2M environment by allowing them to sleep for most of their time. So far, only the super-peer architecture has allowed some of the P2P nodes to enter sleep mode by allocating at least one node in a cluster as a super-peer node while the others have been ordinary peer nodes that can sleep when no activity is required from them. The super-peer model is problematic if all nodes in the subnet are battery-powered constrained-capacity nodes: if the super-peer node was one of the constrained-capacity nodes in a M2M cluster, its battery life would most likely become unfeasibly short. ADHT prevents this problem by rotating the peer agent, which role is equivalent to super-peer node, between the sub-peer nodes. The analysis of the architecture indicates that the power consumption savings are significant when compared to the flat P2P architecture and that the model removes the problem related to extensive super-peer energy consumption when compared to the super-peer architecture.

Overall, the results successfully address the research questions by: 1) providing a detailed analysis on how different traffic patterns affect the energy consumption of mobile devices (RQ1, Papers I–III), 2) introducing methods that exploit energy-awareness in load distribution among peer nodes and shows how much energy can be saved by using these methods (RQ2, Papers IV, V), and 3) introducing an optimized network topology for reducing mobile peer node resource usage and achieving energy savings with it (RQ3, Paper VI).

4.2 Limitations and generalizability

The energy consumption measurements in Papers I–II and energy profiles in Paper IV can be considered accurate with the technologies used in the studies, but they can also be considered generic under specific circumstances. Considering the wireless communication technologies, the validity of the energy profiles is limited to devices using 3G and WLAN radio interfaces. To achieve the best accuracy, the parameters such as timeout intervals and packet size thresholds for energy profiles need to be separately defined for each network using, e.g. empirical tests. This is needed since it seems that different network operators use different settings in their networks. Since most mobile devices use generic radio chipsets, the energy profiles obtained for a specific device type such as a smartphone should provide quite accurate estimation for other devices of the same type.

The results of Papers III and V are specific to the used DHT algorithms, P2P protocols, load balancing algorithms, network interfaces and application scenarios. As concluded in Paper V, the results successfully demonstrate the potential of energy-aware load balancing, but performance depends heavily on operational circumstances. Thus, case-specific evaluations are needed to figure out the suitability of the approach for different application scenarios.

The results presented in Paper VI are based on a comparative analysis based on the literature and the measurements conducted in Papers I–II. Thus, the results can be considered to be approximate. The analysis in Paper VI provides approximations on the relative energy consumption between different models. To obtain more accurate results, measurements with a real-life prototype and a more detailed performance analysis are needed.

4.3 Significance of the results and future work

Since a growing share of Internet nodes is mobile, the requirement of energy-efficiency will be emphasized in future communication systems. Today's mobile devices and networks already incorporate advanced energy-saving functions in lower layers, concerning both processing architectures and wireless network technologies. It is important that higher-layer solutions can maximally exploit these functions.

P2P networking provides a decentralized, self-organizing and scalable alternative for traditional client/server-based networking (Androutsellis-Theotokis & Spinellis 2004). P2P systems have also high potential for optimizing content delivery in otherwise client/server-based systems, such as in cloud computing, where P2P

can alleviate server load by taking into use the spare computational and storage resources of end-user devices (Korzun & Gurtov 2013, Babaoglu *et al.* 2012, Trajkovska *et al.* 2010, Zhelev & Georgiev 2011). However, the high energy consumption at peer nodes (Nurminen & Nöyränen 2008) and the limitations concerning the mobile environment (Kellerer *et al.* 2006, Mäkinen & Nurminen 2008) have so far made P2P networking unfeasible in mobile environment. Considering this context, the overall contributions of this thesis and their significance on the high level is assessed below.

The basic findings of the empirical study provided by Papers I–III have been confirmed and complemented by various other papers published concurrently or later in the area and the papers have been cited so far by more than sixty other scientific papers in the literature. The broad number of citations gives a strong indication of the scientific impact of the results.

e-Aware (Paper IV) accelerates the development of energy-efficient networking solutions by reducing the need for time-consuming iterations between system development and evaluations with real-life networks and devices. e-Aware was among the first energy consumption models published in the literature for estimating the energy consumption of a mobile device in different communication scenarios. Many related works have later cited Paper IV, either when explaining the energy consumption characteristics of 3G or WLAN radio interfaces or as related work when proposing their own energy consumption models. For instance, Vergara & Tehrani (2013) and Vergara *et al.* (2014) have later presented an energy consumption model that, instead of traffic parameters, uses real-world traffic traces in estimating the energy consumption of a mobile node in communication scenarios. It is, however, notable that the model was fundamentally based on the same principle of using an empirically defined energy profile to generate the energy consumption estimation for a mobile node, originally proposed by us.

e-Mon (Paper V) facilitates the participation of battery-powered devices in P2P and other distributed networks by enabling energy-aware load balancing so that energy-critical mobile nodes carry less load than non-energy-critical, mostly fixed, nodes. Fair load balancing is very important in mobile P2P systems since the excess load of a mobile device affects both the short-term and long-term user experience through lower performance and shorter battery life (Ickin *et al.* 2012, Doherty & Sorenson 2015, Kujala *et al.* 2011, Heikkinen & Nurminen 2010). So far, there are very few studies in the literature considering energy-awareness in P2P load balancing, although some studies have provided models for energy-aware load dispatch-

ing among servers in distributed server clusters. Paper V fills this very gap by introducing an energy-aware load monitoring system developed specifically for P2P systems, including mobile and constrained-capacity peer nodes.

ADHT (Paper VI) facilitates the participation of constrained-capacity wireless devices, such as M2M nodes, in P2P networks by allowing them to sleep for most of their time. Most of the existing M2M processing architectures and wireless protocols achieve ultra-low power connectivity by extensive use of sleep modes (Jurdak *et al.* 2010, Rault *et al.* 2014, Xing *et al.* 2007, Gomez & Paradells 2010), which is basically impossible with current P2P protocols due to their frequent signalling traffic. By addressing this problem, ADHT contributes to an important research topic, since it is predicted that M2M device connections worldwide will grow rapidly during the coming years (Analysys Mason 2013, CISCO Infographic 2013). Regarding to related work, there are some concurrently made proposals with some similarities. For instance, Bianzino *et al.* (2014) have proposed an energy-aware node clustering model with a proxy rotation scheme. The idea of a rotating proxy is somewhat similar to ADHT's rotating peer agent, but the context is outside P2P and thus the rotation logic and data management are more straightforward. However, the model included also other features, such as taking into account sensitive node activities that may be disturbed by the proxy rotation. This type of functionality would also be worth considering in the future versions of ADHT, which along with energy-aware peer node selection algorithm would provide improved load balance between sub-peer nodes.

This thesis has made several contributions towards enabling energy-aware P2P networking in mobile environments, which in turn enables the broader use of P2P networking as a general-purpose optimization method in different computing systems, such as cloud computing. The contributions also provide a scientific foundation for further work in the area. The future work includes creating energy profiles for a broader range of wireless network interfaces and device types. The most relevant wireless communication technologies include 4G and various short-range wireless communication technologies. In addition to the P2P protocols, it would also be beneficial to study the energy profiles of other distributed computing protocols and client/server protocols. According to the e-Aware model, the current parametric approach in estimating the energy consumption of an application scenario should be complemented with a traffic trace-based approach for improved accuracy when more detailed traffic traces are available. Finally, implementing a real-world prototype of the ADHT model, together with more advanced peer node selection

algorithm, would improve the fairness and energy-efficiency of the model and provide more detailed results on its energy-efficiency compared to traditional solutions.

5 Conclusions

Energy efficiency is a powerful measure for promoting sustainability in the technological evolution and ensuring feasible battery life of end-user devices in mobile computing. Peer-to-peer technology provides a decentralized, self-organizing and scalable architecture to distribute content between devices in networks that scale up almost infinitely. Peer-to-peer technology also has high potential for optimizing content delivery in systems fundamentally based on client/server model, such as cloud-based content delivery networks, where peer-to-peer technology can alleviate server load by taking into use the spare resources of end-user devices. However, peer-to-peer networking may require lots of resources from peer nodes, which in turn may lead to increased energy consumption on mobile devices. For this reason, peer-to-peer networking has so far been considered unfeasible for mobile environment. This thesis explored mechanisms for enabling energy-aware peer-to-peer networking in order to facilitate the use of peer-to-peer systems in mobile environments.

The thesis commenced with an empirical study to understand the energy consumption characteristics of radio interfaces and typical composition of traffic in structured peer-to-peer networks. This was done in order to identify the most essential obstacles for utilizing peer-to-peer technology in mobile environments. The study established an empirical basis for defining energy consumption profiles for a mobile device connected to 3G and WLAN networks and traffic profiles for structured peer-to-peer overlay networks. Next, the thesis proposed a set of models for distributing load in an energy-aware manner among mobile devices so that energy-critical nodes would carry less load than non-energy-critical nodes. The e-Aware model was proposed for estimating the energy consumption of a mobile device in different distributed application scenarios. It was empirically verified to achieve 3-21% error in comparison to real-life measurements. The model accelerates the development of energy-efficient networking solutions by reducing the need for time-consuming iterations between system development and evaluations with real-life networks and devices. The e-Mon model was proposed for the energy-aware load monitoring of peer nodes. The model facilitates the participation of battery-powered devices in peer-to-peer and other distributed networks by enabling energy-aware load balancing where energy-critical mobile nodes carry less load than non-energy-critical nodes. It was empirically demonstrated to improve the battery life of mobile peer nodes up to 470%. Finally, the ADHT concept of mobile agent based

virtual peer load balancing was proposed for sharing the peer responsibilities between constrained-capacity mobile peer nodes in a subnet so that they can participate in a peer-to-peer overlay without compromising their battery life. The results indicate that the ADHT has a great potential to remove the super-peer overload problem while decreasing the average power consumption of nodes in a subnet.

The thesis gives valuable insight into implementing energy-efficient peer-to-peer systems in mobile environments. The results of the thesis help enabling energy-aware peer-to-peer networking in mobile environments, which in turn enables the broader use of peer-to-peer networking as a general-purpose optimization method in different computing systems, such as cloud computing. The thesis contributes to a topic of a growing importance, since a growing share of Internet nodes is mobile and it is foreseen that machine-to-machine device connections worldwide will grow rapidly during the coming years.

References

- Akyildiz IF, Melodia T & Chowdhury KR (2007) A survey on wireless multimedia sensor networks. *International Journal of Computer and Telecommunications Networking* 51(4): 921–960.
- Alzoubi K, Li, XY, Wang Y, Wan PJ & Frieder O (2003) Geometric spanners for wireless ad hoc networks. *IEEE Transactions on Parallel Distributed Systems* 14(4): 408–421.
- Analysys Mason (2013) Forecast report – M2M device connections and revenue: Worldwide forecast 2013–2023. URI: http://www.analysysmason.com/Research/Content/Reports/M2M-connections-forecast-Aug2013_RDME0/.
- Anastasi G, Conti M, Di Francesco M & Passarella A (2012) Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks* 7(3): 537–568.
- Anceaume E, Gradinariu M & Ravoaja A (2005) Incentives for P2P fair resource sharing. *Proc. 5th IEEE International Conference on Peer-to-Peer Computing*. Konstanz, Germany: 253–260.
- Androutsellis-Theotokis S & Spinellis D (2004) A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys* 36(4): 335–371.
- Artigas M, Lopez P & Karmeta A (2005) Cyclone: a novel design schema for hierarchical DHTs. *Proc. 5th IEEE International Conference on Peer-to-Peer Computing*. Konstanz, Germany: 49–56.
- Ayoub ZT, Ouni S & Kamoun F (2011) Association criteria to optimize energy consumption and latency for IEEE 802.15.4/ZigBee wireless sensor networks. *Proc. 4th Joint IFIP Conference on Wireless and Mobile Networking*. Toulouse, France: 1–4.
- Aziz SM & Pham DM (2013) Energy efficient image transmission in wireless multimedia sensor networks. *IEEE Communications Letters* 17(6): 1084–1087.
- Ba H, Demirkol I & Heinzelman W (2013) Passive wake-up radios: from devices to applications. *Ad Hoc Networks* 11: 2605–2621.
- Babaoglu O, Marzolla M & Tamburini M (2012) Design and implementation of a P2P cloud system. *Proc. 27th ACM Annual Symposium on Applied Computing*. Riva (Trento), Italy: 412–417.
- Balasubramanian B, Balasubramanian A & Venkataramani A (2009) Energy consumption in mobile phones: A measurement study and implications for network applications. *Proc. 9th ACM SIGCOMM Internet Measurement Conference*. Chicago, IL, USA: 280–293.
- Basole RC & Karla J (2011) On the evolution of mobile platform ecosystem structure and strategy. *Business & Information Systems Engineering* 3(5): 313–322.
- Bharambe A, Agrawal M & Seshan S (2004) Mercury: Supporting scalable multi-attribute range queries. *Proc. ACM International Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York, NY, USA: 353–366.
- Bharambe AR, Herley C & Padmanabhan VN (2006) Analyzing and improving a BitTorrent networks performance mechanisms. *Proc. 25th IEEE International Conference on Computer Communications*. Barcelona, Spain: 1–12.

- Bianchi S, Serbu S, Felber P & Kropf P (2006) Adaptive load balancing for DHT lookups. Proc. 15th International Conference on Computer Communications and Networks. Arlington, VA, USA: 411–418.
- Bianzino AP, Asplund M, Vergara EJ & Nadjm Tehrani S (2014) Cooperative proxies: Optimally trading energy and quality of service in mobile devices. *Computer Networks* 75(A): 297–312.
- Bienkowski M, Korzeniowski M & Heide FM (2005) Dynamic load balancing in distributed hash tables. *Lecture Notes in Computer Science* 3640: 217–225.
- Bluetooth (2014) Bluetooth core specification 4.2. URI: https://www.bluetooth.org/DocMan/handlers/DownloadDoc.ashx?doc_id=286439.
- Bontu CS & Illidge E (2009) DRX mechanism for power saving in LTE – Topics in radio communications. *IEEE Communications Magazine* 47: 48–55.
- Bormann C, Castellani A.P & Shelby Z (2012) CoAP: An application protocol for billions of tiny internet nodes. *IEEE Internet Computing* 16(2): 62–67.
- Boukhelef D & Kitagawa H (2009) Dynamic load balancing in RCAN content addressable network. Proc. 3rd International Conference on Ubiquitous Information Management and Communication. Suwon, Korea: 98–106.
- Brands EHTB & Kragiannis G (2009) Taxonomy of P2P applications. Proc. IEEE Workshop on Enabling Service-Oriented Internet. Hawaii, HI, USA: 1–8.
- Breitgand D, Cohen R, Nahir A & Raz D (2010) On cost-aware monitoring for self-adaptive load sharing. *IEEE Journal on Selected Areas in Communications* 28(1): 70–83.
- Bryhni H, Klovning E & Kure O (2000) A comparison of load balancing techniques for scalable Web servers. *IEEE Network Magazine* 14(4): 58–64.
- Byers J, Considine J & Mitzenmacher M (2002) Simple load balancing for distributed hash tables. *Lecture Notes in Computer Science* 2735: 80–87.
- Cardellini V, Colajanni M & Yu PS (1999) Dynamic load balancing on web-server systems. *IEEE Internet Computing* 3(3): 28–39.
- Carroll A & Heiser G (2010) An analysis of power consumption in a smartphone. Proc. USENIX Annual Technical Conference. Boston, MA, USA: 21.
- Celtic-Plus (2012) Purple Book – Celtic-Plus programme of possible and recommended research items. URI: <https://www.celticplus.eu/celtic-plus-purple-book/>.
- Ciraci S, Körpeogly I & Ulusoy Ö (2009) Reducing query overhead through route learning in unstructured peer-to-peer network. *Journal of Network and Computer Applications* 32(3): 550–567.
- CISCO (2014) Cisco visual networking index: global mobile data traffic forecast update 2014–2019. URI: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html.
- CISCO Infographic (2013) The internet of things. URI: <http://share.cisco.com/internet-of-things/>.
- Cohen E & Schenker S (2002) Replication strategies in unstructured peer-to-peer networks. Proc. ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. Pittsburgh, PA, USA: 177–190.

- Cooper MD (1983) Response time variations in an online search system. *Journal of The American Society for Information Science* 34(6): 374–380.
- De Mil P, Allemeersch T, Moerman I, Demeester P & De Kimpe W (2008) A scalable low-power WSN solution for large-scale building automation. *Proc. IEEE International Conference on Communications*. Beijing, China: 3130–3135.
- Dementyev A, Hodges S, Taylor S & Smith J (2013) Power consumption analysis of Bluetooth low energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario. *Proc. IEEE International Wireless Symposium*. Beijing, China, 1–4.
- Doherty RA & Sorenson P (2015) Keeping users in the flow: Mapping system responsiveness with user experience. *Procedia Manufacturing* 3: 4384–4391.
- Ehsan S & Hamdaoui B (2012) A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks. *IEEE Communications Surveys & Tutorials* 14(2): 265–278.
- Eronen P (2008) TCP Wake-up: reducing keep-alive traffic in mobile IPv4 and IPsec NAT traversal. Nokia Research Center, Helsinki, Finland, Rep. NRC-TR-2008-002.
- Faik J (2005) A model for resource-aware load balancing on heterogeneous and non-dedicated clusters. Ph.D. dissertation, Dept. Computer Science, Rensselaer Polytechnic Institute, Troy, USA: 90.
- Felber P, Kropf P, Schiller E & Serbu S (2014) Survey on load balancing in peer-to-peer distributed hash tables. *IEEE Communications Surveys & Tutorials* 16(1): 473–492.
- Ferrari D & Zhou S (1987) An empirical investigation of load indices for load balancing applications. *Proc. 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*. Amsterdam, Netherlands: 515–528.
- Gao B, He C & Jiang L (2008) Modeling and analysis of IEEE 802.15.4 CSMA/CA with sleep mode enabled. *Proc. 11th IEEE Singapore International Conference on Communication Systems*. Guangzhou, China: 6–11.
- Godfrey B & Stoica I (2005) Heterogeneity and load balance in distributed hash tables. *Proc IEEE INFOCOM*. Miami, USA 1: 596–606.
- Gomez C & Paradells J (2010) Wireless home automation networks: a survey of architectures and technologies. *IEEE Communications Magazine* 48(6): 92–101.
- Gomez C, Oller J & Paradells J (2012) Overview and evaluation of Bluetooth low energy: An emerging low-power wireless technology. *Sensors (Basel)* 12(9): 11734–11753.
- Gubbi J, Buyya R, Marusic S & Palaniswami M (2013) Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems* 29(7): 1645–1660.
- Halepovic E & Deters R (2003) The costs using JXTA. *Proc. 3rd International Conference on Peer-to-Peer Computing*. Linköping, Sweden: 160–167.
- Harjula E, Ala-Kurikka J, Howie D & Ylianttila M (2006) Analysis of peer-to-peer SIP in a distributed mobile middleware system. *Proc. 49th Global Telecommunications Conference*. San Francisco, CA, USA: 1–6.

- Harjula E, Hautakorpi J, Beijar N & Ylianttila M (2009) Peer-to-peer SIP for mobile computing: challenges and solutions. In: Seet BC (ed) *Mobile Peer-to-Peer Computing for Next Generation Distributed Environments: Advancing Conceptual and Algorithmic Applications*. Hershey, PA, USA, IGI Publish: 326–347.
- Harjula E, Kassinen O & Ylianttila M (2012) Energy consumption model for mobile devices in 3G and WLAN networks. *Proc. 9th IEEE International Conference on Consumer Communications and Networking*. Las Vegas, NV, USA: 532–527.
- Harjula E, Ylianttila M, Ala-Kurikka J, Riekk J & Sauvola J (2004) Plug-and-play application platform: Towards mobile peer-to-peer. *Proc. 3rd International Conference on Mobile and Ubiquitous Multimedia*. College Park, MD, USA: 63–69.
- Hautakorpi J & Mäenpää J (2010) Load balancing for structured P2P networks using the advanced finger selection algorithm (AFSA). *Proc. ACM Symposium on Applied Computing*. Sierre, Switzerland: 655–662.
- Haverinen H, Siren J & Eronen P (2007) Energy consumption of always-on applications in WCDMA networks. *Proc. 65th IEEE Vehicular Technology Conference*. Dublin, Ireland: 964–968.
- Hefeeda M & Saleh O (2008) Traffic modeling and proportional partial caching for peer-to-peer systems. *IEEE/ACM Transactions on Networking* 16(6): 1447–1460.
- Heikkinen MVJ & Nurminen JK (2010) Consumer attitudes towards energy consumption of mobile phones and services. *Proc. 72nd Vehicular Technology Conference*. Ottawa, Canada, 1–5.
- Hoque MA, Siekkinen M & Nurminen JK (2014) Energy efficient multimedia streaming to mobile devices – A survey. *IEEE Communications Surveys & Tutorials* 16(1): 579–597.
- Hoque MA, Siekkinen M, Nurminen JK, Aalto M & Tarkoma S (2015) Mobile multimedia streaming techniques: QoE and energy consumption perspective. *Pervasive and Mobile Computing* 16(A): 96–114.
- Hoßfeld T, Tutschku K, Andersen FU, de Meer H & Oberender JO (2005) Simulative performance evaluation of a mobile peer-to-peer file-sharing system. *Next Generation Internet Networks*. Rome, Italy: 281–287.
- Huang J, Quian F, Guo Y, Zhou Y, Xu Q, Mao ZM, Sen S & Spatscheck O (2013) An in-depth study of LTE: Effect of network protocol and application behavior on performance. *Proc. 13th ACM SIGCOMM Conference on Internet measurement*. Hong Kong: 363–374.
- Ickin S, Wac K, Fiedler M, Janowski L, Hong J-H & Dey AK (2012) Factors influencing quality of experience of commonly used mobile applications. *IEEE Communications Magazine* 50(4): 48–56.
- IEEE TG 15.4 (2006) Part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs). *IEEE Standard for Information Technology*.
- Jain N, Mahajan P, Kit D, Yalagandula P, Dahlin M & Zhang Y (2008) Network imprecision: A new consistency metric for scalable monitoring. *Proc. 8th USENIX conference on Operating systems design and implementation*. San Diego, CA, USA: 87–102.

- Jennings C, Lowekamp B, Rescorla E, Baset S & Schulzrinne H (2013) Resource location and discovery (RELOAD) base protocol – draft-ietf-p2psip-base-26. IETF draft (expired).
- Jeong H, Kim J & Yoo Y (2011) Adaptive broadcasting method using neighbor type information in wireless sensor networks. *Sensors* 11(6): 5952–5967.
- Jurdak R, Ruzzelli AG & O'Hare GMP (2010) Radio sleep mode optimization in wireless sensor networks. *IEEE Transactions on Mobile Computing* 9(7): 955–968.
- Karger D & Ruhl M (2004) Simple efficient load balancing algorithms for peer-to-peer systems. *Proc. 16th Annual ACM Symposium on Parallelism in Algorithms and Architectures*. Barcelona, Spain: 36–43.
- Karger D, Lehman E, Leighton F, Levine M, Lewin D & Panigrahy R (1997) Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web. *Proc. 29th Annual ACM Symposium on Theory of Computing*. El Paso, TX, USA: 654–663.
- Kelényi I & Nurminen JK (2008) Energy aspects of peer cooperation – Measurements with a mobile DHT system. *Proc. Cognitive and Cooperative Wireless Networks Workshop in the IEEE International Conference on Communications 2008*. Beijing, China: 164–168.
- Kelenyi I & Nurminen JK (2008) Optimizing energy consumption of mobile nodes in heterogeneous Kademia-based distributed hash tables. *Proc. Next Generation Mobile Applications, Services and Technologies*. Cardiff, UK: 70–75.
- Kellerer W, Kunzmann G, Schollmeier R & Zölsb (2006) Structured peer-to-peer systems for telecommunications and mobile environments. *International Journal of Electronics and Communications* 60(1): 25–29.
- Khssibi S, Idoudi H, Van Den Bossche A & Saidane L (2013) Presentation and analysis of a new technology for low-power wireless sensor network. *International Journal of Digital Information and Wireless Communications* 3(1): 75–86.
- Kim K & Park D (2006) Heterogeneity aware P2P algorithm by using mobile nodeID. *Springer Lecture Notes in Computer Science* 3961: 975–984.
- Kimura N & Latifi S (2005) A survey on data compression in wireless sensor networks. *Proc. International Conference on Information Technology: Coding and Computing*. Las Vegas, NV, USA: 8–13.
- Korzun D & Gurtov A (2011) Survey on hierarchical routing schemes in “flat” distributed hash tables. *Peer-to-Peer Networking and Applications* 4(4): 346–375.
- Korzun D & Gurtov A (2013) Structured peer-to-peer Systems – Fundamentals of hierarchical organization, routing, scaling, and security. Springer, Heidelberg.
- Koskela T & Vätjus-Anttila J (2015) Optimization techniques for 3D graphics deployment on mobile devices. *3D Research – February 2015* 6(1): 1–27.
- Koskela T, Kassinen O, Harjula E & Ylianttila (2013) P2P group management systems: A conceptual analysis. *ACM Computing Surveys* 45(2): 20.
- Kosta S, Aucinas A, Hui P & Mortier R (2012) ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading. *Proc. 31st Annual IEEE International Conference on Computer Communications*. Orlando, FL, USA, 945–953.

- KPCB (2015) Internet Trends 2015 – Code Conference. URI: <http://www.kpcb.com/internet-trends>.
- Kramers A, Höjer M, Lövehagen N & Wangel J (2014) Smart sustainable cities – Exploring ICT solutions for reduced energy use in cities. *Environmental Modelling & Software* 56: 52–62.
- Krashinsky R & Balakrishnan H (2002) Minimizing energy for wireless web access with bounded slowdown. *Proc. 8th Annual International Conference on Mobile Computing and Networking*. Atlanta, GA, USA: 119–130.
- Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E & Sinnelä A (2011) UX Curve: A method for evaluating long-term user experience. *Interacting with Computers* 23: 473–483.
- Kumar K & Lu YH (2010) Cloud computing for mobile users: Can offloading computation save energy?, *Computer* 43(4): 51–56.
- Kumar K, Liu J, Lu Y-H & Bhargawa B (2013) A Survey of computation offloading for mobile systems. *Mobile Networks and Applications* 18(1): 129–140.
- Ledlie J & Seltzer M (2005) Distributed, secure load balancing with skew, heterogeneity and churn. *Proc. 24th Annual Joint Conference of the IEEE Computer and Communications Societies*. Miami, FL, USA 2: 1419–1430
- Lee JH, Song JW, Lee YH & Yang SB (2009) An energy-effective routing protocol for mobile P2P systems. *Proc. 4th International Conference on Wireless Pervasive Computing*. Melbourne, Australia: 1–5.
- Lee JW, Schulzrinne H, Kellerer W & Despotovic Z (2009) mDHT: Multicast-augmented DHT Architecture for High Availability and Immunity to Churn. *Proc. 6th IEEE Consumer Communications and Networking Conference*. Las Vegas, NV, USA: 1–5.
- Leppänen T, Liu M, Harjula E, Ramalingam A, Ylioja J, Närhi P, Riekkä J & Ojala T (2013) Mobile agents for integration of internet of things and wireless sensor networks. 2013 IEEE International Conference on Systems, Man, and Cybernetics. Manchester, UK: 14–21.
- Lu R, Li X, Liang X, Shen X & Lin X (2011) GRS: The green, reliability, and security of emerging machine to machine communications. *IEEE Communications Magazine* 49(4): 28–35.
- Lua EK, Crowcroft J, Pias M, Sharma R & Lim S (2005) A survey and comparison of peer-to-peer overlay network schemes. *IEEE Communications Surveys* 7(2): 72–93.
- Ma T, Hempel M, Peng D & Sharif H (2013) A Survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems. *IEEE Communications Surveys & Tutorials* 15(3): 963–972.
- Mahmud K, Inoue M, Murakami H, Hasegawa M & Morikawa H (2004) Measurement and usage of power consumption parameters of wireless interfaces in energy-aware multi-service mobile terminals. *Proc. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. Barcelona, Spain 2: 1090–1094.
- Mäkinen L & Nurminen JK (2008) Measurements on the feasibility of TCP NAT traversal in cellular networks. *Proc. 4th EURO-NGI Conference on Next Generation Internet Networks*. Krakow, Poland: 261–267.

- Manweiler J & Choudhury RR (2011) Avoiding the rush hours: WiFi energy management via traffic isolation. Proc. 9th International Conference on Mobile Systems, Applications and Services. Washington DC, USA: 253–266.
- Maymounkov P & Mazieres D (2002) Kademlia: A peer-to-peer information system based on the XOR metric. Proc. International Workshop on Peer-to-Peer Systems. Cambridge, MA, USA: 1–6.
- Mitzenmacher M (2001) The power of two choices in randomized load balancing. IEEE Transactions on Parallel and Distributed Systems 12(10): 1094–1104.
- NIST (2012) Secure hash standard. Federal Information Processing Standards Publication 180–4.
- Nurminen JK & Nöyränen (2009) Parallel data transfer with voice calls for energy-efficient mobile services. Mobile Wireless Middleware, Operating Systems, and Applications 7: 87–100.
- Nurminen JK & Nöyränen JK (2008) Energy-consumption in mobile peer-to-peer – quantitative results from file sharing. Proc. 5th IEEE Consumer Communications and Networking Conference. Las Vegas, NV, USA: 729–733.
- Nurminen JK (2010) Parallel connections and their effect to battery consumption of a mobile phone. Proc. 7th IEEE Consumer Communications and Networking Conference. Las Vegas, NV, USA: 1–5.
- Olsson J (2014) 6LoWPAN demystified. URI: <http://www.ti.com/lit/wp/swry013/swry013.pdf>.
- Ou Z (2010) Structured peer-to-peer networks: Hierarchical architecture and performance evaluation. Ph.D. dissertation, Dept. Electrical and Information Engineering, University of Oulu, Finland.
- Ou Z, Harjula E, Kassinen O & Ylianttila M (2010) Performance evaluation of a Kademlia-based communication-oriented P2P system under churn. Computer Networks 54(5): 689–705.
- Ou Z, Harjula E, Koskela T & Ylianttila M (2010) GTPP: General truncated pyramid peer-to-peer architecture over structured DHT networks. Mobile Networks and Applications 15(5): 729–749.
- Pathak A, Hu YC & Zhang M (2012) Where is the energy spent inside my app?: Fine grained energy accounting on smartphones with Eprof. Proc. 7th European Conference on Computer Systems. Bern, Switzerland: 29–42.
- Perälä PHJ, Barbuzzi A, Boggia G & Pentikousis K (2009) Theory and practice of RRC state transitions in UMTS networks. Proc. IEEE Broadband Wireless Access Workshop. Hawaii, USA: 1–6.
- Pering T, Agarwal Y, Gupta R & Want R (2006) CoolSpots: Reducing the power consumption of wireless mobile devices with multiple radio interfaces. Proc. 4th International Conference on Mobile Systems. Uppsala, Sweden: 220–232.
- Pouwelse J, Garbacki P, Epema D & Sips H (2005) The Bittorrent P2P file-sharing system: Measurements and analysis. Lecture Notes in Computer Science 3640: 205–216.

- Putrycz E (2003) Design and implementation of a portable and adaptable load balancing framework. Proc. 13th Conference of the Centre for Advanced Studies on Collaborative Research. Toronto, Ontario, Canada, 2003: 238–252.
- Pyles AJ, Qi X, Zhou G, Keally M & Liu X (2012) SAPSM: Smart adaptive 802.11 PSM for smartphones. Proc. 2012 ACM Conference on Ubiquitous Computing. Pittsburgh, PA, USA: 11–20.
- Qian F, Wang Z, Gao Y, Huang J, Gerber A, Mao Z.M, Sen S & Spatscheck O (2012) Periodic transfers in mobile applications: Network-wide origin, impact, and optimization. Proc. 21st international conference on World Wide Web. Lyon, France: 51–60.
- Quin X (2008) Performance comparisons of load balancing algorithms for I/O- intensive workloads on clusters. ACM Journal of Network and Computer Applications 31(1): 32–46.
- Rajagopalan R & Varshney PK (2006) Data-aggregation techniques in sensor networks: A survey. IEEE Communications Surveys & Tutorials 8(4): 48–63.
- Rao A, Lakshminarayanan K, Surana S, Karp R & Stoica I. (2003) Load balancing in structured P2P Systems. Proc. International Workshop on Peer-to-Peer Systems. Berkeley, CA, USA: 68–79.
- Ratnasamy S, Francis P, Handley M, Karp R & Shenker S (2001) A scalable content-addressable network. Proc. Conference on Applications, Technologies, Architectures and Protocols for Computer Communications. San Diego, CA, USA: 161–172.
- Rault T, Bouabdallah A & Challal Y (2014) Energy efficiency in wireless sensor networks: A top-down survey. Computer Networks 67: 104–122.
- Ravi N, Scott J, Han L & Iftode L (2008) Context-aware battery management for mobile phones. Proc. IEEE 6th Annual Conference on Pervasive Computing and Communications. Hong Kong: 224–233.
- Rhea S (2005) OpenDHT: A Public DHT Service. Ph.D. dissertation, Dept. Computer Science, University of California, Berkeley, USA.
- Rieche S, Niedermayer H, Götz S & Wehrle K (2005) Reliability and load balancing in distributed hash tables. Lecture Notes in Computer Science 3485: 119–135.
- Rieche S, Petrak L & Wehrle K (2004) A thermal-dissipation-based approach for balancing data load in distributed hash tables. Proc. 29th Annual IEEE Conference on Local Computer Networks. Tampa, FL, USA: 15–23.
- Rowstron A & Druschel P (2001) Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. Lecture Notes in Computer Science 2218: 329–350.
- Saariinen A, Siekkinen M, Xiao Y, Nurminen JK, Kemppainen M & Hui P (2012) Can offloading save energy for popular apps?. Proc. 7th ACM international Workshop on Mobility in the Evolving Internet Architecture. Istanbul, Turkey: 3–10.
- Sakaryan G & Unger H (2004) Topology evolution in P2P distributed networks. Proc. IASTED International Multi-Conference on Applied Informatics. Innsbruck, Austria: 791–796.

- Saravanakumar E & Prathima G (2010) A novel load balancing algorithm for computational grid. *Proc. International Conference on Innovative Computing Technologies*. New Delhi, India: 1–6.
- Schlomann B, Eichhammer W & Stobbe L (2015) Energy saving potential of information and communication technology. *International Journal of Decision Support Systems*, 1(2): 152–163.
- Serbu S, Bianchi S, Kropf P & Felber P (2007) Dynamic load sharing in peer-to-peer systems: When some peers are more equal than others. *IEEE Journal of Internet Computing* 11(4): 53–61.
- Shen H & Xu CZ (2008) Hash-based proximity clustering for load balancing in heterogeneous DHT networks. *Journal of Parallel and Distributed Computing* 68(5): 686–702.
- Shye A, Scholbrock B & Memik G (2009) Into the wild: Studying real user activity patterns to guide power optimizations for mobile architectures. *Proc. 42nd Annual IEEE/ACM International Symposium on Microarchitecture*. New York, NY, USA: 168–178.
- Siekkinen M, Hiienkari M, Nurminen JK & Nieminen J (2012) How low energy is Bluetooth low energy? Comparative measurements with zigbee/802.15.4. *Proc. IEEE Wireless Communications and Networking Conference Workshops*. Paris, France: 232–237.
- Siekkinen M, Hoque M, Nurminen J & Aalto M (2013) Streaming over 3G and LTE: How to save smartphone energy in radio access network-friendly way. *Proc. 5th Workshop on Mobile Video*. Oslo, Norway: 13–18.
- Soltani N, Mousavi K, Sharifi M & Mirtaheri SL (2012) A dynamic popularity-aware load balancing algorithm for structured P2P systems. *Lecture Notes in Computer Science* 7513: 77–84.
- Song J-W & Yang S-B (2006) Improved load balancing algorithms in DHT-based dynamic P2P systems. *Proc. 4th IASTED International Conference on Communications, Internet, and Information Technology*. St. Thomas, Virgin Islands, USA: 218–224.
- Srisuresh P & Egevang K (2001) The traditional IP network address translator (Traditional NAT). *IETF RFC* 3022.
- Stoica I, Morris R, Karger D, Kaashoek M & Balakrishnan H (2001) Chord: a scalable peer-to-peer lookup service for internet applications. *Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. San Diego, CA, USA: 149–160.
- Stutzbach D & Rejaie R (2006) Understanding churn in peer-to-peer networks. *Proc. 6th ACM SIGCOMM Conference on Internet measurement*. Rio de Janeiro, Brazil: 189–201.
- Surana S, Godfrey B, Lakshminarayanan K, Karp RM & Stoica I (2006) Load balancing in dynamic structured peer-to-peer systems. *P2P Computing Systems*, 63(3): 217–240.
- Trajkovska I, Rodriguez J-S & Velasco A-M (2010) A novel P2P and cloud computing hybrid architecture for multimedia streaming with QoS cost functions. *Proc. 18th International Conference on Multimedia*. Klagenfurt, Austria: 1227–1230.
- Veciana G & Yang X (2003) Fairness, incentives and performance in peer-to-peer networks. *Proc. 41st Annual Allerton Conference on Communication, Control and Computing*. Monticello, IL, USA: 367–378.

- Vergara EJ, Nadjm-Tehrani S & Prihodko M (2014) EnergyBox: Disclosing the wireless transmission energy cost for mobile devices. *Sustainable Computing: Informatics and Systems* 4(2): 118–135.
- Vergara EJ, Nadjm-Tehrani S (2013) EnergyBox: A Trace-Driven Tool for Data Transmission Energy Consumption Studies. In: Pierson JM, Da Costa G, Dittmann L (ed) *Energy Efficiency in Large Scale Distributed Systems (LNCS 8046)*. Springer, Germany: 19–34.
- Wang S, Vasilakos A, Jiang H, Ma X, Liu W, Peng K, Liu B & Dong Y (2011) Energy efficient broadcasting using network coding aware protocol in wireless ad hoc network. *Proc. IEEE International Conference on Communications*. Kyoto, Japan, 1–5.
- Xing G, Lu C, Zhang Y, Huang Q & Pless R (2007) Minimum power configuration for wireless communication in sensor networks. *ACM Transactions on Sensor Networking* 3(2): 11.
- Xu Z, Min R & Hu Y (2003) HIERAS: a DHT based hierarchical P2P routing algorithm. *Proc International Conference on Parallel Processing*. Kaohsiung, Taiwan: 187–194.
- Yang B & Garcia-Molina H (2003) Designing a super-peer network. *Proc International Conference on Data Engineering*. Bangalore, India: 49–60.
- Yang M & Fei, Z (2009) A novel approach to improving search efficiency in unstructured peer-to-peer networks. *Journal of Parallel and Distributed Computing* 69(11): 877–884.
- Zhang X & Shin KG (2012) E-MiLi: Energy-minimizing idle listening in wireless networks. *IEEE Transactions on Mobile Computing* 11(9): 1441–1454.
- Zhang Y, Shakhsheer Y, Barth, A, Powell HC, Ridenour SA, Hanson MA, Lach J & Calhoun BH (2011) Energy efficient design for body sensor nodes. *Low Power Electronics and Applications* 1(1), 109–130.
- Zhao B, Huang L, Stribling J, Rhea S, Joseph A & Kubiawicz J (2004) Tapestry: a resilient global-scale overlay for service deployment. *IEEE Journal of Selected Areas in Communications* 22(1): 41–53
- Zhelev R & Georgiev V (2011) A DHT-based scalable and fault-tolerant cloud information service. *Proc. 5th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. Lisbon, Portugal: 55–60.
- Zhou Y & Koyanagi K (2013) Improving routing load fairness in structured P2P overlay networks. *International Conference on Advanced Communication Technology*. Pyeongchang, Korea: 724–728.
- Zhu Y & Hu Y (2004) Towards efficient load balancing in structured P2P systems. *Proc. 18th International Parallel and Distributed Processing Symposium*. Santa Fe, NM, USA: 349–361.
- Zhu Y & Hu Y (2005) Efficient, proximity-aware load balancing for DHT-based P2P systems. *IEEE Transactions on Parallel and Distributed Systems* 16(4): 349–361.

Original publications

- I Kassinen O, Harjula E, Korhonen J & Ylianttila M (2009) Battery Life of Mobile Peers with UMTS and WLAN in a Kademlia-based P2P Overlay. Proc. 20th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. Tokyo, Japan: 662–665.
- II Ou Z, Harjula E, Kassinen O & Ylianttila M (2010) Performance Evaluation of a Kademlia-based Communication-oriented P2P System under Churn. Computer Networks 54 (5): 689–705.
- III Harjula E, Koskela T & Ylianttila M (2011) Comparing the Performance and Efficiency of Two Popular DHTs in Interpersonal Communication. Proc. IEEE Wireless Communications and Networking Conference. Cancun, Mexico: 2173–2178.
- IV Harjula E, Kassinen O & Ylianttila M (2012) Energy Consumption Model for Mobile Devices in 3G and WLAN Networks. Proc. 9th IEEE Int. Conference on Consumer Communications and Networking. Las Vegas, NV, USA: 532–537.
- V Harjula E, Gurtov A, Koskela T, Ojala T & Ylianttila M (2016) Energy-Aware Load Monitoring for Improving Battery Life of Mobile Peer-to-Peer Nodes. *Manuscript*.
- VI Harjula E, Leppänen T, Ojala T & Ylianttila M (2014) ADHT: Agent-based DHT Architecture for Constrained Devices. Proc. 2014 IEEE Global Communications Conference. Austin, TX, USA: 2763–2769.

Reprinted with permission from IEEE (I, III, IV, VI) and Elsevier (II).

Original publications are not included in the electronic version of the dissertation.

- 556. Omran, Mamdouh (2015) Microwave dephosphorisation of high phosphorus iron ores of the Aswan region, Egypt : developing a novel process for high phosphorus iron ore utilization
- 557. Kiljander, Jussi (2016) Semantic interoperability framework for smart spaces
- 558. Aula, Matti (2016) Optical emission from electric arc furnaces
- 559. Ferdinand, Nuwan Suresh (2016) Low complexity lattice codes for communication networks
- 560. Xue, Qiang (2016) Analysis of near-optimal relaying schemes for wireless tandem and multicast relay networks
- 561. Rautio, Anne-Riikka (2016) On the stability of carbon nanotube and titania nanowire based catalyst materials : from synthesis to applications
- 562. Kuokkanen, Ville (2016) Utilization of electrocoagulation for water and wastewater treatment and nutrient recovery : techno-economic studies
- 563. Huusko, Jarkko (2016) Communication performance prediction and link adaptation based on a statistical radio channel model
- 564. Nguyen, Vu Thuy Dan (2016) Transmission strategies for full-duplex multiuser MIMO communications systems
- 565. Keränen, Pekka (2016) High precision time-to-digital converters for applications requiring a wide measurement range
- 566. Koivuranta, Elisa (2016) Optical monitoring of flocs and filaments in the activated sludge process
- 567. Lohikoski, Päivi (2016) Information processing in global virtual NPD projects
- 568. Kauppila, Osmo (2016) Integrated quality evaluation in higher education
- 569. Kisko, Anna (2016) Microstructure and properties of reversion treated low-Ni high-Mn austenitic stainless steels
- 570. Postila, Heini (2016) Peat extraction runoff water purification in treatment wetlands constructed on drained peatlands in a cold climate
- 571. Happonen, Tuomas (2016) Reliability studies on printed conductors on flexible substrates under cyclic bending
- 572. Soderi, Simone (2016) Evaluation of industrial wireless communications systems' security

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

S E R I E S E D I T O R S

A SCIENTIAE RERUM NATURALIUM

Professor Esa Hohtola

B HUMANIORA

University Lecturer Santeri Palviainen

C TECHNICA

Postdoctoral research fellow Sanna Taskila

D MEDICA

Professor Olli Vuolteenaho

E SCIENTIAE RERUM SOCIALIUM

University Lecturer Veli-Matti Ulvinen

F SCRIPTA ACADEMICA

Director Sinikka Eskelinen

G OECONOMICA

Professor Jari Juga

H ARCHITECTONICA

University Lecturer Anu Soikkeli

EDITOR IN CHIEF

Professor Olli Vuolteenaho

PUBLICATIONS EDITOR

Publications Editor Kirsti Nurkkala

UNIVERSITY of OULU
OULUN YLIOPISTO



ISBN 978-952-62-1248-7 (Paperback)

ISBN 978-952-62-1249-4 (PDF)

ISSN 0355-3213 (Print)

ISSN 1796-2226 (Online)