Department of Communications and Networking

On Providing Energy-efficient Data Transmission to Mobile Devices

Le Wang

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 40/2016

On Providing Energy-efficient Data Transmission to Mobile Devices

Le Wang

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Electrical Engineering for public examination and debate in Auditorium S1 at the Aalto University School of School of Electrical Engineering (Espoo, Finland) on the 1st of April 2016 at noon.

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

Supervising professor
Professor Jukka Manner

Preliminary examiners
Professor Mikko Valkama, Tampere University of Technology, Finland
Professor Navid Nikaein, Eurecom, France

Opponent
Professor Tommi Mikkonen, Tampere University of Technology, Finland

Aalto University publication series
DOCTORAL DISSERTATIONS 40/2016

ISBN 978-952-60-6685-1 (printed)
ISBN 978-952-60-6686-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6686-8

Unigrafia Oy
Helsinki 2016

Finland

441 697
Printed matter

Aalto University

Abstract

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

Author
Le Wang

Name of the doctoral dissertation
On Providing Energy-efficient Data Transmission to Mobile Devices

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 40/2016

Field of research Networking Technology

Manuscript submitted 18 March 2015 **Date of the defence** 1 April 2016

Permission to publish granted (date) 9 November 2015 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

The transformation from telephony to mobile Internet has fundamentally changed the way we interact with the world by delivering ubiquitous Internet access and reasonable cost of connectivity. The mobile networks and Internet services are supportive of each other and together drive a fast development of new services and the whole ecosystem. As a result, the number of mobile subscribers has skyrocketed to a magnitude of billions, and the volume of mobile traffic has boomed up to a scale no-one has seen before with exponential growth predictions.

However, the opportunities and problems are both rising. Therefore, to enable sustainable growth of the mobile Internet and continued mobile service adaption, this thesis proposes solutions to ensure that the reduction of overall environmental presence and the level of QoE are mutually addressed by providing energy-effient data transmission to mobile devices.

It is important to understand the characteristics of power consumption of mobile data transmission to find opportunities to balance the energy consumption and the growth of mobile services and the data volumes. This research started with power consumption measurements of various radio interfaces and investigation of the trade-off between computation and communication on modern mobile devices. Power consumption models, state machines and the conditions for energy-efficient mobile data transmission were proposed to guide the development of energy-saving solutions.

This research has then employed the defined guideline to optimise data tranmission for energy-efficient mobile web access. Proxy-based solutions are presented in this thesis, utilising several strategies: bundling-enabled traffic shaping to optimse TCP behaviour over congested wireless links and keep the radio interface in low power consumption states as much as possible, offloading HTTP-object fetching to shorten the time of DNS lookups and web content downloading, and applying selective compression on HTTP payload to further reduce energy consumption of mobile data transmission. As a result, the solutions dramatically reduce the energy consumption of mobile web access and download time, yet maintain or even increase user experience.

Keywords Power consumption, Mobile devices, Power modeling, Proxy

ISBN (printed) 978-952-60-6685-1 **ISBN (pdf)** 978-952-60-6686-8

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

Location of publisher Helsinki **Location of printing** Espoo **Year** 2016

Pages 200 **urn** http://urn.fi/URN:ISBN:978-952-60-6686-8

Preface

This thesis was carried out in the Department of Communications and Networking at Aalto University, and performed within the Tekes (Finnish Funding Agency for Technology and Innovation) and industry funded projects: the FI SHOK (Future Internet) project and the ECEWA (Energy and Cost Efficiency in Wireless Access) project, which all are gratefully acknowledged.

I want to address my sincere thanks to all people having been part of the journey, helping and supporting along the way for me to complete the work. I express my deepest gratitude to Prof. Jukka Manner, who has supervised the work through my journey towards the completion of this thesis. I thank Jukka for giving me the opportunity to start my doctoral study at Aalto University, working there, guiding me, and being always supportive. I specially appreciate that all his constructive feedback to my work and help to manuscript preparation of this thesis.

Dr. Anna Ukhanova, Dr. Evgeny Belyaev, Dr. Edward Mutafungwa and Mr. Eero Sillasto deserve special thanks for invaluable collaboration and rewarding discussions that make this work possible. In addition, I also want to thank Dr. Tero Isotalo for providing help and assistant to perform crucial measurements.

My warm thanks go to Prof. Mikko Valkama from Tampere University of Technology, Finland and Dr. Navid Nikaein from Eurecom, France for reviewing the manuscript of this thesis, and giving invaluable comments and suggestions, which are insightful and help me improve the didactical parts and structure of the thesis.

My thanks are extended to my colleagues at the Department of Communications and Networking, with whom I have had enlightening discussions about the various topics of this work. Special thanks go to Sebastian Sonntag, Timo Kiravuo, Lennart Schulte, Gautam Moktan and

Nuutti Varis for all the feedback, discussions, knowledge sharing, motivation and friendship. I would like to take this chance to thank Mr. Viktor Nässi for providing helping to setup measurement environment and being always available for support and discussion. I also thank the personnel of the Department for a pleasant and inspiring working atmosphere.

I own heartfelt thanks to my friends and parents for being so supportive, patient and a great source of strength. Thesis writing together with a daily job and family life is always a challenge during the last few year of the work. Therefore, special thanks belong to my beloved Dudu, first a girlfriend, then my fiancée, and now my wife and kid's mother, for the unwavering love, company, support and encouragement during all these years.

Le Wang

Helsinki, February 1, 2016,

Le Wang

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Le Wang, Jukka Manner. Energy Consumption Analysis of WLAN, 2G and 3G interfaces. In *Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications & Interntional Conference on Cyber, Physical and Social Computing*, Hangzhou, China, pp. 300-307, December 2010.

II Le Wang, Jukka Manner. Evaluation of Data Compression for Energy-aware Communication in Mobile Networks. In *Proceedings of the IEEE International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, Zhangjiajie, China, pp. 69-76, October 2009.

III Eero Sillasto, Le Wang, Jukka Manner. Using compression energy efficiently in mobile environment. In *Proceedings of the IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing*, Hangzhou, China, pp. 9-16, December 2010.

IV Iiro Jantunen, Joni Jantunen, Harald Kaaja, Sergey Boldyrev, Le Wang, Jyri Hämäläinen. System Architecture for High-speed Close-proximity Low-power RF Memory Tags and Wireless Internet Access. *International Journal On Advances in Telecommunications*, Vol. 4, Iss. 34, pp. 217-228, November 2011.

V Le Wang, Anna Ukhanova, Evgeny Belyaev. Power consumption analysis of constant bit rate data transmission over 3G mobile wireless networks. In *Proceedings of the 11th IEEE International Conference on ITS Telecommunications (ITST)*, St. Petersburg, Russia, pp 217-223, August 2011.

VI Anna Ukhanova, Evgeny Belyaev, Le Wang, Søren Forchhammer. Power consumption analysis of constant bit rate video transmission over 3G networks. *Computer Communications*, Vol. 35, Iss. 14, Elsevier, pp. 1695-1706, August 2012.

VII Le Wang, Bin Yu, Jukka Manner. Proxies for Energy-Efficient Web Access Revisited. In *Proceedings of the 2nd IEEE International Conference on Energy-Efficient Computing and Networking*, New York, USA, pp. 55-58, May 2011.

VIII Le Wang, Edward Mutafungwa, Puvvala Yeswanth, Jukka Manner. Strategies for Energy-Efficient Mobile Web Access An East African Case Study. In *Proceedings of the 3rd International ICST Conference on e-Infrastructure and e-Services for Developing Countries*, Zanzibar, Tanzania, pp. 74-83, November 2011.

IX Le Wang, Jukka Manner. Energy-efficient mobile web in a bundle. *Computer Communications*, Vol. 57, Iss. 17, Elsevier, pp 3581-3600, December 2013.

Author's Contribution

Publication I: "Energy Consumption Analysis of WLAN, 2G and 3G interfaces"

Wang and Manner created the idea for the article together. Wang was responsible for designing the evaluation system, conducting measurements and writing the most of the article.

Publication II: "Evaluation of Data Compression for Energy-aware Communication in Mobile Networks"

Manner initialised the idea of the article. Wang's contributions consisted of designing the evaluation system, performing experiments, analysing the results, and acting as the main author of the article.

Publication III: "Using compression energy efficiently in mobile environment"

Sillasto and Wang proposed the idea for the article together. Sillasto developed the model of partial compression initially and wrote Sections 2, 3 and 4. Wang further developed the model, designed the measurement system, conducted the evaluation and wrote Section 5. Wang also reviewed and edited the rest of the sections.

Publication IV: "System Architecture for High-speed Close-proximity Low-power RF Memory Tags and Wireless Internet Access"

I.Jantunen was the main driver of the article. Together with J. Jantunen, Kaaja and Boldyrev, he contributed to drafting and building the system architecture. Wang was mainly responsible for power consumption analysis and measurement of the architecture, and wrote Section IV.

Publication V: "Power consumption analysis of constant bit rate data transmission over 3G mobile wireless networks"

The research was done in cooperation between the authors. Wang and Ukhanova proposed the idea together. Wang created the model of power consumption of data transmission and verified the model with real measurements. Belyaev was the main driver of the uplink power consumption modelling. Wang and Ukhanova contributed to create a power model for RRC states. In this article, Wang wrote Section II and III and commented on the other sections. Wang also reviewed and edited the other sections.

Publication VI: "Power consumption analysis of constant bit rate video transmission over 3G networks"

This publication is an extended work of Publication V. The research was done in cooperation between the authors. Ukhanova was mainly responsible for Sections 1, 6 and 7. Belyaev was the main driver of Section 5 and Wang was mainly responsible for creating a power consumption model of data transmission, collaborating with other authors to create a power consumption model for RRC transition state machine and writing Sections 3 and 4. Wang also reviewed and edited the other sections.

Publication VII: "Proxies for Energy-Efficient Web Access Revisited"

Wang and Manner created the idea of the article together. Yu was responsible for the implementation and provided measurement results. Wang contributed to the system design, data analysing and acted as the main author of the article.

Publication VIII: "Strategies for Energy-Efficient Mobile Web Access An East African Case Study"

Mutafungwa proposed the idea of the article and wrote Sections 1 and 2. Yeswanth performed the measurements and provided data. Wang, together with Manner, designed the main structure of the article. Wang designed and implemented the system. He performed system evaluation and comparison as well as wrote Sections 3, 4 and 5.

Publication IX: "Energy-efficient mobile web in a bundle"

Wang was the main driver of the article. He contributed to provide the main idea of the paper, design and implement the system, conduct part of the experiment and act as the main editor of the paper.

List of Figures

List of Abbreviations

$1G$ The First Generation mobile phone networks

$2G$ The Second Generation mobile phone networks

$3G$ The Third Generation mobile phone networks

$3GPP$ Third Generation Partnership Project

$4G$ The Fourth Generation mobile phone networks

$AJAX$ Asynchronous JavaScript and XML

$AMPS$ Advanced Mobile Phone System

AP Access Point

$ARPU$ Average Revenue Per User

$CAGR$ Compound Annual Growth Rate

CAM Continuously Active Mode

$CDMA$ Code Division Multiple Access

$CDMA2000$ A family of 3G mobile technology standards

CSG Closed Subscriber Group

CSS Cascading Style Sheets

DNS Domain Name System

DOM Document Object Mode

DRX Discontinuous Reception

$EDGE$ Enhanced Data rate for GSM Evolution

EAP Explicitly Authenticated Proxy

EEP Energy- Efficient Proxy

$EV-DO$ Enhanced Voice-Data Optimised

FSM Finite State Machine

GHG GreenHouse Gas

$GIPS$ Giga-Instructions Per Second

$GPRS$ General Packet Radio Service

GPS Global Positioning System

GSM Global System for Mobile Communication

HBI Human-Battery Interaction

$HSPA$ High Speed Packet Access

$HTML$ HyperText Markup Language

$HTTP$ Hypertext Transfer Protocol

IBI Interactive Batter Interface

ICT Information and Communication Technology

IoT Internet of Things

$IS-136$ Interim Standard 136, a second-generation mobile phone system

$IS-95$ Interim Standard 95, a second-generation mobile phone system

ITU International Telecommunication Union

LTE Long-Term Evolution

$M2M$ Machine-to-Machine

MAC Media Access Control

NCP Network Connectivity Proxy

NEP Nokia Energy Profiler

NFC Near Field Communication

NFC Near Field Communication

NMT Nordic Mobile Telephone

$OFDMA$ Orthogonal Frequency-Division Multiple Access

$PAWP$ Power Aware Web Proxy

PDC Personal Digital Cellular

$PDCCH$ Physical Downlink Control Channel

PEP Performance Enhanced Proxy

PSM Power Saving Mode

QoE Quality of Experience

RAN Radio Access Network

$REST$ Representational State Transfer

RLC Radio Link Controller

RNC Radio Network Controller

RRC Radio Resource Control

RSS Really Simple Syndication

$TACS$ Total Access Communications System

$TD-SCDMA$ Time- Division-Synchronous CDMA

$TDMA$ Time Division Multiple Access

TIM Traffic Indication Map

TOP Tail Optimisation Protocol

TTI Transmission Time Interval

$UMTS$ Universal Mobile Telecommunications System

$UWBLEE$ Ultra-wideband Low End Extension, a wireless technology developed within MINAmI project

$VLSI$ Very Large Scale Integration

VMP Virtual-Machine based Proxy

$VoIP$ Voice over IP

WAP Wireless Application Protocol

$WBAN$ Wireless Body Area Network

$WCDMA$ Wideband Code Division Multiple Access

$WLAN$ Wireless Local Area Network

$WPAN$ Wireless Personal Area Network

$XHTML$ Extensible HTML

1. Introduction

We have been witnessing a decline in the sales of traditional desktops and notebooks, and the rise of mobile devices, signifying that the Post-PC era is about to begin. The Post-PC devices are featured with fast connectivity, portability, intuitive user interfaces, sensory perception to the surrounding world and easy accessibility to Internet services. These characteristics offer a more additive way for mobile users to consume Internet content, be in touch and stay distinguished.

Mobile devices are influencing people dramatically in many aspects. This is particularly true in regions where life and business already have widespread access to PCs. They are often served as a time filler for users' daily fragmented leisure time, while waiting or relaxing, and has also become one of the motors of the 21st century economy, providing ubiquitous means to reach global audiences and interact with customers. More importantly, mobile devices are enablers for Information and Communication Technologies (ICTs), to penetrate countries in all regions of the world, bridging the digital divide between information haves and have-nots. Nowadays, millions of users are only connected to the Internet through mobile devices, especially in the most emerging areas of Asia and Africa, where the penetration of fixed-line Internet is minuscule, and electricity infrastructures are falling behind [1]. Mobile networks provide much wider coverage for Internet connectivity, thus enabling constant access to information and increasing the level of access to the information for a larger number of users. Easy access to the Internet lowers the barrier of being connected with services, education, health care, civic engagement and much more.

As a chemical reaction of the Internet and mobile usage, mobile Internet has dramatically and profoundly changed the way we learn, think and react with the world. Not since Johannes Gutenberg invented the printing press, or Alexander Graham Bell the telephone, has a human invention empowered so many and offered such great possibility for benefiting humankind. Mobile technologies lift Internet services into the next level. Meanwhile, the advance is pushing mobile devices to have faster CPU/GPU, more powerful hardware, higher resolution display, bigger storage, and more powerful software. However, a few areas that are still lacking and under development are battery- and power-saving technologies.

1.1 Research Motivation, Methodology and Goals

The fast development of mobile services, along with the advance of radio communication, hardware manufacture and integration technologies, is pushing mobile devices to hold powerful computing processors, massive storage memories, radio interfaces and many different kinds of hardware components. Intuitively, the average number of applications per smartphone is 41, up from 32 last year [2], and the battery life of a smartphone lasts barely over a day, because the average user looks at the phone 150 times a day according to Tomi Ahonen's speech during the Mobile Web Africa conference, 2013. Needless to say, the increasing number of hardware components and installed software are together making the mobile devices much more power-hungry than ever before. The concerns over the fast development pace of mobile devices and Internet services are not only limited to short battery life, but also cover other areas, which lead to the ***motivations*** of this research work as listed below:

1. The ever-increasing demand for mobile devices and wireless services leads to increased energy consumption on mobile devices. Therefore, reduction in energy consumption is of great importance. The focus of energy-saving techniques has been on energy conservation in mobile systems, essentially due to limited battery technology. The major technological challenge is to store a large amount of energy in batteries for increasingly complex mobile devices and yet still deliver reasonable size and weight. Nonetheless, so long as batteries continue to be based on electro-chemical processes, limitations of power density and limited lifetime will be difficult to overcome, making it hard to cater to mobile devices with power-hungry features. Even though new battery technologies may eventually come, designing more energy-efficient systems will

still be important. This has already presented a significant barrier to the continued adoption of mobile Internet services and sustainability of an acceptable Quality of Experience (QoE) for mobile Internet.

2. The fast development of the mobile communication industry is also at the cost of significant carbon footprint and electricity cost. The whole ICT sector has been estimated to represent 3% of total carbon emissions in the world [3] and the total electricity consumption in ICT consumes $5 \sim 10\%$ of the total worldwide electricity consumption [4]. Thus, there is a strong environmental and economic incentive to reduce energy consumption in this area. Even though mobile devices account for a small fraction of the total energy consumption, and electricity cost is not a prime confer of mobile users, it becomes a clear expenditure considering that the rising number of mobile users and devices can lead to a large aggregate electricity consumption and GreenHouse Gas (GHG) emission.

3. Moreover, energy-saving techniques play a critical role, and have been gaining social impact and benefits to the society at large for third world countries. In some Asian and African countries, the lack of readily available access to electricity is proving to be a major barrier to both adoption and usage of mobile Internet. Throughout East Africa, the fraction of the population with mobile Internet access, but no access to electricity, is growing, particularly in rural areas, where less than 3% of the rural population has access to electricity [1]. It is clear that the very limited access to electricity and unreliable electricity supply worsens the problem in these regions. Therefore, energy-saving solutions that prolong the mobile battery life are now very essential.

In order to cater to the above-mentioned concerns, energy saving technologies have been broadly studied by industry and academia, which roughly fall into the following categories: hardware design, operating system, middleware, application- and user-related solutions. Irrespective of which solution for energy savings, a principle ***research methodology*** [5] is followed: “1) a rich measurement and monitoring infrastructure; 2) accurate analysis tools and models that predict resource usage and identify trends and causal relationships, and provide prescriptive feedback; 3) control algorithms and policies that leverages the analysis above to meaningfully control power (and heat), ideally coordinated across layers”. As a holistic system, the mobile device can be broken down by its main hardware components into CPU, memory, touchscreen, graphics hardware, audio, storage, and various networking interfaces, which are the main energy consumers of the device. A high level of hardware integration is able to reduce the power consumption, size and weight of devices. Advanced mobile chips are integrated CPU, graphics accelerators as well as Global Positioning System (GPS) chips and much more. However, a major problem with current mobile devices is high power consumption when using networking interfaces to transmit data, especially with a Wireless Local Area Network (WLAN) interface and cellular interfaces. The prior study [6] demonstrates that the networking interfaces are one of the biggest energy consumers. Thus, this thesis focuses on the ***research scope*** of mobile data transmission on mobile devices. By the time this research was conducted, the Fourth generation (4G)/Long Term Evolution (LTE) networks were neither largely deployed nor available to us. Thus, the energy consumption in LTE is out of the scope. Yet, we discuss the applicability of our results in an 4G/LTE environment in Chapter 5. Besides, this dissertation emphasises on investigation of mobile data transmission and its optimisation for energy saving on mobile devices. Thus, the comparison of different mobile operating systems, CPU architectures, and radio access networks on the energy consumption of mobile devices is out of the scope as well. More specifically, the work is categorised into the following ***research areas***: 1) understanding energy consumption characteristics of networking interfaces for mobile data transmission; 2) providing energy-efficient mobile data transmission.

1.2 Contributions

Generally, the research was conducted in four distinct directions that complemented each other. The structure of the research areas and the focus area of each publication can be seen in Figure 1.1.

In-depth understanding of power consumption characteristics of real mobile devices is a prerequisite of building energy consumption models, developing energy-efficient protocols, algorithms, and energy saving solutions. The research started with understanding how energy is consumed when data are transmitted over wireless network interfaces.

Contribution 1

The observations presented in Publication I and Publication IV clearly

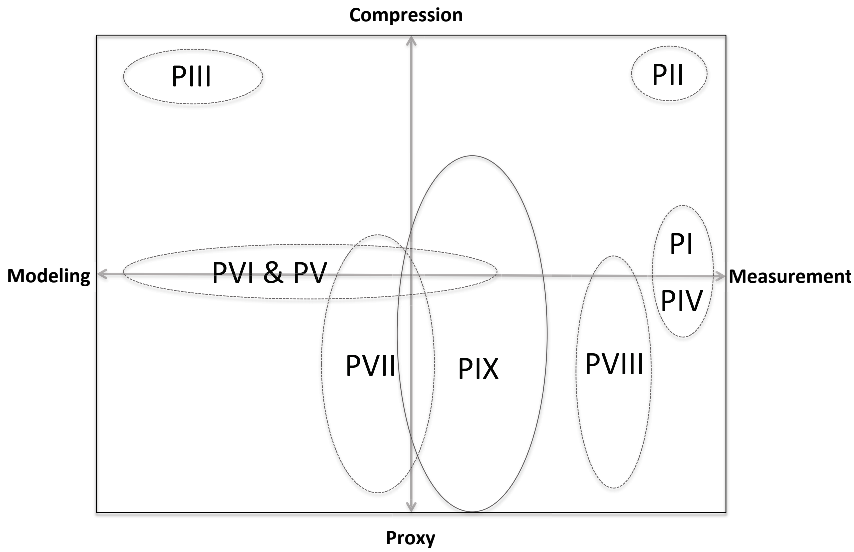


Figure 1.1. Research contributions and publications

suggests that transmission over the air is highly energy consuming and the transmission should be shaped into bursty chunks in order to keep radio interfaces in a low power consumption state as long as possible. Moreover, based on the fact that the energy consumed on a single bit transmission over wireless is over 1000 times greater than a single 32-bit CPU computation [7], we evaluated the trade-off between computation and communication on modern mobile devices for both uplink and downlink in Publication II, which shows that compression can be adaptively used to gain energy benefit when fulfilling certain conditions. Another observation from Publication V and Publication VI reflects that Radio Resource Control (RRC) in the Third Generation (3G) networks leads in efficient power consumption of data transmission for downlink.

Prior art

By the time the research was conducted, there were several studies already in the area. The early research related to measurement of power consumption of WLAN interfaces was reported in [8], that provides detailed results of the energy consumption of IEEE 802.11 wireless network interface in ad hoc network, and linear equations and some suggestions were given for designing energy-efficient protocols. In the research done by Ebert et al. [9], the power dissipation of wireless interface was measured for detailed power consumption pattern of sending and receiving packets with various transmission rate, packet size and RF transmission power in IEEE 802.11 wireless network. In paper conducted by Perrucci et al. [10], the authors measured power consumed in sending text messages and using voice services in the Second Generation (2G) and 3G networks. Another research conducted by Balasubramanian et al. [11] presents a measurement study of energy consumption of TCP data downloads in Global System for Mobile Communication (GSM), 3G and IEEE 802.11 wireless networks and proposes a protocol named TailEnder to reduce energy consumption of common mobile applications. In paper returned by Sharma et al. [12], the authors analysed energy consumption characteristics of General Packet Radio Services (GPRS)/ Enhanced Data rate for GSM Evolution (EDGE)/3G and WiFi radios on smartphones and proposes an architecture named Cool-Tether that builds a WiFi hotspot with a cloud-based gatherer and an energy-aware striper to provide energy-efficient, affordable connectivity.

In comparison to these studies, the focus of this dissertation is on investigating power consumption per bit of user data when sending or receiving data over various wireless links. As power consumption on hand-held devices differs from each other due to hardware and software related factors, an evaluation over modern hand-held devices provides a more timely understanding of data transmission in the view of energy efficiency, and it is also possible to offer a chance to explore new approaches for more energy savings. Therefore, it is necessary to conduct experiments based on the latest mobile device at the time. The most valuable contributions in Publication I and Publication IV are comprehensive measurements of power consumption and energy consumed per bit of 2G, 3G, IEEE 802.11 and short-range wireless interfaces when sending and receiving packets and corresponding analysis and comparisons.

Prior art [7, 13, 14] present that it is viable to explore new approaches for energy savings by applying data compression to mobile communication. As power consumption on hand-held devices differs from each other due to hardware and software related factors, an evaluation over modern hand-held devices provides a fresh understanding of data transmission and compression. The contribution in the work is to provide a timely evaluation of a wide number of compression schemes on various types of web content on modern mobile devices.

In previous work [15, 16], the the Third Generation Partnership Project (3GPP) transition state machine was analysed based on measurements, and the effect of different timer values on power consumption of devices was examined. Comparably, the work in this dissertation is to analyse the power consumption of each RRC state and propose a power consumption model for one of the most energy consuming state for downlink transmission. Based on the model, a parameter selection mechanism can be proposed to minimise power consumption of constant bit rate transmission on mobile devices.

Contribution 2

Based on the three main observations, the work proposed several solutions to improve the energy efficiency of mobile data transmission correspondingly. Publication II formalises conditions for energy-efficient compression in mobile data transmission and suggests having partially compressed data for uplink data transmission. Publication V presents a parameter selection criteria, taking signal overhead and transition delay into consideration for 3GPP state transition machine to minimise power consumption of constant bit rate transmission on mobile devices. The work was extended in Publication VI to reduce energy consumption of video streams.

Prior art

Compared to previously mentioned studies [7, 13, 14], the contribution in Publication II takes into account that there are limitations both for communication and compression on mobile devices. These factors have to be reconsidered when developing an energy-efficient way of utilising data compression. The work formalises compression conditions for energy-efficient data transmission, and proposes to use partial compression.

Several works [15, 16, 17] investigate the optimisation task of the RRC state machine parameter selection and explores the optimal timer values to save energy. The work presented in Publication V and Publication VI propose a power consumption model for the RRC transition state machine and present a parameter selection criteria taking signal overhead and transition delay into consideration. Furthermore, the work extended to video transmission, where experimental results show that in this case the proposed solution allows to save power on video transmission.

Contribution 3

Since mobile web content is taking a considerate amount of Internet traffic, Publication VII and Publication VIII analyse and evaluate different energy-saving strategies for energy-efficient web access, which led the research direction to designing and implementing a proxy-based architecture for energy-efficient mobile web access. After that, the research focused on solving the problem holistically from several levels to achieve better efficiency. Thus, Publication IX proposes a solution taking consideration of RRC state in the MAC layer, traffic scheduling in the Transport layer and header compression in the Application layer.

Prior art

In a previous study done by Qian et al. [18], the Tail Optimisation Protocol (TOP) dynamically determines the values of the inactivity timers and terminates the tail energy if no further data transmission is needed. The approach predicts the end of traffic transmission to utilise fast dormancy to configure the radio to low power consumption states. Another type of solution is to aggregate traffic with prefetching, such as TailEnder [11], which aggregates prefetched data of delay-tolerant applications into large ones so that the tail energy is reduced. TailTheft [19, 20] uses a virtual tail time mechanism for making better decisions on when to perform prefetching and when to terminate tail transmission in order to fully utilise unused tail time and reduce total transmission time.

In comparison with the work in this dissertation, our study utilises the principle of split TCP to optimise Hypertext Transfer Protocol (HTTP) downloading over wireless links, and focuses on leveraging RLC buffer threshold to keep the mobile device in lower power consumption state.

Some other studies of energy-efficient web browsing have also been reported in prior work. For example, the Power Aware Web Proxy (PAWP) [21] designs an architecture to schedule web traffic so that WLAN interface can be turned off and remain in a low power state for longer periods after active data exchange between the mobile device and proxy. The Network Connectivity Proxy (NCP) [22] proposes a SOCKS-based proxy on behalf of a mobile device to maintain full network presence, allowing the device to stay idle and in a low power consumption state. The proxy preserves TCP connections and UDP flows for the sleeping device to achieve significant energy savings. Another approach is reported on paper [23] by Zhao et al., which proposes an architecture called Virtual-Machine based Proxy (VMP). VMP shifts computation from the mobile device to the proxy in 3G networks, where the proxy handles HTTP requests, replies, execution of JavaScript and rendering of web objects. Then, a screenshot of the rendered web page is compressed, transferred and displayed on the mobile device. Since the heavy lifting is offloaded to the proxy, energy savings become possible.

In comparison with the proxy-based solutions for energy-efficient web browsing, our solution utilises bundling and header compression to cater to the energy consumption characteristics of WLAN and 3G networks. The selective compression applied is lossless compression, which does not alter original web content and still provides significant improvement of energy consumption along with other techniques. In addition, the solution does not require any modification on web browser and web servers, thus it can be deployed incrementally.

The contributions in this dissertation are primarily seeking to enable lower energy consumption for devices operating already in current networks, without needs to modify the basic operation and standardisation of the existing radio networks. However, the insights and results presented in the dissertation can be valuable inputs for future development of radio technologies and standardisation organisations.

1.3 Structure of the Thesis

This dissertation consists of a summary and nine original articles. The rest of the thesis is structured as follows: in Chapter 2, the status of mobile communication and services as well as the rising issues regarding energy consumption, are presented. Chapter 3 presents the understanding of power consumption of mobile data transmission in the perspective of power consumption characteristics of radio interfaces, and depicts the energy trade-off between computation and communication. In Chapter 4, two solutions to reduce energy consumption of mobile data transmission are presented, namely using compression and using a performance-enhanced proxy. After that, Chapter 5 summarises the research results, discusses a few open questions and presents future research directions.

2. Evolution of Mobile Internet

The Internet is increasingly wireless and continues its explosive growth with non-PC devices from mobile phones to tablets, wearable electronic devices to Machine-to-Machine (M2M) devices. With the fast growth of mobile Internet, almost half of all IP traffic will originate with non-PC devices by 2017, raising new opportunities and challenges for mobile operators, service providers as well as mobile users [24]. This chapter starts with presenting the evolution of mobile Internet in Section 2.1 to elaborate the development of mobile wireless communications and the corresponding adaption of Internet services. Then, Section 2.2 focuses on the pains and challenges along with the evolution, especially from an energy consumption point of view. After that, Section 2.3 shortly summarises this chapter.

2.1 From Telephony towards Mobile Internet

As reflected by the following listed facts, mobile device uptake has grown at a strong pace around the world.

- Global PC shipments dropped 11.2% to 79.2 million units in the first quarter of 2013 compared to the same period in 2012 - the steepest decline since 1994 [25]. There is no clear sign of recovery since the shipments only reached 79.4 million units in the third quarter of 2014 [26].

- There were 7.1 billion mobile subscriptions worldwide in 2014. The growth is led by China and India, which now account for over 30% of world subscriptions [27, 28].

- By 2020, the number of mobile devices is expected to surpass the world's population and reach 9.5 billion [28].

- Mobile data traffic will surpass long-haul traffic in 2015 and will continue to grow and account for 64% of tattle IP traffic in 2018 [24].

- There were over 1.2 Billion people accessing web content from their mobile phones in 2013 [27].

- In 2014, Facebook was receiving fewer PC visitors than mobile visitors, showing a clear sign of the transformation in social network's business [29].

- 25% of US web users, 59% of India web users and 85% of African web users are mobile-only web users [27].

As can be seen, the Internet traffic characteristics, the carrier of the traffic and the way users access the Internet have been dramatically changing during recent years. As a result of high demand for growing subscriber base and emerging Internet services while moving from telephony to mobile Internet, infrastructure of mobile network is fundamentally changing to be more service-centric rather than transport-centric. Internet services and applications are on the rise, allowing numerous service and content providers to be more creative in offering new services that meet the user demands and desires. The mobile networks and Internet services are supportive of each other for fast development. The following sections describe the evolution of mobile Internet from its mobile networks to its services.

2.1.1 Evolution of Mobile Communication Networks

This section describes the generations of mobile communication networks, in particular the evolution of radio technologies, and the other wireless technologies as a complementary system. As a whole, all the radio technologies should be integrated to deliver services across different networks with high spectral and bandwidth efficiency.

The Evolution of the Mobile Network From 1G to 4G

During the early '80s, the First Generation (1G) mobile systems, based on analog radio transmission techniques, were deployed to provide voice services using Frequency Division Multiple Access (FDMA), and used circuit-switched technologies in the network core.

The evolution actually started in the early '90s, with the replacement of the analog mobile network with the digital one, 2G mobile systems, which are still in wide use today to provide data and voice services. This generation allows mobile users to be accommodated in radio spectrum through either FDMA (IS-95) and Time Division Multiple Access (TDMA) (GSM, IS-54, PDC). GSM as one of the 2G digital wireless telephone technologies was initially from Europe but has been widely spread to almost all countries. It was originally based on circuit switched network optimised for full duplex voice telephony. The "2.5G", GPRS keeps the GSM radio modulation, frequency bands and frame structure, but implements a packet-switched network domain in addition to circuit-switched domain. EDGE is considered as "2.75G" technology, with a new radio modulation scheme introduced to triple the bandwidth offered by GPRS [30].

To provide a truly mobile broadband experience globally, 3G was defined by International Telecommunication Union (ITU) with the IMT-2000 standard, which has been gradually fulfilled by 3GPP [31]. Two main proposed systems for 3G are Code Division Multiple Access (CMDA) multi-carrier based CDMA2000, and FDD and TDD based Universal Mobile Telecommunications (UMTS), which deploys Wide-band CDMA (WCDMA) and Time-Division-Synchronous CDMA (TD-SCDMA) separately. Later on, High Speed Packet Access (HSPA) utilises higher order modulation (64QAM) and multiple-antenna technique (MIMO for "Multiple-Input and Multiple-Output") to achieve high speed in both downlink and uplink [32].

As a successor of 3G, 4G mobile network is to accomplish new levels of user experience of data communications using an All-IP design with "freedom and flexibility to select any desired service with reasonable QoS and affordable price anytime, anywhere" specified by ITU-R as IMT-Advanced specification. As defined in 3GPP, LTE-Advanced is based on an all-IP packet-switched network including Orthogonal Frequency-Division Multiple Access (OFDMA), MIMO, scalable channel bandwidth usage and link spectral efficiency to provide data rates up to 1.5 Gbit/s for uplink and up to 3 Gbit/s for downlink. Also, IEEE is evolving Worldwide Interoperability for Microwave Access (WiMAX) through IEEE 802.16m to meet 4G requirements [33].

The evolution path of mobile networks is elaborated in Figure 2.1. As the radio technologies advanced from CDMA and TDMA to OFDM and MIMO, the mobile network architecture has also been developed from circuit-switched network to packet-switched network, and towards All-IP with layered network architectures. As backhaul networks shift from the access layer to the distribution layer, the circuit-switched domain is eliminated, and an efficient delivery of packet-oriented multimedia services with higher data rates and lower latency is enabled [31].

WLAN and Other Wireless Technologies as a Complementary System

Cellular networks are currently limited by insufficient spectrum allocation, cell size trade-offs and costly infrastructures, which have become show-stoppers of cellular networks to be pervasive [34]. Femtocells have so far been deployed as coverage enhancements of cellular networks, especially for indoor users. However, the cost of data usage is likely to remain high, as the technologies are licensed spectrum-based. This, in turn, requires complementary access technologies to augment both coverage and capacity for affordable, flexible and ubiquitous communications. As predicted, mobile offload increases from 33% in 2012 to 46% in 2017, reaching 9.6 exabytes/month [24].

WLAN as one of the most prevalent unlicensed wireless technologies has been standardised with IEEEE 802.11 and branded as “Wi-Fi". IEEE 802.11n can provide bit-rates up to 600 Mbit/s and IEEE 802.11ac is able to support bit-rates up to almost 7 Gbit/s . Wi-Fi provides mobile devices Internet access with coverage of private homes, businesses or public spaces. Wi-Fi hotspots are also often considered as a key part of mobile infrastructure to offload data from 3G/4G networks.

Most of the mobile Internet traffic is generated by cellular and WLAN network users, but mobile traffic has also led to growth by communications between machines, sensors or mobile phones. M2M technologies are being used across a broad spectrum of industries, such as in smart grid for automated monitoring and control, vehicular telematics for navigation and diagnostics, and healthcare for recording a patient's blood pressure, heart rate and body temperature. These machine-generated data are automatically transmitted from machines to M2M servers to support cloud-based mass devices management and services either 1) directly though cellular/WLAN networks or 2) through short-range wireless networks. e.g. Wireless Personal Area Network (PAN) or Wireless Body Area Network (WBAN) networks, as the devices are sensitive to cost or power consumption. With short-range wireless technologies, an M2M gateway can collect and aggregate all the data from the devices, allowing a final up-

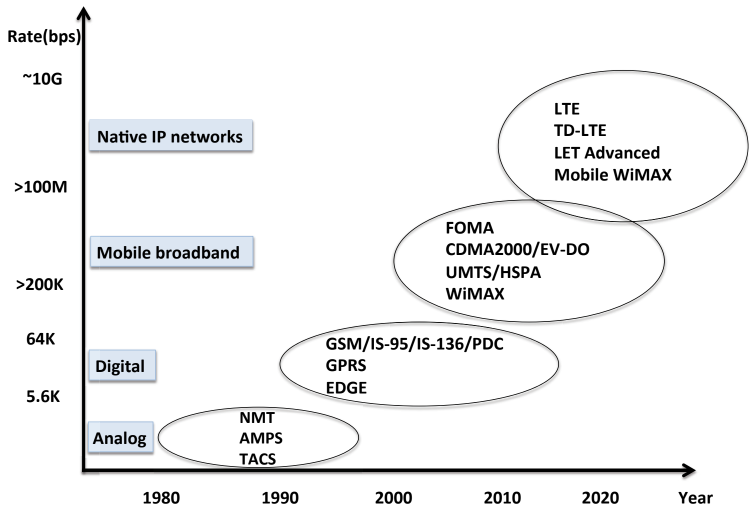


Figure 2.1. Evolution of mobile networks

link through cellular or WiFi connections. There are various technologies including Bluetooth, ZigBee, ANT, Near Field Communication (NFC) and Ultra-wideband Low End Extension (UWBLEE), whose specifications are listed in Table 2.1.

Satellite communication provides maritime, broadcasting, navigational, meteorological, aeronautical and mobile satellite services. Even though it has many advantages, such as large coverage and no geographic limitation, the power and bandwidth availability are severely limited under the mobile satellite communications environment. Therefore, satellite communication has been used as a complementary system and gap fillers, covering remote areas where there is no fixed or cellular networks [35]. Satellite communication has also proved to be an inalienable part of the mobile communication system in case of serious damage to infrastructure of terrestrial mobile communication is caused by natural disasters. However, satellite communication has the potential to become an alternative for ubiquitous communications since integration with terrestrial communication system, capacity, performance, spectrum efficiency and coverage are expected to be significantly improved in coming years [36].

Currently, cellular networks provide full coverage and consistent connectivity. Wi-Fi networks as hotspots offer high bit-rates and affordable access, and short-range networks provide interconnectivity between de-

Table 2.1. Wireless technologies with unlicensed-band

Technologies	Frequency Band	Max Rate	Range	Power	Standards
WiFi	2.45GHz, 3.6GHz and 5GHz	540Mbit/s	100m	High	IEEE 802.11
Bluetooth	2.4GHz	3Mbit/s	1m, 10m and 100m	Medium	IEEE 802.15.1
Bluetooth LE	2.4GHz	1Mbit/s	5-15m	Low	
ZigBee	2.4GHz, 868MHz, and 915MHz	250kbit/s	50m	Low	IEEE 802.15.4
ANT	2.45GHz	1Mbit/s	5m		Proprietary
NFC	13.65GHz	442kbit/s	2cm	Low	ISO 14443
RFID	860-930MHz, and2.45GHz	4Mbit/s	5cm	Low	ISO 18000-4
UWBLEE	900MHz, 7.9GHz	112Mbit/s	10cm	Low	

vices. The evolution direction of mobile networks is pervasive, spectrum efficient and with high bit-rates and also cheap costs. When the mobile network is moving towards becoming service-centric, it requires the network to transparently deliver differentiated services across a fully seamless network operating on diverse wireless technologies, with an IP-based backhaul in an optimum way, and to be able to handle rapidly increasing traffic in its backhaul.

2.1.2 Drivers of Mobile Internet Usage

Technical advances, both large and small, continue to reform mobile devices, transforming mobile phones from huge brick-like devices into stylish smartphones carried with us everyday. In particular, we have seen steady advances in mobile Internet services, bringing convenience, health, a new lifestyle and entertainment to people, productivity and cost efficiency to businesses, and safety and sustainability to societies. With the increasing number of mobile devices and services today, mobile usage is expanding rapidly with web content, audio, video and emergence of connected cars, drones and wearable electronics [37]. Fast declining costs of connectivity and ubiquitous Internet access is the fundamental technical enabler for modern Internet usage, and there are several other enablers to skyrocket mobile connected devices to a magnitude of billions [38], which will be discussed in the following paragraphs.

Mobile Web: Since Sir Tim Berner-Lee invented the first web browser in 1990 [39], web technology has shifted from Web 1.0 to Web 2.0, from a static, non-interactive way of accessing Internet information to a social revolution in the use of web technologies [40]. In the era of Web 2.0, the rise of social networking and user-generated content has engaged users in accessing web-based services. There are several web-related concepts encompassed by the umbrella of Web 2.0, including techniques (blogs, Really Simple Syndication (RSS), WiKis, mashups, tags (folksonomies) and social networking), standards (XHTML, CSS, REST, and HTML5), and tools (AJAX, mashup APIs, WiKi engines) [41].

Web 2.0's extravagance is not suited to mobile devices's limited battery life, relatively small screens and computing resources. These facts lead to mobile Web 2.0 as a successor to Web 2.0 to cope with the limitations and leverage the opportunities of location-based and other environment-aware services [42]. After the first commercial mobile web browser, NetHopper was launched in 1996 [39], microbrowsers such as Wireless Application Protocol (WAP) [1] and NTT DoCoMo's i-mode browser [2], enabled mobile users to interact with mobile service providers via cellular networks. Nowadays, WebKit [3], Presto [4] and Gecko [5]-based mobile web browsers, together with Web 2.0 trends, introduce new QoE of Internet services, leading a transition towards mobile Web 2.0 [43]. Mobile Web 2.0 is a framework of mobile Internet services with emphasis on delivering Web 2.0 services, especially mobile instant messaging, location-based services, mobile search and social networking to users via mobile web browsers [42].

The rise of mobile Web 2.0 and user generated content has accelerated the growth of mobile usage. The mobile device is an inherently personal device, which contains a huge amount of personal information and whereabouts, making it a logical extension for social networks and other collaborative Web 2.0 services. In 2011, 50% of the total active Twitter users were mobile users and they contributed 40% of all tweets [38]. According to Mary Meeker's report of "2013 Internet Trends" [37], mobile has helped Facebook increase mobile subscriptions by 54% and revenue by 43%. With more efficient advertisement, based on personal information collected from mobile devices, the rising mobile Average Revenue Per User (ARPU) has offset declining desktop ARPU, early 2013.

"Semantic Web" proposed by Tim Berners-Lee, Jim Handler and Ora Lassila in 2001 [44], is a framework to link and structure data on the

[1] WAP, http://technical.openmobilealliance.org/tech/affiliates/wap/wapindex.html
[2] i-mode browser, http://www.nttdocomo.co.jp/english/service/developer/make/content/browser/
[3] WebKit, http://www.webkit.org/
[4] Presto, http://www.opera.com/docs/specs/presto2.12/
[5] Gecko, https://developer.mozilla.org/en-US/docs/Mozilla/Gecko?redirectlocale=en-US&redirectslug=Geckowebkit

web, defined in such a way that they can be understood and exchanged not only between humans but also machines. Web 3.0, the next phase of the web evolution, is considered as an extension of Web 2.0 applications using semantic web technologies and linked data [45, 46]. With Web 3.0, rich Web 2.0 applications and social media will be brought to machines, especially mobile devices. The aggregation of human-generated data and machine-generated data will enable a new level of mobile Internet usage.

Mobile Apps: Mobile apps as pieces of software running on mobile devices were originally to provide add-on functionalities to mobile operating systems for general productivity, and the distribution of mobile content and services was dominated by mobile network operators. Even though NTT DoCoMo's i-mode environment had long been an example of success in mobile content distribution, the surge of mobile apps started when Apple released the iPhone in 2007 and the subsequent launch of the Apple App Store. The App store introduced a simple access to app marketplaces and an attractive revenue share model for developers [47, 48]. Since then, mobile app stores have become a primary way of distributing mobile apps, which is understandable, given that 300.000 apps were available in the Apple App Store and more than 160.000 were available at Google's Android Play (formerly Marketplace) at the end of 2010 [49]. By April of 2012, more than 25 billion apps were downloaded from the App Store and 15 billion downloads from Google Play, and the total number of app downloads is predicted to be over 44 billion by 2016 [50].

Mobile apps themselves not only become one of the major channels to deliver digital content and services to end users, but also drive the way for end users to communicate, shop, play and work, accelerating mobile Internet usage. Mobile apps, as one of the primary drivers of mobile Internet usage, have become a gateway to the Internet due to its convenience and effective delivery of personalised information [51]. Video traffic created by mobile apps like YouTube, combined with social services like Viddy[6] and collaborative services like Skype, are contributing a tremendous amount of mobile traffic. Besides, mobile app powered search, commerce, social networking, instant messaging, context-aware services and others are the main drivers in the rise of big data. For example, the total user base consuming location-based services will reach 1.4 billion, and mobile e-mail users are expected to reach 713 million by 2014 [52].

Cloud, M2M and New Opportunities: Cloud computing utilises vir-

[6]Viddy, http://http://www.viddy.com/

tualisation technologies and computing hardware to enable web-based, value-added services on a resource-shared infrastructure [53]. Cloud services in general have three predominant service models, utilising infrastructure, platform and software-as-a-service [54]. Despite most benefits, such as lowering cost of operation, centralised security control, agility of provisioning resources, cloud computing shifts heavy back-end development from developers and service providers to cloud, enabling cloud clients to interact with cloud services with web browsers or browser-based mobile apps. The cloud application uses a thin client on a mobile device, while the service logic and data reside in the cloud; Google Maps, YouTube, Wikipedia and thousands of others have been mobile-enabled in this way. Chromebook [7] is an extreme example of using thin clients and cloud services. In the last few years, cloud computing has had a momentous and remarkable growth. Up to 30% of top global companies will broker more than two cloud services by 2014, and 40% of mobile apps developed will leverage cloud mobile back-end services by 2016 [55]. The rise of cloud computing has created expectations of consuming cloud services anytime and anywhere from desktops, laptops and mobile devices.

In parallel to cloud computing, M2M communication is able to connect billions of sensors and other machines to the Internet. By leveraging the power of cloud computing, this communication has been introduced as "Internet of Things" (IoT) [56]. M2M empowers the areas of automotive navigation, telematics, metering, healthcare, tracking, payment, vending, security and more with centralised decision making and management within the cloud. Propelled by the development of IP-enabled devices and the advance of global mobile connectivity, the explosion of bandwidth-intensive M2M communication will fuel big data growth [57]. According to the Cisco Visual Networking Index [24], industrial segments of healthcare and automotive are expected to experience 74% and 42% Compound Annual Growth Rate (CAGR) from 2012 to 2017. Moreover, sensor-enabled wearable and flyable attributes [37], augmented reality [58], and mobile payments are also catalysts for boosting mobile Internet usage.

2.1.3 Trends of Mobile Internet Usage

The development of mobile communication networks and related technologies is booming up the volume of mobile traffic, which is expected to

[7] Chromebook, http://www.google.com/intl/en/chrome/devices/

experience an immense explosion in the following years based on the current trends of the number of mobile subscriptions and devices, and growing mobile Internet services. Global mobile Internet traffic grew 70%, reaching 885 petabytes per month in 2012, and the traffic is expected to increase 13-fold by 2017, reaching 11.2 exabytes monthly. The mobile Internet traffic will grow at a CAGR of 66% from 2012 to 2017 and contributes 68% of the total Internet traffic by 2017 [24, 59].

Among various types of traffic, video is the largest contributor to mobile Internet traffic. The amount of mobile video traffic will increase 16-fold between 2012 and 2017, accounting for 66% of total mobile Internet traffic by 2017 [24]. The boosting mobile video traffic is foreseen to be driven by: 1) emerging fast network speed (HSPA and LTE), 2) increasing video quality (HDTV and 3D), 3) larger screens of mobile devices, 4) more convenient video transmission technologies(HTML5 and WebRTC), and 5) continual growth in video content and services(video conferencing, VoD, virtual reality sharing and gaming).

In 2012, web browsing accounted for 30% of all web traffic and is expected to increase 50% by 2014 [24]. By 2018, web browsing will constitute 10% of the total mobile data traffic [59]. Another equivalent contributor is social networking, which will account for 9% of mobile Internet traffic by 2018 [59]. Social networking is a collection of segmented information generated spontaneously. It is more natural for users to update their social network statues via mobile devices. Moreover, social networking has become a primary channel of advertising, business campaign, and integration of online gaming. Meanwhile, its non-social networking functionalities are also boosting the volume of traffic, such as social network authentications and search.

Even though video traffic dominates the share of mobile Internet traffic, M2M has the potential to lead the traffic volume, considering the amount of mobile traffic from various scenarios, especially from bandwidth-intensive application, such as real-time information monitoring. It is predicted that there will be 225 million cellular M2M devices, resulting in significant mobile traffic by 2014 [60] and presenting 5% of global mobile data traffic by 2016 [24].

2.2 Challenges in the Mobile Internet Evolution

The booming of mobile traffic can be foreseen while wireless network infrastructure, mobile devices and various applications are advancing. Meanwhile, the insatiable demand for mobile data worldwide is creating challenges and pains for operators and content providers. Moreover, the ever-increasing demand for mobile devices and skyrocketing amount of mobile Internet traffic can also lead to issues for end users and the whole society at large. As this dissertation focuses on energy efficiency and energy savings, the following sections will describe energy-efficiency-related challenges, mainly focusing on CO_2 emission, electricity consumption and the impact on QoE.

2.2.1 Rising CO_2 Footprint and Energy Consumption of Mobile Internet

ICT has been one of the fastest growing sectors of the economy, and is expected to continue to grow at a rapid rate in coming years, but at the price of increased carbon footprint. The ICT footprint implies the environmental impact created by all individual ICT devices and networks. In 2007, the ICT sector was accountable for 1.3% of worldwide CO_2 emissions, which equals 620 Mt of CO_2. The study [61] found that GreenHouse Gas (GHG) generated per average user has decreased from about 300 kg CO_2e in 1995 to about 100 kg in 2007, and is estimated to drop further to 80 kg by 2020 due to improving energy efficiency of ICT equipment. Meanwhile, the carbon footprint per gigabyte also shows a decline from about 75 kg/GB 1995 down to about 7 kg/GB in 2007 [62]. However, the estimated total CO_2 will increase to 1.9%, giving about 1100 Mt of CO_2 by 2020 [63, 64]. This is mainly because the number of Internet-connected devices is foreseen to be more than doubled in 2020. Strong evidence shows that climate change is happening, and the GHG emission is identified as the root of this change and most air pollution. In order to achieve a low CO_2 and sustainable economy, the EU is committed to taking urgent action by reducing GHG emission to a manageable level that would limit the global temperature increase to $2\,^{\circ}\mathrm{C}$ compared to pre-industrial levels [65].

The increasing GHG emissions are produced from the fossil-fuel-generated electricity that is used to power all the ICT devices and networks. The ICT energy consumption is becoming a significant portion of the energy consumption worldwide, and this portion is expected to grow dramatically over the coming years. The current estimation is that the ICT sector consumes around $6\sim10\%$ of the world's energy [66] and is foreseen to increase by 60% from 2007 to 2020 [61]. All ICT devices, networks and services are dependent on electricity to function. Oil and gas prices have doubled over the past three years, with electricity prices following [67]. The increasing electricity cost has already been a huge presence of ICT operating expense.

Typically, the ICT footprint refers to the environmental impact created by wireless and fixed telecommunication networks, data centres, and all equipment connected to the networks including mobile phones, tablets and PCs. Currently, the most significant ICT footprint may be accounted to PCs and data centres [62, 61]. The GHG emissions and energy consumption per PC were dropped due to the change from cathode ray tube screens to flat panels and from desktops to laptops. The presence of PCs is expected to decrease in the future, considering the fast growth of mobile Internet and the emerging cloud computing. Cloud computing provides software, platform and infrastructure as a service with elasticity, reliability and constant availability, requiring running servers, cooling systems, power supplies and voltage converters, which are all powered by electricity. The consumption has introduced high electricity costs and GHG emissions. In 2007, the global data centre footprint was around 90 Mt CO_2 and is expected to grow to 259 Mt CO_2 by 2020, making data centres the fastest growing [62, 68]. Fixed-line networks, including local area networks and data transport networks, contribute around 15% of the total GHG emissions of ICT, and the rate is not expected to see a high increase in the future [62].

In addition to the environmental and economic cost of data centres, fixed networks and PCs, there is a a strong incentive to reduce energy consumption of mobile communications given the rising number of mobile devices and network infrastructures. The GHG emissions of mobile networks, including wireless access points, is expected to be 235 Mt CO_2 by 2020, where the footprint of Radio Access Network (RAN) dominates the overall GHG footprint. The average RAN electricity consumption per subscription was about 17 kWh and decreases about 8% every year. The amount of energy consumption is predicted to be 88 TWh/year by 2020. Compared to the RAN, the power consumption of a femto cell is around $8\sim10$ W. It can be assumed that the power consumption would drop to 5 W by 2020, and total energy consumption will be less than 5% of that consumed by the global RAN [69]. With the emerging convergence between cloud computing and wireless communications, an increasing number of users access cloud services from anywhere, at anytime wirelessly. A study [70] indicates that wireless access technologies, such as WLAN and LTE, will be the dominant methods for accessing cloud services instead of wired connections, and the density of wireless base stations will increase by 1000 times to meet the demand of huge mobile traffic volume. Therefore, the total energy consumption of cloud services accessed via wireless networks (wireless cloud energy consumption) could reach between 32 TWh and 43 TWh by 2015, where wireless communications, including mobile communications and WLAN technologies, would account for 90% of total wireless cloud energy consumption, while data centres account for only around 9%. One study [69] estimated that a mobile device generates 18 kg CO_2 for manufacturing and 2 kWh/year for operating on average. Although the energy consumption and footprints of mobile devices are relatively small, it is still essential to keep the energy consumption as low as possible since users require connection with cloud services all the time via mobile devices, which are always power starving due to the performance limitation of batteries.

2.2.2 Quality of User Experience

The fast growth of mobile Internet services and mobile data traffic is not only at the cost of GHG footprints and energy consumption, but also presents a significant barrier to continued adoption of mobile Internet services and sustainability of an acceptable QoE for mobile Internet. With the transition from wired networks to wireless networks, mobile devices are treated as a gateway to one's daily life, providing not only entertainment but also access to work. However, the ubiquitousness and mobility is compromised by limited battery life of mobile devices.

There have been various studies on Human-Battery Interaction (HBI) to investigate how mobile phone users behave with limited battery lifetime by conducting user behaviour surveys and tracking their mobile phone statues, such as charging activity and battery level. A typical scenario almost all users ran into is that mobile users feel disturbed when mobile devices were running out of battery, and more disturbed when the devices turn off and users lose important phone calls due to unpredictable battery life. According to the study [71], most mobile users consistently charged their devices during the day and again overnight, and were not satisfied with the longevity of their devices's battery. Similar frustration can be found from another study [72] as well. Nearly one-fifth of the users experienced a dead battery at least once a week, and about half of them reported it one or more times per month. The study also shows that 63% of mobile users reported low-battery warning at least 1∼2 times per week with 18% of those seeing one between 3∼9 times per week.

There are several reasons of causing such degraded QoE, some of which are elaborated as follows.

- **Battery Technology**: According to the HBI studies, mobile users have limited understanding and little indication about how to manage battery life and energy-consuming applications. This can be improved by providing fine-grained information and Interactive Battery Interface (IBI) to effectively deal with the limited battery lifetime. However, a non-neglectable fact is that current battery technologies are topping out in capacity, while demands of mobile devices for capabilities and performance are driving higher power consumption.

  The state-of-the-art integrated circuits doubles processing every two years, more or less following Moore's Law. However, the law does not apply to battery technologies due to some challenges, one of which is today's lithium-ion batteries have limitations in storing large amounts of energy with reasonable size and weight. More specifically, each battery has a graphite electrode and a metal oxide electric. The charge stored by the battery is released when lithium-ions move from one electrode to the other. However, the graphite anode that the battery generally uses has to be fairly large to store enough power. Thus, so long as batteries continue to be based on electro-chemical processes, limitations of power density will be difficult to overcome.

  Great efforts of improving battery capacity continue. Recently, a team of the University of Maryland replaced the graphite anodes with silicon and grew beads of silicon on a Carbon NanoTube (CNT). New chemical processes have been developed to create a resilient structure for silicon to be charged with lithium-ions [73]. The breakthrough may lead to vastly improved power density and more charge/discharge cycles than it does today. There are also others working on finding less bulky replacement material. However, there is still much to do until the technologies can be applied to commercial mobile phone batteries.

Due to the lagging behind of battery technology, research and development has been focusing on energy conservation and saving techniques on mobile devices.

- **Various Radio Interfaces**: A common situation found in HBI studies is that the battery lasts no more than a few hours when a mobile device is continuously transmitting data, such as watching video streaming, using a mobile device as a modem, or downloading large files. This is a simple reflection of the fact that radio chipsets are the most power-consuming components and needed in various occasions to transmit bits in mobile devices nowadays.

  A race is already happening among mobile devices manufactories, who have realised that just offering voice, SMS and a colour display nowadays is far from enough. Products have to seamlessly enable support for multiple radio interfaces for providing "always-on" Internet connectivity and higher data rates via either 2G, 3G, 4G or WLAN. Due to requirements of high data rates, the complexity of radio interfaces doubles every 2.5 years. The Very Large Scale Integration (VSLI) horsepower grows from 0.1 Giga-Instructions per second Giga-Instructions per second (GIPS) for GSM, to 2 GIPS for UMTS, and beyond 10 GIPS for LTE [74]. Moreover, the products also need high computational power and storage to keep pace with this trend. Last but not least, a number of sensors and short-range radios are equipped to provide cutting-edge services, using GPS to develop location-aware applications, accelerometer for motion tracking, Near Field Communication (NFC) for mobile payment, and Bluetooth for short-range and energy-efficient transmission and connecting to other hardware within the range. Various radio interfaces and sensors increase the feature-set of a mobile device. However, as a consequence, their processing power increases power constrains, which is bottlenecked by limited battery life.

  Interesting research has been done in the area. A quantitative study [75] presents a trace-driven simulation on the performance of 3G mobile data offloading to WiFi networks, indicating that WiFi offloaded about 65% of the total mobile data traffic and saved 55% of battery power by the time the study was conducted in 2012. The delay of data transfer can further achieve higher energy saving. Another study [76] presents a system called Wiffler, which brings two key ideas: leveraging delay tolerance and fast switching to reduce 3G usage of moving vehicles in cities. The

Wiffler predicts WiFi connectivity based on the average throughput offered by an AP and the number of APs that will be encountered until a given future time interval. Then, the prediction results instruct the system when to delay transfer and offload data from 3G networks to WiFi networks. By combining different networks, the total cost of 3G data transfer can be reduced by almost half for a delay tolerance of 1 minute.

- **Non-optimised Mobile Applications and Services**: Nowadays, the SDKs provided by vendors like Apple and Google give an easy entry for software developers to make mobile applications. However, on one hand, limited power consumption information exposed by the operating systems and non-optimised system-level power management set up obstacles for developers to address energy consumption issues in the first place; on the another hand, many developers have limited experience with energy-constrained mobile operating systems, which leads to unintentional and unfortunate power-hungry software design decisions. Thus, power consumption information, together with processing power, display size and input capability, should be considered as one of the most important limitations in developing applications and services for mobile devices [77]. For example, heartbeat messages are often used by mobile applications and service backends to maintain connections between each other and update their status. Intuitively, the more frequently the heartbeats are sent, the better synchronisation of services is. However, frequent heartbeats are one of the causes of the limited battery life, since the data transmission keeps radio interfaces always active. For iOS devices, background applications do not generate heartbeat messages when the screen is switched off. Due to the lack of a unified heartbeat mechanism in system-level of Android devices, the number of connections is 15 times that of iOS devices when a mobile device is in connected status [78]. Besides, heartbeat messages, together with a fast dormancy feature of cellular networks, also increase access request and paging signalling in the networks. Two studies [79, 80] give deep insights that always-on type applications can lead to unacceptable short battery lifetimes as well as massive signalling in 3G and 4G networks.

- **Wireless Network**: One unavoidable issue effecting QoE is wireless network latency. In wired networks, network latency is much lower, and QoE can be ensured by traffic engineering and over-provisioning to minimise latency and avoid network congestion. However, wireless Internet connections are different from the wired counterparts. Any number of physical or electromagnetic barriers can introduce interference to wireless signal and adversely impact the effective bandwidth. The mobility of mobile users can even worsen the quality. Publication IX presents solid measurement results showing wireless data transmission experiences high and variable latency, and TCP throughput is fluctuated. Latency inflation could lead to high retransmission, potential TCP SYN timeout, high recovery time and packet losses. The high latency can severely affect QoE of services such as visiting a website easily. Since HTTP message exchange is based on request and reply between a mobile device and web server, including DNS (Domain Name System) lookup messages, a wireless link creates various latency for all of these back and forth transmissions, and dramatically increase page downloading time.

Poor connectivity and signal coverage are always obstacles for mobile data transfer. One way [81] to tackle the problem is to combine multiple network interfaces on mobile devices. The solution uses Open vSwitch to stitch multiple networks together at the same time for higher throughput, minimised loss, delay and power consumption without reestablishing TCP state due to handovers. Moreover, the study done by Ding et al. [82] quantifies the power consumption on data transfer induced by poor wireless signal strength, and introduces a system-call-driven power model to incorporate the signal strength factor. The results show that delaying background traffic can reduce the total energy consumption of data communication by up to 23.7% and 21.5% under WiFi and 3G respectively, with a maximum delay of 12 hours. In another study done by Ra et al. [83], an optimal online algorithm was presented for energy-delay tradeoff using the Lyapunov optimisation framework, that can achieve near-optimal power consumption by automatically adapting to three factors, namely wireless channel conditions, transmission energy and the volume of backlogged data, to decide whether and when to defer a data transmission.

2.3 Summary

With the fast development pace of mobile technologies, the opportunities and problems are both rising as discussed in this chapter. To understand the impact of mobile Internet booming can give a perspective on the accomplishments of technologies and the challenges we are facing towards the future. The growth in power requirements and levels of CO_2 emissions render the current state unsustainable. The ICT sector has been regarded as a negative environmental impact. But it can also make positive impacts by helping other sectors to reduce the environmental impacts via improving production efficiency, intelligent process control, such as e-health, e-learning and e-banking, and favouring renewables and low-carbon conversion technologies for electricity, heating and cooling, and so on [84]. ICT-enabled solutions could reduce global CO_2 emissions by 16.5% by 2020 [61].

The fast growth can also become clear expenditures for telecom operators and a cause of QoE degradation for mobile users in terms of battery life. Thus, integrating the fast change and energy consumption will therefore ensure that they are mutually reinforcing to reduce overall environmental presence and increase QoE. This thesis focuses on providing insights and solutions based on measurements, modelling and optimisation of mobile data transmission to reduce data and signalling transmitted over wireless links. These can not only help in saving energy on mobile devices but also in decreasing the energy consumption in wireless access networks. The detailed approaches are elaborated in the following chapters.

3. Understanding the Power Consumption of Mobile Data Transmission

Due to the growth of mobile Internet and the increase in traffic, it is important to understand the characteristics of power consumption of mobile data transmission in order to find opportunities to balance the energy consumption and the growth of mobile users and the data volumes, which are covered in the results of Publications I, II, III, IV, V and VI. This chapter first introduces techniques and methodologies of measuring power consumption of mobile devices in Section 3.1. Then, the power consumption characteristics of radio interfaces are illustrated in Section 3.2. Last, Section 3.3 presents the potential of using data compression in a mobile environment to save energy for mobile devices.

3.1 Power Consumption Measurement

As mobile devices use power, the power consumption must be made an integral part of product design and testing. Thus, the request of a better understanding of the power consumption characteristics is placing high demands on power consumption measurements to provide essential knowledge for optimising, evaluating and validating. Power consumption is defined as the amount of energy per unit of time, and the basic unit of power is watt (W), while the joule (J) is a derived unit of energy. Instead of analysing pulse or peak power, this dissertation focuses on average power consumption, which is the average value of the accumulated product of instantaneous voltage and current integrated over a specific time period of measurement. The battery life we refer to is the longevity of a mobile device running on a single charge of a battery power source.

It is important that power measurements can provide accurate results, and be repeated at different times and at different places to cater different measurement scenarios. This leads to various measurement techniques

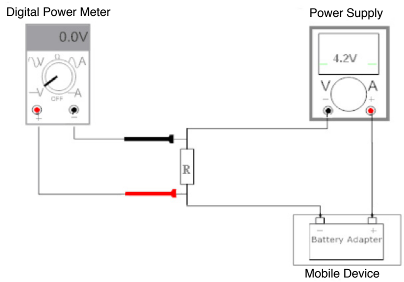


Figure 3.1. Measurement logic

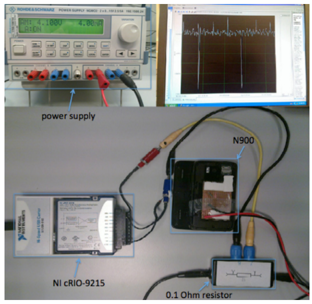


Figure 3.2. Power measurement setup

for mobile devices, which are compared as follows.

3.1.1 Hardware-Based Measurement

Highly accurate measurement results require well-behaved equipment and measurement techniques. Direct measurement of mobile devices with external digital multimeters assures significantly improved accessibility of fine-grained energy consumption information. A commercial power meter features high measurement accuracy and high sampling rates. The typical measurement setup of power consumption measurement includes a digital multimeter connected to a mobile device for current or voltage sampling, and a PC running with special software to collect, store and analyse the samples. This approach is widely adopted in existing studies [85, 10, 86]. An energy consumption monitoring framework was proposed in the study done by Keranidis et al. [87], which was built on a distributed network of low-cost, but accurate devices with full integration with large-scale wireless testbed. The framework can characterise the power consumption of realistic wireless experiments, and monitor experiment execution.

To make sure the measurement results are systematic when repeating measurements, and uninterrupted when making a measurement for a long period, it is necessary that the power source of the mobile device remains stable. In this thesis work, the batteries of examined mobile devices were replaced by battery adaptors, which connected to an external stead power supply. A high-speed sampling data acquisition device NI cRIO-9215[1] was then used to collect voltage fluctuations with a rate of 1000 samples per second across a 0.1 Ohm resistor. With a known resistance and measured voltage drop, the current can be determined by Ohm's

[1]NI cRIO-9215, http://sine.ni.com/nips/cds/view/p/lang/sv/nid/208793

law. The voltage samples were sent via NI USB-9162[2] to a PC running NI-DAQmx software[3] to analyse recorded data and calculate real-time power consumption. Figure 3.1 and Figure 3.2 illustrate the logic of the setup and its real implementation. This methodology is the main approach to conduct accurate power consumption measurements in this dissertation, which has been applied in Publications I, II, III, V, VI, VIII, IV and VI.

External hardware-based measurement can provide high accurate results. However, it also has a clear drawback for measuring power consumption of mobile devices. Regardless, the cost of the hardware, external hardware limits a phone's mobility, restricting real-world measurement and mobile scenarios. A counter-solution is to use accurate battery sensors that provide accurate readings of battery voltage and an instantaneous current from the mobile OS. Nokia Energy Profiler (NEP) [4] is a very typical example of this. It is an application with built-in power profiling running on Nokia's Symbian and Meego phones, allowing power consumption measurements without external hardware. Besides power readings, NEP also provides temperature, signal strength, CPU, memory and networking usage. Compared to hardware-based solutions, NEP only has a maximum sampling rate at 4 samples per second. Nevertheless, NEP is proven to be a reliable power consumption measurement technique by many studies [10, 77, 11], showing that the accuracy is accurate enough to replace external hardware as the source of power measurements. This approach was also used in Publication IX.

3.1.2 Component Level Measurement

The measurement methodologies just discussed give the overall power consumption at system level. Inside a mobile phone system, each component, such as CPU, memory, display, radio interfaces and various application-specific accelerators, contributes to the overall consumption. It is also important to identify and deeply study the most power-consuming components on a battery-powered and resource-limited mobile device by breaking down the system into major subsystems. In one study [85], a special mobile phone, Openmoko Neo Freerunner[5], provides free circuit schematics and enables the researchers to produce a breakdown of power distribu-

[2] NI USB-9162, http://sine.ni.com/nips/cds/view/p/lang/sv/nid/204178
[3] NI-DAQmx software, http://www.ni.com/dataacquisition/nidaqmx.htm
[4] NEP, http://store.ovi.com/content/73969
[5] Freerunner, http://wiki.openmoko.org/wiki/Main_Page

tion to various components. However, this approach is not widely adaptable for other commercial device models. Instead of acquiring power consumption from a special mobile device, another way to get coarse-grained estimation of the component is switching off all other elements that might consume energy, yet are not vital to keep the OS running when executing the workload of the component. The estimated power consumption can then be calculated by subtracting the power consumption of the mobile device in an idle state from the measured result. There are many studies [77, 88, 89], including the work of this thesis, that use this approach to breakdown the power consumption of a mobile device.

3.1.3 Power Consumption Modelling

The inconvenience, cost, and complexity of external power measurement hardware or special requirement of on-board battery sensors limit measurement cases and scenarios for mobile devices. Thus, many research efforts are dedicated to creating power modelling tools. Another reason for modelling research is to build power models for applications and certain types of network traffic in order to design power provisioning and energy savings based on the models.

There is a wealth of research studies on power models in existing literature. One kind of research focuses on power modelling based on determining the Finite State Machine(FSM) of a mobile device. The approach breaks down a mobile device into subsystems described by FSM states and creates a model that maps a fixed power consumption value to each state. The power consumption of a subsystem can be formulated by regression model as a function of residence time of states and the power cost associated with each state. The study [90] by Pathak et al. is one of the examples that proposes a system utilisation-power-state correlation. It collects utilisation statistics of individual components via OS to build a linear-regression model to correlate the sampled values. Once the model is constructed, it uses system calls to determine the power state of each component. Its extended work [90] presents an energy profiler for mobile devices named Eprof. In a study [77] by Zhang et al., PowerTutor was proposed to provide real-time power estimations for mobile devices, whose core engine is a power model named PowerBooter. It uses a set of training programs to determine the relationship between each power state and power consumption for each relevant hardware(CPU, LCD, GPS, Wi-Fi and cellular interfaces). Other studies [91, 89] apply the similar approach

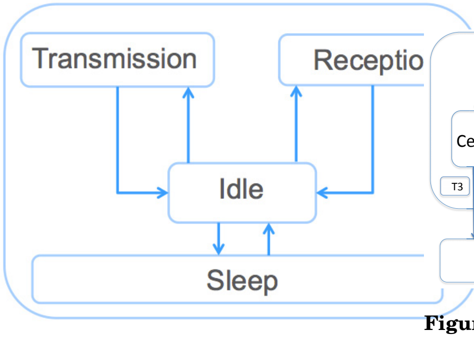


Figure 3.3. Power consumption states of WLAN interface

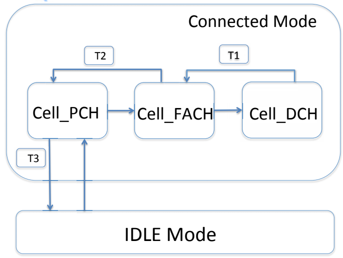


Figure 3.4. Power consumption states of 3G interface

but focus on power modelling of individual applications by estimating the state residence time from the behaviour of applications.

Other approaches focus on building statistical power models. Sesame [92] presents a statistical power model that uses the battery interface to build an adaptive and self-learning model. It is based on model moulding and has predictor transformation to improve accuracy. Carat [93] uses a rather different approach than the existing studies by collecting instrumentation data from mobile devices and sending them to a Carat server, where Carat builds power models and diagnoses anomalies. By comparing application behaviour with the same application running on other mobile devices, the system can detect anomalies, and quantify error and confidence bounds, then provide recommended actions to improve battery life.

Still, it is challenging for power modelling to provide precise readings for many reasons: 1) readings provided by hardware and software performance counters may not be accurate; 2) accuracy is limited by training environment and modelling of each component; 3) particularly, for machine learning-based models, model correction is needed for power anomalies. In summary, each measurement methodology has its advantages and drawbacks, and measurement cases define methodology selection.

3.2 Power Consumption Characteristics of Radio Interfaces

This section starts with power consumption states of WLAN and 3G radio interfaces on mobile devices. Then it elaborates the characteristics of these radio interfaces with the results in Publications I, IV, VI, VI and IX.

3.2.1 Power Consumption States

Most existing WLAN power-saving mechanisms are based on deactivating WLAN NIC in periods of no data transmission. IEEE 802.11 standards [94, 95] define that 802.11 WLAN-capable devices operate either in Continuously Active Mode (CAM) or Power Saving Mode (PSM). The PSM was designed to improve the power saving of a WLAN interface by switching the interface from Active state to the Sleep state as soon as data transmission is completed. To be precise, a WLAN interface can operate in four states, namely Transmission, Reception, Idle or Sleep states, as shown in Figure 3.3, each of which presents different power consumption. The Idle state means that the interface is powered and ready to transmit or receive data, consuming significant amount of power. When the WLAN interface starts to send or receive data, it enters the Transmission or Reception states, which are together known as the Active state and present the highest amount of energy consumption. A mobile device synchronises with an infrastructure, such as Access Point (AP). If there is no traffic, the interface stays in the lowest power consumption state, namely the Sleep state, and only wakes up every beacon interval for a beacon to decide to wake up or not depending on whether the frame contains a Traffic Indication Map (TIM) message, which indicates that the interface buffered data frame at the AP, is ready to receive. Thus, keeping the interface in the Sleep state as much as possible is the goal of many techniques, such as traffic shaping, ON/OFF switching of WLAN interface and processor, packet pacing, and MAC-level download scheduling by access points and so on.

3G networks have more sophistic resource management, which uses an RRC state machine [96] to control 3G interfaces. There are several states: IDLE state, Cell Paging Channel (Cell_PCH) state, Cell Forward Access Channel (Cell_FACH) state and Cell Dedicated Channel (Cell_DCH) state. The IDLE state enables 3G interface to only receive paging messages from the Radio Network Controller (RNC) and is the lowest power consumption state. In the Cell_PCH state, the interface monitors the paging control channel, and is still not able to have uplink activity. Packet Data Protocol (PDP) context is maintained so a session could be reconnected rapidly. Since there is no real data traffic transmitted in the Cell_PCH state, the power consumption of the state is also low. The Cell_FACH state allows low data rate transmission via a common or shared transport

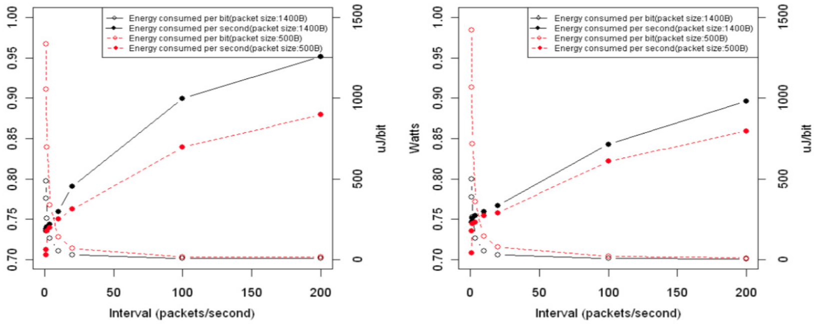


Figure 3.5. Consumed energy on packets with different transmission intervals in an HSPA network

channel. In the Cell_DCH state, the RRC connection is fully established, and dedicated transport channels are assigned to downlink and uplink for full-speed transmission. Due to the dedicated resources and high data rate traffic, this state presents the highest power consumption. Compared to this state, the Cell_FACH state consumes roughly 50% of that in the Cell_DCH state, and the Cell_PCH state only consumes about $1{\sim}2\%$ of the operating power of the Cell_DCH state. As shown in Figure 3.4, the states promote when switching from lower power consumption states to higher power consumption states, and the states demote when switching happens in the reverse direction. The state promotion from the Cell_IDLE or Cell_PCH state to Cell_FACH state is triggered by transmission activity(T1, T2 and T3). The promotion to Cell_DCH state happens when the data volume exceeds the Radio Link Control (RLC) buffer threshold. The state promotion normally only takes $1{\sim}2$ seconds [97]. The demotion is triggered by in-activity timers or controlled directly by Fast Dormancy, which is a feature in 3GPP specifications for a mobile device to demote to IDLE state by sending an RRC control message to the RNC [98].

3.2.2 Power Consumption Characteristics

To gain the understanding of the power consumption characteristics of radio interfaces, especially the most power-consuming ones, namely WLAN and cellular interfaces, thorough measurements have been conducted in Publication I. The publication used the measurement methodology described in Section 3.1.1 to analyse the power consumption of wireless data transfer over EDGE, HSPA and WLAN. Instead of analysing a particular application, this study focuses on packet transmission patterns to provide insights for power consumption modelling and answers the questions regarding how much energy a certain service consumes on a mobile device caused by communications. The study analyses the impacts of different packet sizes, packet-sending intervals on the power consumption, and presents the dissipation results. An example result is shown in Figure 3.5, where the power consumption and consumed energy changes dramatically with the increase of sending intervals. Furthermore, it compares the power consumption difference between IEEE 802.11b and IEEE 802.11g, as well as investigates how different data service packages and the operator's network affect the power consumption of cellular interface. All the measurements are quantified by their power consumption and energy consumption per bit when the radio interfaces send or receive traffic.

The results suggest that it is important to transfer data at full capacity of radio links, since the fixed overhead of transmission is significant when the radio interfaces are in a communication state. So the packet size and sending interval should be set as high as possible to minimise the transmission time and per-bit energy consumption. When designing Internet services or programming mobile applications, data should be sent in burst to extent the residence time of radio interfaces in a low power consumption states.

By extending Publication I, we analysed the power consumption in the case of uplink transmission in 3G networks and showed how the power consumption is determined by different transmission parameters. This research is presented in Publication V and it continued in Publication VI. The studies break down the power consumption of 3G interfaces into each RRC state and deeply analyse the influence of packet-sending intervals and packet size on power consumption. The size of the transport block determines the maximum payload that can be transmitted once every Transmission Time Interval (TTI), and TTI determines the maximum packet sending or receiving rate. These two parameters together influence the maximum throughput and packet sending or receiving pattern in the Physical layer. A packet-sending or -receiving interval of application directly affects the transiting interval in the physical layer, and the size of packet determines whether packet segmentation happens or not. Thus, the power consumption of a radio interface increases proportionally to the number of transport block sets sent and received over one radio interface. In the paper, we proposed the following power consumption model for UE to send or receive packets in state Cell_DCH. As shown in the equation 3.1, the power consumption consists of three main power contributors. The power consumption of maintaining Cell_DCH state is defined as P_{DCH} in watt and considered to be an approximately constant value. The power consumption of sending or receiving packets is defined as P_{peak} in watt. Also, the power consumption for encapsulation or decapsulation $P_{enc}(s)$ for packet size s is treated as the incremental power that is proportional to the size of the packet.

$$P = P_{DCH} + P_{peak} + P_{enc}(s). \quad (3.1)$$

Meanwhile, we define the number of transport blocks needed for sending one IP packet as

$$N = \left\lceil \frac{s}{MTBS} \right\rceil, \quad (3.2)$$

where MTBS is Maximum Transport Block Size.

When more than one transport block is needed for sending or receiving one IP packet, the time spent on processing this packet is $N \cdot \tau$, where τ is defined as the value of TTI. Normally, a packet-sending interval I is much larger than the packet processing time. Thus,

$$P_{peak} = \frac{N}{I} \cdot E_{peak}, \text{ when } I > N \cdot \tau. \quad (3.3)$$

Where E_{peak} is defined as energy consumption of sending or receiving one peak in Joule.

Then taking into account (3.2) and (3.3), power consumption in the Cell_DCH state can be written as

$$P = P_{DCH} + \frac{E_{peak}}{I} \left(\left\lceil \frac{s}{MTBS} \right\rceil \right) + P_{enc}(s) \quad (3.4)$$

A more-detailed detailed explanation can be found in Publication V and Publication VI. Moreover, the model was validated against real measurement and a reference model. Furthermore, the publications analyse the RRC transition state machine for the uplink power consumption. According to the state machine, an RRC parameters selection algorithm was proposed to optimise power saving for data transmission. The selection algorithm also considers the constrains regarding amount of signalling traffic, RLC buffer size, and buffering latency.

Furthermore, Publication IX investigates transmission issues over wireless networks and its impact on energy consumption. In the beginning of the study, TCP performance issues are presented based on a thorough mobile measurements from the Nettitutka platform[6]. Due to severe error rates caused by external radio interferences, going out-of-range, or

[6]Nettitutka, http://www.nettitutka.fi

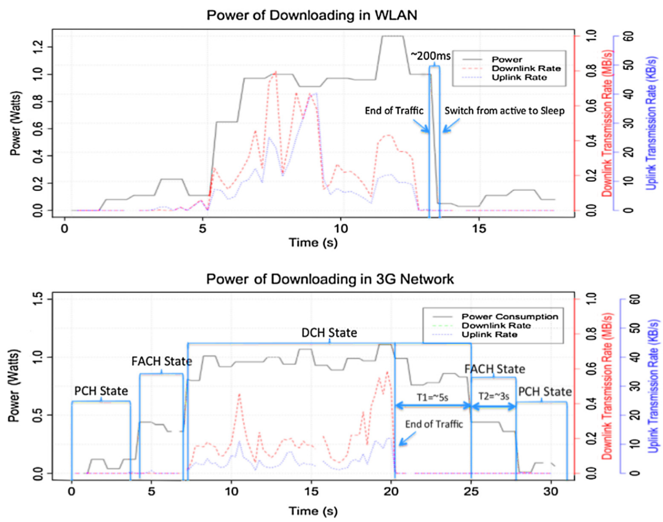


Figure 3.6. Power consumption in different power consumption states in WLAN and 3G networks

blocking of signal, and large delay of wireless networks, TCP suffers from significant throughput degradation, link capacity underutilisation, and excessive interruption of data transmissions. All the issues have a negative impact on power consumption of mobile devices. The work presents the impact of these issues on HTTP traffic. Moreover, the work quantifies the power consumption of the RRC state of 3G networks and compares it with the power consumption characteristic of WLAN, as shown in Figure 3.6. The 3G link exhibits significant residual energy consumption due to the inactivity timers. In order to help further work in making design decisions, the work also describes how to identify the values of each inactivity timers and the RLC buffer threshold that determines the amount of data triggering the RRC state transition.

Since WLAN and 3G interfaces are typically the most power-hungry components for a mobile device, due to the high power consumption overhead, it is worth looking into low-energy radio technologies to provide the best solution for different scenarios. In Publication IV, an open architecture platform for using passive RFID tags in close proximity environment is proposed. Maximising throughput and minimising power consumption are critical requirements for these kinds of use cases. Thus, the work looks into several radio technology alternatives and compares 3G and WLAN with Bluetooth, NFC and UWBLEE technologies. Example results are shown in Figure 3.7, where the time spent on downloading a 50 MB movie trailer is only 8 seconds, and the energy consumption of

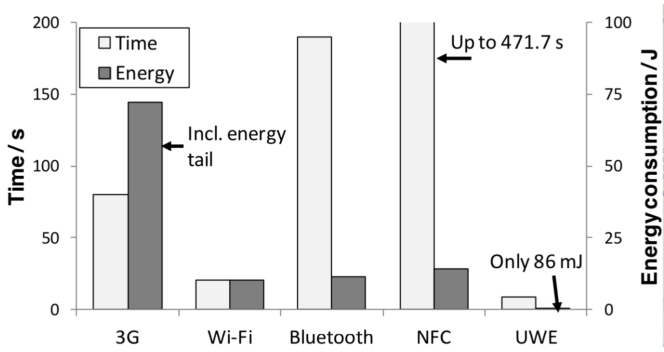


Figure 3.7. Time and energy consumed of using different radio technologies

RF front-end is 0.043 J when using UWBLEE, while all the other radio technologies are more time and energy consuming.

3.3 Energy Trade-off between Computation and Communication

Data compression is a solution to decrease communication costs in terms of the number of bits transmitted. Compression algorithms can be divided into two categories, namely, lossy and lossless compression. Since lossy compression introduces differences to reconstructed data in exchange for a better compression ratio, the compression algorithms investigated in this dissertation fall into the category of lossless compression. Lossless compression is widely applied in the Transport layer to minimise the amount of data and reduce the transmission time. For example, packet header compression has been used to improve the throughput over weaker wireless links, such as TCP/IP, UDP/IP header compression and HTTP compression.

The power consumption of transmitting data over wireless links is expensive, as shown in the study [99] by Barr et al, where the consumption of sending one bit over the air is over 1000 times than that of 32-bit CPU computation. To tackle the issue, one research direction [14, 13] is to use data compression for energy-efficient communications. However, compression schemes involve tradeoff due to the intensive computation and memory access to compress and decompress data. The consequence might be that more energy is consumed than when simply transmitting the raw data. Furthermore, the transmission rate in wireless networks may give different results in energy consumption and affect the decision on whether to deploy compression schemes or not. As power consump-

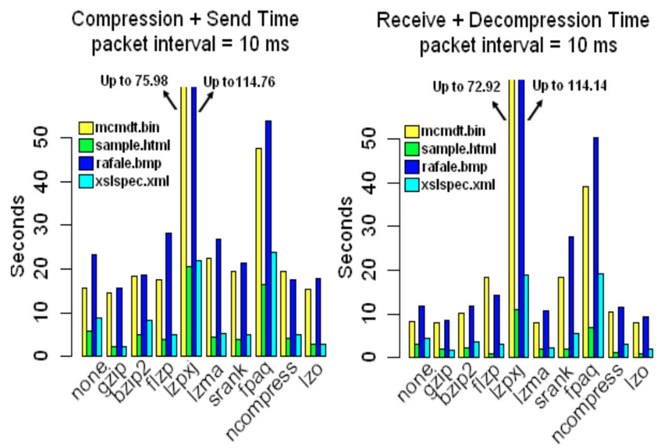


Figure 3.8. Time required to compress and send BIN, HTML, BMP, and XML files

tion on mobile devices differs from each other, due to hardware and software related factors, an evaluation of compression algorithms over modern hand-held devices provides a timely understanding of data transmission and compression in the perspective of energy efficiency for more energy savings.

In Publication II, energy-efficient ways to utilise compression have been re-evaluated to answer two key questions: 1) what data should be compressed and how, and 2) what are the limitations and restrictions when optimising communication and compression together. The study evaluates nine compression schemes that are the representatives of widely used compression algorithms, such as statistical compression, dictionary compression and predictive compression. We examined a set of the most common file types in the Internet, divided into three categories: hard-to-compress files (e.g. JPG, MP3, EXE and WMA files), compressible files (e.g. PDF, SWF files) and easy-to-compress files (e.g. BIN, HTML, BMP, and XML files). Figure 3.8 shows an example of comparison results of the easy-to-compress files. As shown, most of the compression schemas provide energy-efficient transmission of the files. However, lzpxj and fpaq demand an extremely long time and consume a lot of energy to compress as well as decompress. Overall, gzip offers the best results for all the files from the energy-consumption perspective.

More sophisticated compression algorithms may take longer computation time to achieve smaller file sizes of certain files. The reduction of file size may shorten transmission time over the air, thus an overall reduction of transmission time (including the time spent on compression, de-

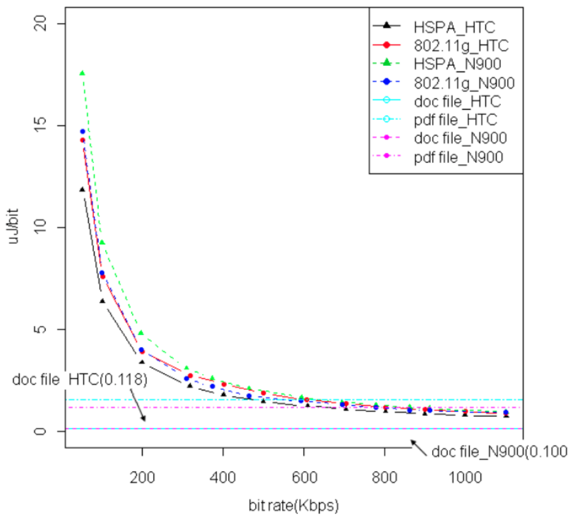


Figure 3.9. Compression energy conditions for HTC Hero and Nokia N900

compression and data transmission) is possible. However, all depend on the compression algorithm, file size, file type and link speed. Therefore, the study evaluates the power consumption of compressing and decompressing each file type with different compression programs, and looks into the trade-off between computation and communication regarding the above-mentioned aspects. In order to show the benefits in energy savings achieved by using compression with asymmetric patterns, the study also evaluates a series of webpages.

In summary, the contributions of this publication are the analysis of a wide number of compression schemes on many types of web content on a modern mobile device, identifying the trade-offs when using compression for energy-efficient data transmission and the discussion of the deployment issues related to enabling data compression on the Internet and for the mobile users.

Publication III extends the study Publication II and takes boundaries, such as bandwidth and hardware, into consideration when utilising compression for the energy-efficient data transmission. Practically, there is a maximum bit rate of communication due to the limitations of processing, radio communications technologies, and conditions of wireless links. Also, compression has a maximum information bit rate due to the nature of compression algorithms and capacity of hardware.This publication formulates the condition when to transmit compressed data instead of just sending plain data for energy savings. It proposes partial compression to increase energy efficiency when using data compression but with the limitations of compression, and communications. Instead of fully compressing all data, only part of the data flow is compressed and the rest is communicated uncompressed. To elaborate the proposal, the publication formulates the power consumption and energy efficiency of transiting compressed data, and the compression conditions in a mobile environment under 1) communication limit and 2) when adding new data flows to existing communications.The compression conditions have been verified through experiments and measurements on the Nokia N900 and HTC Hero in both cellular and WLAN networks. The results verified the linear approximations of the compression conditions and showed the condition for adding a new data flow. Figure 3.9 shows one of the results from the publication, where the conditions of applying compression is illustrated. As shown, the thresholds of compressing .pdf and .doc files or not on the Nokia N900 and HTC Hero are drawn in horizontal lines, indicating that it is worthwhile to compress the .doc file for both devices at all measured bit rates in either a WLAN or HSPA network. As for the .pdf file, it is not worth compressing if the bit rate is over 500 kbps and 900 kbps for the Hero and N900 respectively in the HSPA network. Similarly, it does not bring energy savings if the bit rate is over 600 kbps for the Hero and 800 kbps for the N900 in the WLAN network.

As previously said, when the quality of the radio link goes down, even small savings in file size can lead to substantial energy savings, since energy consumption per bit becomes increasingly significant. However, energy saving through data compression needs to fulfil certain conditions, which includes considerations of link quality, computation load, file type, compression algorithms, compression and communication limits.

3.4 Summary

In order to provide effective methods and solutions for energy-efficient communication, it is fundamentally important to understand the power consumption characteristics of radio communications. This chapter started with the tools and techniques of measuring the power consumption of mobile devices. With the methodologies, accurate power consumption of a mobile device as a whole and the consumption break down become possible, enabling this work to analyse and model the power consumption of radio interfaces when transmitting data. Furthermore, the work dives deep into the RRC states in UMTS and provides a power model for the

RRC transition state machine for potential power saving. Moreover, thorough evaluation of using data compression for mobile data transmission is introduced, and the conditions for when to use data compression for energy-efficient mobile data transmission are formulated and discussed. With the tools and knowledge discovered in this chapter, energy-saving solutions are introduced in the following chapter.

4. Proxy-based Solution for Energy-efficient Mobile Web Access

This chapter introduces proxy-based solutions for energy-efficient mobile web access, utilising the results discovered in previous studies. It presents the results of Publications VII, VIII and IX. Firstly, Section 4.1 shows existing energy-saving solutions for mobile web access. As elaborated in Section 3.2 and Section 3.3, energy saving can be achieved by shaping traffic patterns according to power consumption characteristics of wireless networks, and compressing data adaptively. By taking the two discussed approaches into consideration, Section 4.2 presents an architecture of a proxy-based solution for energy-efficient mobile web access. Then the section elaborates the ways of implementation and shows the results for the proxy-based energy-efficient mobile web access.

4.1 Overview of Energy-efficient Mobile Web Access

As discussed in Section 2.1.3, web traffic delivery over mobile networks is rapidly growing. It is increasingly important to improve QoE end-to-end from web servers across the fixed Internet and the mobile networks to the mobile devices. To assure QoE and secure operators and web content providers' business, it is crucial to shorten page loading time as well as lower the power consumption of web access to enhance mobile users' satisfaction. Compared to desktop browsers, mobile browsers are limited by computational resources, power supply, unstable network connectivity and small screen size. The ways of enhancing QoE is to accelerate mobile web content delivery and reduce power consumption through one or a combination of the following common strategies.

- **Mobile Web Optimisation**: Since the majority of web content on the Internet are meant for PCs, one of the strategies for mobile web access is through content adaptations that reconstruct and tailor the web pages for mobile devices and mobile networks, with techniques such as removing the site header, advertisements, resizing or removing all images, customising the site with style changes and web page layout adaptation. The layout adaptation segments the web page based on its structure and regenerates a page for mobile browsing according to the hierarchy of the web elements [100, 101]. An alternative is to create a mobile version of a website so that the optimised web content can be more efficiently delivered to mobile users. For instance, .mobi [102] sites are optimised for mobile devices with special capabilities and restriction of screen size, input/output options, and so on, providing a top-level domain access and engaging mobile users with mobile compatible content and ubiquitous experiences.

Mobile web optimisation helps to reduce data volume of web traffic, thus on one hand, alleviating congestion for mobile networks; on the other hand, reducing downloading and rendering time, and power consumption for mobile devices. However, web content adaptation relies on simplified web elements and modified content, which may lead to reduction of QoE for mobile users. Furthermore, it forces content providers to maintain two versions of the same content.

- **Compression**: Webpage compression techniques reduce the data redundancies of web content. As defined in RFC 2616 [103], HTTP compression uses lossless compression to transmit HTTP request and response messages in compact format. The technique also applies to textual files, which normally are HTML, XML, JavaScript, CSS or binary content. Lossy compression usually applies to multimedia contents, such as icons, pictures, and videos. For example, Opera Mini [104] conducts transcoding for images and other multimedia web content before forwarding to the web browser. Besides minimising the content within a webpage, Delta ending [105] introduces a technique to identify the difference between sequential requested resources and only the data differences are transmitted to avoid the unnecessary network traffic caused by frequent web content updates and modifications. The solutions were designed for accelerating webpage fetching by altering original web content, which, unlike the .mobi version of the site, may not necessarily be what the web content owners intend for the mobile audience. As mentioned in Section 3.3, certain conditions need to be fulfilled so that these techniques can assure both fast content delivery and reduction of power consumption for mobile web contents.

- **Web Caching and Prefetching**: A further energy-saving technique is web caching, which keeps copies of web content either on a browser cached locally or a proxy cache remotely. When subsequent HTTP requests for the same content are made, the cache returns with either a hit or a miss to indicate the existence of content on the cache. If it is a hit, the web content is transmitted from the cache directly instead of from web server. In mobile networks, web caching is crucial to speed up content delivery and reduce mobile network traffic, as a cache proxy in a mobile network typically serves many users, avoiding repeated requests of the same content from the original content source. On the other hand, the reduction of delivery time leads to reduced power consumption of mobile web access and notable user experience improvement. As indicated in the study [106] by Qian et al., the redundant contents contribute about 20% of the total mobile HTTP traffic volume and are responsible for 7% of the radio energy consumption. However, the challenge remains on how to efficiently maintain consistency between the cached content and the frequently changed data source. Thus, it is important to improve the hit ratio of not only static content but also dynamic content to further reduce download latency and power consumption. Increasing the cache size only will not significantly improve the effectiveness of the hit ratio on a mobile browser though [107]. Therefore, research has been focused on improving the replacement algorithms and how to cache style and layout data for Document Object Mode (DOM) elements to reduce style formatting and layout calculation time [108].

  While web caching utilises the temporal locality of web objects, another technique often combined with caching is web prefetching, which utilises spatial locality of the web objects. Prefetching predicts which web page user will visit in the neat future and download the pages beforehand based on the user's visiting history or the content of visited pages [109, 110].

- **Radio Resource Allocation**: In radio networks, the RRC states determine the allocation of radio resources and power consumption state of mobile devices as described in Section 3.2.1. The interplay between mobile applications and the state machine of RRC behaviour causes inefficiencies of the resources including radio resources, network signalling traffic, device energy consumption and performance [17]. Statically configured inactivity timers may lead to either frequent state promotions and its corresponding transition delays and signal overheads if the timers are too short, or to over-occupation of radio resources and energy consumption of mobile devices. Thus, recent research has been focused on determining the optimal values of the inactivity timers and mitigating energy tail time effect. Finding the optimal values of the timer and tuning them is an effort to balance the energy wasted in waiting for the timers to expire and the effort by state promotions and demotions.

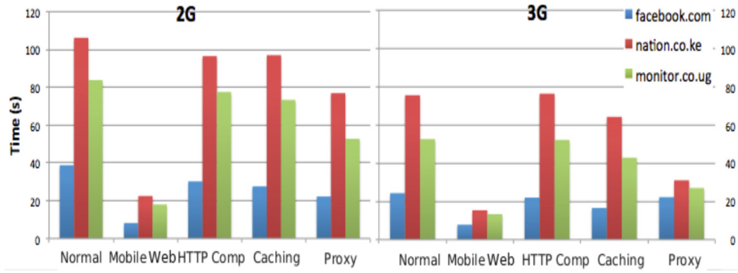
- **Performance Enhanced Proxy (PEP)**: Proxies have also been utilised to assist in energy saving. As an intermediary between mobile devices and web servers, the PEP is able to introduce a series of power saving assisted features, such as scheduling data packets for more energy-efficient traffic patterns, content adaptation for web browsing, prefetching, computation offloading and so on.

4.2 Using Proxy for Energy-Efficient Web Access

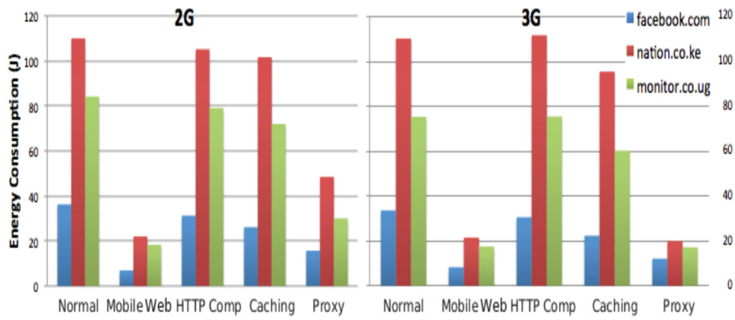
The previous sections described our understanding of power consumption of mobile data transmission, power consumption characteristics of various radio interfaces, as well as the trade-off between compression and data transmission. Based on the deep understanding, this section presents the architecture and design of a proxy-based solution for energy-efficient web access and the performance analysis.

4.2.1 Architecture of Energy-efficient Web Proxy

In order to design an energy-efficient proxy for web access, it is crucial to tackle the challenges in transmitting web content over wireless networks and shorten the transmissions on high-power consumption states as mush as possible. In addition, the solution has to be generic and transparent between mobile devices and web servers, and independent of mobile browser applications to accelerate deployment of the solution. Publication VII initialised basic requirements of how to design such a proxy-based architec-



(a) Time spent on fetching the sample web pages



(b) Energy consumed in fetching the sample web pages

Figure 4.1. Time and energy of fetching three sample web pages with different techniques

ture, taking compression, caching and bundling into consideration. The work evaluated and compared the performance of both using and not using proxy, proxy with compression, bundling,or both. The results show that using the proxy with bundling and compression decreases the delivery time of web content between mobile devices and web proxy, and its energy consumption, due to minimising the side-effect of TCP throughput caused by a potentially large delay between mobile devices and web sites in unpredictable wireless network environments. The results promise great potential, yet more work needs to be done to improve the design based on each radio link to enable more precise compression and bundling decisions, and power consumption reduction.

Thus, Publication VIII takes three East African countries as a case study to further evaluate different strategies for energy-efficient web access on mobile devices. By comparing the proxy-based solution with mobile optimisation, HTTP compression and web caching, the proposed solution reduces the energy consumption of accessing web content up to more than 59% for 2G networks and 74% for 3G networks, and the corresponding downloading time decreases up to 60%, as shown in Figure 4.1.

After the proxy for energy-efficient web access has been revisited, Publi-

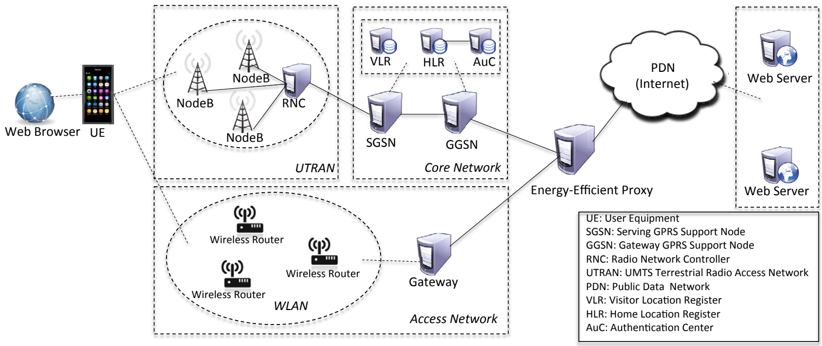


Figure 4.2. Architecture of energy-efficient proxy

cation IX proposes a newly designed architecture named Energy- Efficient Proxy (EEP), with a scheme of delivering web content to a mobile device as a whole instead of separate objects, RRC state-based header compression and selective content compression to keep radio in a low power state for longer durations and shorten downloading time. As a result, a huge reduction of energy consumption and increased QoE are achieved.

The architecture of the EEP is shown in Figure 4.2. Ideally, the proxy can be deployed by network operators enabling the proxy to be located as close as possible to mobile devices so that the delay between the mobile devices and proxy is minimised. The proxy is introduced between the mobile devices and web servers to split HTTP traffic into two portions, one of which is normal HTTP traffic between the proxy and web servers, the other is optimised content delivery with a number of enhancements over wireless links. The solution improves the energy efficiency of web access from the following aspects.

Firstly, the solution separates the TCP connection between the mobile device and web server. Without the mobile device explicitly requesting all the objects by itself, the proxy fetches the objects on behalf of the device. TCP, as a widely used transport protocol, was initially designed for wired networks, where physical links are reliable, and not for energy saving purpose. High packet loss rates and dramatical changing link quality in wireless networks forces TCP to retransmit in order to recover from errors. In addition, the TCP split results in lower connection overhead, better utilisation of the wireless network bandwidth, and higher robustness against link variances because of low delay of the E2E path. Besides, the mobile device utilises one single TCP connection to effectively retrieve web objects from the proxy instead of multiple persistent HTTP connections from web servers. Since modern websites are integrated with

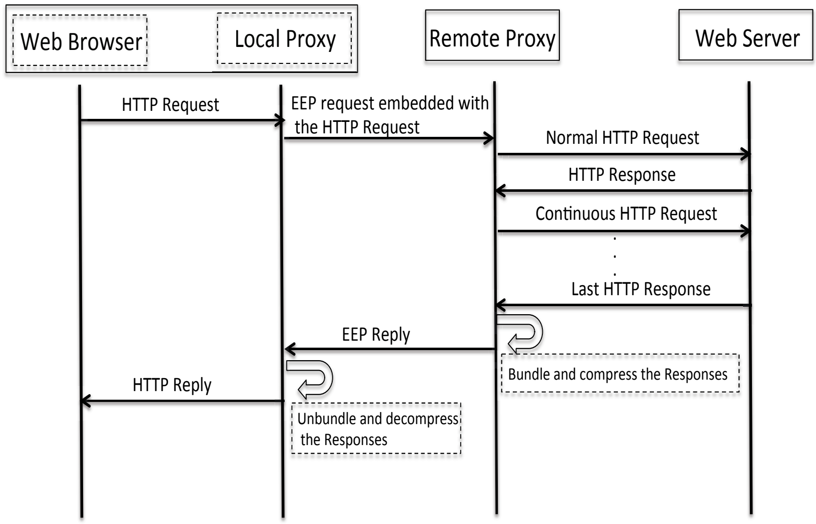


Figure 4.3. Flow chart of message exchange between the web browser, local proxy, remote proxy and web server

third-party content, such as web analytics tools, social media plugins and embedded advertisements, TCP connections have to be set up between the mobile device and multiple domains, resulting in high TCP connection overhead and a significant handshake delay due to the high latency of wireless links. With the proxy, the heavy-lifting can be offloaded from the mobile device to the proxy, where multiple TCP connections can be established fast to download the embedded objects from different servers, and DNS lookups can be accelerated.

Secondly, as seen in Figure 4.3, an HTTP request is forwarded from a mobile browser to the Local Proxy. Then the request is embedded in EEP payload and sent to the Remote Proxy. After the Remote Proxy parses the request, all the web objects associated with the request can be fetched from web servers. Once all the objects are received, the Remote Proxy reorders the sequence of the object request to accelerate rendering according to the DOM tree for each type of mobile web engine before the bundle is sent back to the mobile device. In case of inconsistence or missing objects, the Local Proxy performs requests for the content until the page is fully loaded. The bundling enables the optimisation of TCP behaviour over congested wireless links in order to keep the link utilised during the transmission. Also, the limited computation capability of a mobile device causes the mobile web browser to take a long time to download and process all objects. As a result, the data transmissions are spread along the whole downloading duration, and RRC timers would have never expired. Consequently, the radio interface is always on and radio resource cannot be released. With the bundling, the radio interface is able to enter a low power consumption state during the period of web object fetching in the Remote Proxy to achieve energy reduction.

Thirdly, the solution supports a range of enhancements to further reduce power consumption and download time. Carefully selecting compression on HTTP payload can provide energy saving when fulfilling certain conditions, which include considerations of link quality, computation load, file type and compression algorithms as discussed in Publication III. The solution adopts selective compression to decide whether to compress an object or not, based on the compression ratio of compressing the object and operating power of mobile devices required for decompressing during the web fetching. Also, the mobile devices may require a long time to request one object resulting in a long waiting time for radio interface to receive the object. Thus, caching is not only needed locally on mobile devices, but also needed on the Remote Proxy. If the content has been cached on the proxy, the bundling process retrieves the content from the cache directly; otherwise, the proxy sends requests to web servers. To maximise the cache hit rate, the Remote Proxy utilises content hash to eliminate redundant caching. The caching component generates cache indexes based on content hash rather than URLs to increase the hit rate on the proxy.

Moreover, a protocol named EEP protocol is defined to reduce protocol overhead instead of using HTTP with additional header fields. As a verbose protocol, HTTP is coded in standard, ASCII and the size of cookies could be up to 4096 bytes. Thus, it is necessary to reduce the number of bits sent over the air. The more important incentive to use a more compact format to transmit payload is to keep the size of the request from the mobile device to the Remote Proxy under the RRC state promotion threshold so that the radio interface remains in a low power consumption state while requesting and waiting for the bundles to come back in 3G networks.

4.2.2 Design of Energy-efficient Proxy

Embodying the above-mentioned requirements, two different design principles for the Energy-efficient proxy are presented as follows based on Publication IX.

One of the designs is to implement the Local Proxy as a native appli-

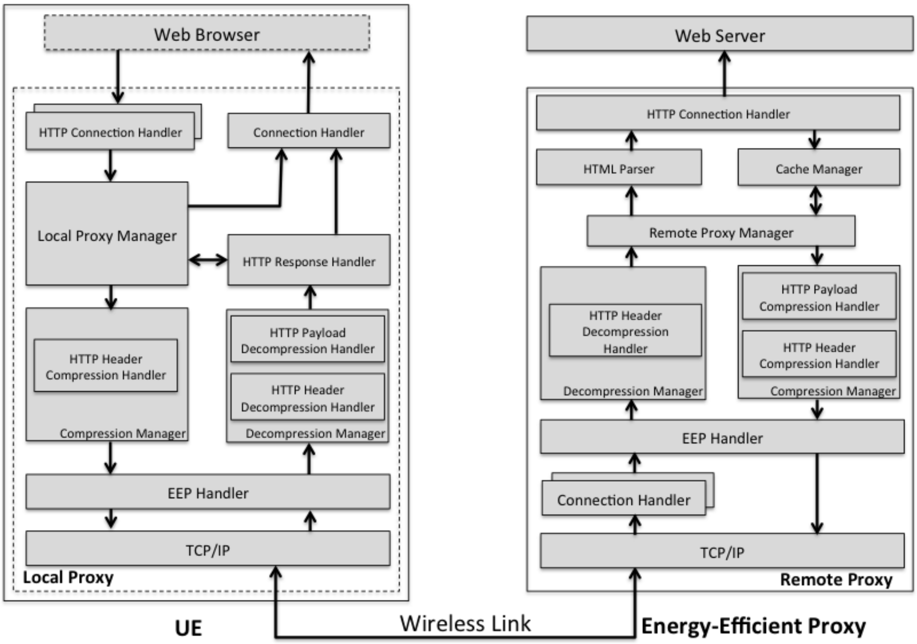


Figure 4.4. System design and components

cation on mobile devices to support described features and communicate with the Remote Proxy, as shown in Figure 4.4. The HTTP Connection Handler spawns itself to accept HTTP requests from the web browser while there is a new incoming request. Then the handler forwards the requests to the Local Proxy Manager, where the other handlers are invoked. The hash of each URL is calculated using SHA-1. The hashed indexes are stored in the Local Proxy Manager to map to the corresponding EEP reply, which consists of EEP header, the URL hash and compressed HTTP response. The hashed URL is analysed by the HTTP Response Handler first to check whether the reply is already stored in the HTTP Response Handler or not. In case of a miss, the Local Proxy Manager invokes the Compression Manager to compress the request before encapsulating it as EEP payload and sending it over the air. Figure 4.5 illustrates the protocol stack of EEP protocol that is enforced by the EEP handler. It enables the Local-Remote communication, where compression algorithms and levels are determined by an estimation of power consumption of compression/decompression, and downloading time for each transmission medium (2G, 3G or WLAN).

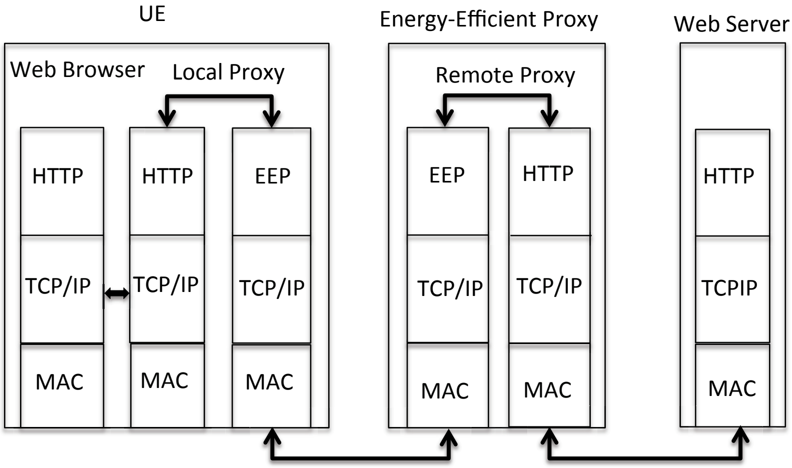


Figure 4.5. Protocol stack of native-based solution

After being received by the Connection Handlers in the Remote Proxy, the EEP requests are examined, and different actions are taken by the EEP Handler depending on the request types. If the type is for web objects, the requests are then forwarded to the Remote Proxy Manager after decompression. Upon each HTTP request, an instance of HTML Parser is invoked to act as a dedicated web engine. A webpage normally contains a number of web objects, not only the HTML page. These eventually create more than one HTTP request after parsing the HTML document. The engine is able to build a DOM tree based on the HTML document, but also able to evaluate JavaScripts, which may generate new requests for web objects. Therefore, all the web objects associated with the request can be fetched through the HTTP Connection Handler. When every HTTP response is received, the handler forwards the response to the parser so that the following HTTP requests can be generated. In the meantime, a copy of the response is forwarded to the Remote Proxy Manager, in which the Compression Manager is invoked to compress the response's header and the payload selectively. Since HTTP is stateless, HTTP cookies and some other header fields are used to maintain consistency between the web browser and web servers. This is the reason that HTTP response headers are also kept in EEP replies. After all the web objects are downloaded, the Remote Proxy Manager sends them back in sequence as a bundle to the Local Proxy.

To install the native application for each and every mobile device that expects to engage with the service is a a challenging deployment issue.

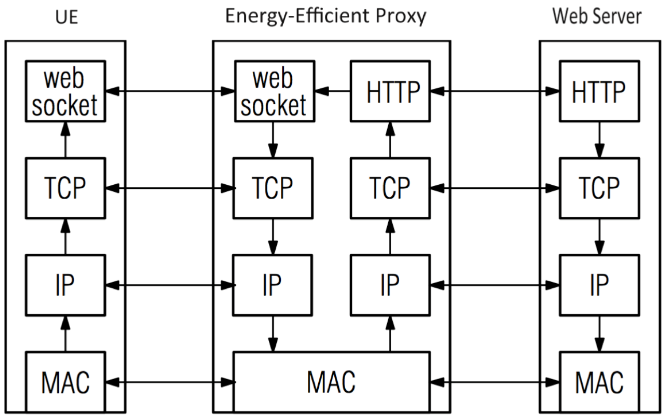


Figure 4.6. Protocol stack of WebSocket-based solution

To overcome the limitation, another design is proposed [111], as shown in Figure 4.6. Instead of requiring installation of native application on mobile devices, this design only requires mobile browsers to support WebSocket [112] and WebStorage [113], which have already been widely supported by most modern mobile browsers. In this design, HTTP requests are sent to the Remote Proxy directly. In response to receiving a request for content from a mobile browser, the Remote Proxy replies with a response containing instructions configured to set up a bi-directional communication channel using WebSocket APIs on the mobile device for communication between the proxy and the device. Meanwhile, a JavaScript library is sent to the mobile browser as well and will act as a handler to receive bundles, unbundle, decompress content, and store the post-processed content on local storage of the mobile browser using WebStorage APIs. Then the Remote Proxy fetches all the objects and sends them in a bundle with all the enhancements to the mobile browser via the established WebSocket, similarly to sending a bundle with EEP protocol. Inside the bundle, the HTML page is modified to support the WebSocket-based solution, where URLs to each object are changed to refer to where the objects are stored in local storage. Once the bundle is processed with the JavaScript library, the modified HTML page is sent to the mobile browser to render all the stored objects from the local storage.

4.2.3 Evaluation and Performance

The Energy-Efficient Proxy was implemented on commercial smartphones and thoroughly evaluated through experiments in both WLAN and 3G networks, with different test cases in order to answer the following ques-

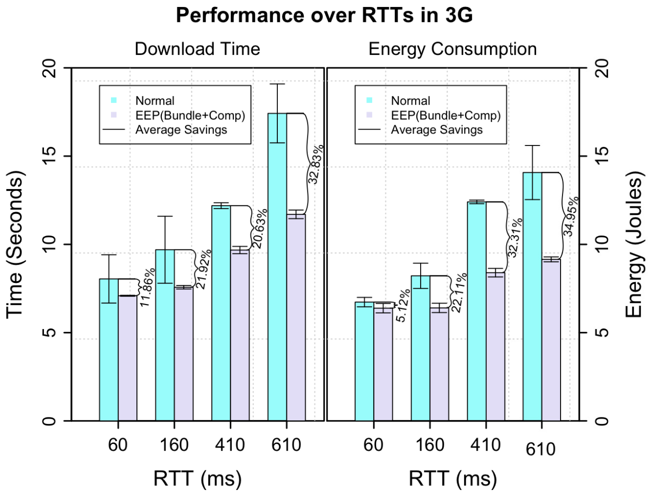


Figure 4.7. Download time and energy consumption of a webpage over different RTTs in 3G

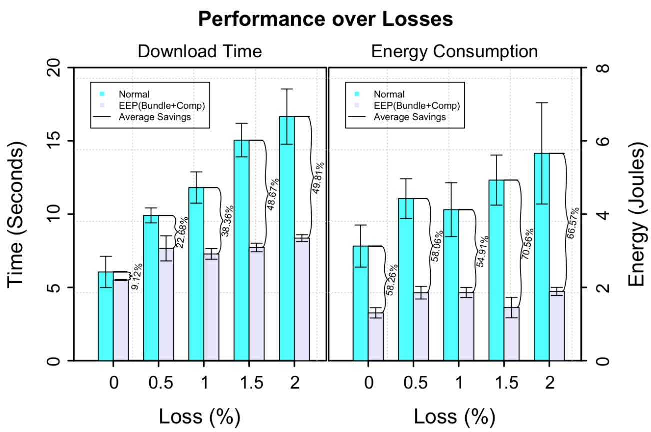


Figure 4.8. Download time and energy consumption of a webpage over different packet loss rates in WLAN

tions: (1) How much can the proxy speed up mobile web access? (2) How much energy can the proxy save? (3) How does web content, network delay and link speed affect the results? (4) How do the inactivity timers affect the results in a 3G network? and (5) How many changes do hardware and OSes cause?

The results show that the performance of downloading and power consumption is tolerant to the network delay and packet losses. Compared to the performance of using normal browsing, the solution can save up to 32% of downloading time and 34% of energy when experiencing huge network latency in 3G networks as shown in Figure 4.7. As the packet loss rate grows from 0% to 2.0%, the time saved by using the proxy increases from 9.12% to nearly 50%. Given the measurement cases, the energy can be saved over 58.26% when there is no packet loss, and increases to nearly 70.56% when the packet loss rate grows over 1.5% in the WLAN network, as shown in Figure 4.8. The similar trends can be found in 3G networks as well. The RRC inactivity timers control the demotions of mobile devices and radio resource release. In the evaluation, the EEP is able to save up to 43% of downloading time and 38% of energy consumption when small values of the inactivity timers are configured. More illustrated results can be found in Publication IX. The solution also favours savings over larger webpages. Moreover, the evaluation shows that the solution gives significant improvement of downloading time and energy savings on both Nokia Meego and Google Android platforms. With more powerful CPU/GPU and modern radio chipset, the better performance the solution offers, due to faster execution of unbundling, decompression, JavaScript execution, page rendering, and lower power consumption of radian interfaces.

4.3 Summary

As already discussed in Chapter 3, it is important to provide effective energy-saving solution for mobile web access to extend battery life, improve QoE, benefit business, and bridge the digital divide at large. Thus, this thesis focuses on providing solutions for energy-efficient mobile web access. As discussed in Section 4.1, the prior energy-saving strategies for mobile web access have been reviewed and categorised in the areas of mobile web optimisation, compression, web caching, prefetching, radio resource allocation and proxy-based solutions. The thesis proposes several energy saving techniques, such as traffic pattern shaping based on the power consumption characteristics of mobile data transmission, adaptive data compression and RRC-state-based web access tuning. Finally, the thesis presents the proxy-based architecture for energy-efficient mobile web access and its implementation that takes the advantages of each proposed technique and is proven to be an effective solution for not only significant energy savings, but also non-neglectable improvement of QoE in terms of faster content retrieval.

5. Conclusion

Mobile Internet is growing at a fast pace, with new opportunities and problems emerging. To enable sustainable mobile Internet growth and continue mobile service adaption, it is important to ensure that the reduction of overall environmental presence and the level of QoE are mutually addressed.

5.1 Summary and Discussion

The high-level objective of this dissertation is to reduce power consumption of mobile devices, extend battery life, yet maintain or even increase user experience. In order to achieve these goals, the first effort is to understand the power consumption characteristics of communications on mobile devices. The research has employed measurements and proposed power models based on thorough measurement data. The work also investigated the impact of data compression technologies on mobile data transmission, and defined the guideline of how to gain energy-efficient communications with data compression. With the deep insights obtained from the study, this research applies the knowledge to favour mobile web access with the proposed architecture to improve energy efficiency of data transmission without hindering QoE. To answer the motivations of this thesis mentioned in Chapter 1, the main contributions are highlighted here:

- Characterising power consumption of mobile data transmission
- Identification of main causes of battery drain of mobile devices
- Modelling power consumption of mobile data transmission and RRC

power consumption states based on thorough measurements

- Evaluation of data compression technologies and identification of conditions for energy-efficient mobile data transmission

- Data transmission optimisation for energy-efficient mobile web access

- Energy-efficient web proxy to reduce power consumption, shorten transmission time and improve QoE for mobile web access

Beyond the focus of this dissertation, there are still several topics worth discussing. One consideration is about security and privacy, which are persistent issues in web access. Privacy considerations have especially drawn too much attention recently. Personal data and browsing behaviour are becoming more sensitive and easy to leak in a cloud environment. As suggested in RFC 7258 [114], pervasive monitoring is a practical approach for analysing Internet traffic, but now it is considered an attack on the privacy of Internet uses and organisations. Some works have been proposed for secure web browsing by modularising the web browser and limiting communication within the modules or subsystems [115]. But the de-facto approach is to enable HTTPS when browsing the Internet. While speeding up the deployment of HTTPS tunnels, it has become difficult to process web traffic on proxies and other gateways for caching, enhancing performance as well as decreasing power consumption for mobile devices. In order to keep the success and the presence of the intermediaries, one proposal [116] is to support Explicitly Authenticated Proxy (EAP), which is an HTTP proxy to intercept the TLS-encrypted connection between a user and a targeting service server, with a certification authenticated and acknowledged by the user. With the user's permission, the proxy is able to continue the enhancements for existing Internet services. When taking privacy into consideration, the design decision in this dissertation is that all HTTPS traffic is bypassed to avoid violating users' privacy at the cost of losing all the enhancements, even including basic caching, instead of generating a certificate for the user to accept and decrypt HTTPS traffic on the proxy. However, as part of future work, the EEP should be extendable to support Explicitly Authenticated Proxy when it becomes mature.

The proxy can be a service offered by an independent third party, or, for example, a telecom operator's serving gateway could integrate the tech-

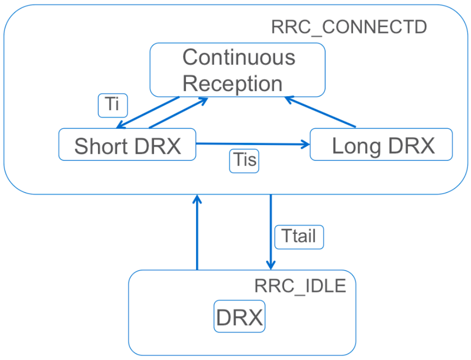


Figure 5.1. Radio Resource Control state machine of LTE

nology to provide the service for their customers. In fact, the solution can also be deployed and integrated as a part of customer premises equipment or femtocells to serve home or corporate users. A further deployment scenario would be to integrate the technology directly into a content server. In this way, the energy-efficient delivery of content can be offered by the content provider without a third party in the middle. Another finding in EEP measurements is that using the proxy is more beneficial when transmitting over slow or congested wireless links.

With the increasing deployment of LTE technology, it is worth discussing how the EEP would perform in LTE networks. Compared to the RRC state machine in UMTS networks, LTE has only two states, namely RRC_CONNECTED and RRC_IDLE, as shown in Figure 5.1. In the RRC_CONNECTED state, a UE can be in one of the three modes: Continuous Reception, Short DRX (Discontinuous Reception) and Long DRX. The Short DRX and the Long DRX have same cycle duration, but with different DRX cycle length, which is the number of frames in the paging cycle; The larger the cycle length is, the lower the UE battery power consumption is. In the RRC_IDLE state, there is no RRC connection and the UE is only in DRX mode. The DRX modes in RRC_CONNECTED and RRC_IDLE operates similarly, but with different parameter settings [117].

In the RRC_IDLE state, the UE can have the following processes: PLMN selection, cell selection and re-selection, location registration, and support for manual CSG (Closed Subscriber Group) selection. When there is a packet transmission, a state promotion from the RRC_IDLE state to the RRC_CONNECTED state occurs with a delay. After being promoted to the RRC_CONNECTED state, the RRC connection of the UE is established with the serving eNodeB. Consequently, the UE enters the Continuous Reception mode and keeps monitoring the PDCCH (Physical Downlink Control Channel) for control messages from eNodeB. Meanwhile, its power consumption follows the DRX procedure. When there is no transmission, a DRX inactivity timer Ti starts. Upon Ti's expiration without seeing any data activity, the UE enters the Short DRX mode, during which it can switch off main RF circuit and reduce power consumption. The Long DRX cycles begin after the Short DRX cycle timer Tis expires, if there is no data activity. When there is still no data transmission, the UE enters the Long DRX mode. The UE always enters the Continuous Reception mode when there is data transmission. Upon the data transmission, the UE starts a tail timer, Ttail, which is reset every time a packet is sent or received. When Ttail expires, the UE releases radio resource and is demoted from the RRC_CONNECTED state to the RRC_IDLE state [118, 119].

As above-mentioned, the RRC states of LTE networks is quite different from the ones in 3G networks with respect to data rate, inactivity timers, power consumption states and the transition among the states. Thus, an estimation would be that the benefits of using the EEP proxy may decrease. For example, the HTTP header compression used in EEP to keep UE in the Cell_FACH state is not valid anymore in 4G/LTE networks. The bundling concept would still be valid but its benefit might decrease due to less time needed for transmitting bundled content. However, with billions of connected devices and complicated use cases, part of network we will experience might be over-congested and perceived data speed might not be as fast as it could. Thus, the bundle and the EEP can provide benefits in LTE networks too, but we need more investigation on the operation and optimisation of the system and how the EEP can be best integrated with the LTE RRC timers and bit rates.

5.2 Further Research

Future work can be elaborated here based on the discoveries and results of this thesis. First, as discussed above, network conditions have a significant impact on power consumption. It remains an open question, though, how to show the impact explicitly in the power models that are designed for application developers. While the EEP protocol is designed for improving HTTP traffic, the theoretical thinking of scheduling traffic in a bundle in this dissertation can be easily extended and applied to other non real-time services. The design of the EEP proxy has the potential to adapt for other kinds of Internet services, which have similar interactions with HTTP between application behaviour and the underlying protocols. The design requires the proxy to be aware of the application types and data transmission mediums in order to optimise the transmission according to the characteristics of the applications.

As cloud computing maximises the effectiveness of shared resources and adopts dynamically to changed service requirements, the deployment of the proxy should be also cloud-based. Virtual machine and Linux container based solutions are often compared to each other. Virtual machines have a full OS with its own memory management installed, running on a resource emulated environment on top of hypervisor (KVM, Xen and HyperV). Due to this nature, a virtual machine has the associated overhead of virtual device drivers. On the other hand, a Linux container, such as Docker container [120], runs as a process of the host system and relies on control groups to manage groups of processes, CPU, memory and block I/O usage. As a lightweight virtualisation technology, Linux containers are therefore faster, less resource demanding and can be launched in just a few seconds while launching a virtual machine can take up to several minutes.

There are advantages and disadvantages for each type of visualisation technology. Depending on the requirements of the execution environment of the proxy service, a virtual machine is able to provide full isolation with guaranteed resources to fulfil the security and privacy requirements of the service. With the deployment of the proxy in containers, the service can be easily and quickly scaled out according to the amount of traffic, the number of requests and the CPU requirements.

Moreover, HTML5 technologies and mobile cloud computing are diversifying and growing at an unprecedented speed. For example, Mozilla's Firefox OS [121] is a web-engine-based mobile operation system, and all its applications are based on HTML5. The adoption of interactive technologies and feature-sets of mobile web browsers is growing and maturing. As discussed in Section 4.2, the WebSocket-based proxy not only unveils the possibility of using HTML5 technologies for fast deployment of the proxy without pre-installing any application, but also presents the power to develop cross-browser and cross-device energy-saving solutions and services seamlessly.

Based on the understanding of this dissertation, some implications can be also drawn for app developers to optimise their services and reduce energy consumption: 1) using the right data compression algorithms; 2) scheduling some transfers based on the available radio technologies; 3) bundling small transfers when possible into a single longer transmission; 4) last but not least, signal strength is always a good indictor for when to transfer data. However, currently mobile application development APIs are more feature-centric, focusing on providing rich set of functions to fulfil implementation requirements rather than performance requirements. Performance optimisation is often done at system level for all running apps. Thus, certain system level information, such as current RRC status and predicted signal strength, should be presented in an easy-to-understand way and exposed to developers as APIs for further optimisation and energy savings.

References

[1] Le Wang, Edward Mutafungwa, Yeswanth Puvvala, and Jukka Manner. Strategies for Energy-Efficient Mobile Web Access: An East African Case Study. The 3rd International ICST Conference on e-Infrastructureand e-Services for Developing Countries, 2011.

[2] Nielsen. STATE OF THE APPNATION-A YEAR OF CHANGE AND GROWTH IN U.S. SMARTPHONES. http://www.nielsen.com/us/en/newswire/2012/state-of-the-appnation-%C3%A2%C2%80%C2%93-a-year-of-change-and-growth-in-u-s-smartphones.html, referred to 18.5.2015.

[3] MOSTI and PIKOM. ICT STRATEGIC REVIEW 2012/13. http://www.pikom.org.my/cms/ICT_Strategic_Review/ICT_Strategic_Review_2012-2013.pdf, referred to 12.5.2015.

[4] Sofie Lambert, Ward van Heddeghem, Willem Vereecken, Bart Lannoo, Didier Colle, and Mario Pickavet. Worldwide electricity consumption of communication networks. The International Online Journal of Optics, Volume: 20, Issue: 26, Page(s): 513-524, 2012.

[5] Parthasarathy Ranganathan. A Recipe for Efficiency? Some Principles of Power-aware Computing. Communications of the ACM, Volume: 53 No. 4, Page(s): 60-67, 2010.

[6] Aaron Carroll and Gernot Heiser. An Analysis of Power Consumption in a Smartphone. Proceedings of the 2010 USENIX conference on USENIX annual technical conference, Page(s): 21, 2010.

[7] Kenneth Barr and Krste Asanovic. Energy-aware lossless data compression. The First International Conference on Mobile Systems, Applications, and Services, Volume: 24, Issue: 3, Page(s): 250-291, 2006.

[8] Laura M. Feeney and Martin Nilsson. Investigating the energy consumption of a wireless network interface in an ad hoc networking environment. Proceedings of the IEEE INFOCOM, Volume: 3, Page(s): 1548-1557, 2001.

[9] Jean-Pierre Ebert, Brian Burns, and Adam Wolisz. A trace-based approach for determining the energy consumption of a wlan network interface. Proceedings of European Wireless, Page(s): 230-236, 2002.

[10] Gian Paolo Perrucci, Frank H.P. Fitzek, Giovanni Sasso, Wolfgang Kellerer, and Jórg Widmer. On the Impact of 2G and 3G Network Usage for Mobile Phones? Battery Life. The 15th IEEE European Wireless Conference, Page(s): 255 - 259, May 2009.

[11] Niranjan Balasubramanian, Aruna Balasubramanian, and Arun Venkataramani. Energy consumption in mobile phones: a measurement study and implications for network applications. Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, Page(s): 280-293, 2009.

[12] Ashish Sharma, Vishnu Navda, Ramachandran Ramjee, Venkat Padmanabhan, and Elizabeth Belding. Cool-Tether: Energy Efficient On-the-fly WiFi Hot-spots using Mobile Phones. Proceedings of the 5th ACM International Conference on Emerging Networking Experiments and Technologies, Page(s): 109-120, 2009.

[13] Rong Xu, Zhiyuan Li, Cheng Wang, and Peifeng Ni. Impact of Data Compression on Energy Consumption of Wireless-Networked Handheld Devices. Proceedings of the 23rd IEEE International Conference on Distributed Computing Systems, Page(s) 302-311, 2003.

[14] Rakan Maddah and Sanaa Sharafeddine. Energy-Aware Adaptive Compression Scheme for Mobile-to-Mobile Communications. The 10th IEEE International Symposium on Spread Spectrum Techniques and Applications, Page(s) 688-691, Aug, 2008.

[15] Jui-Hung Yeh, Jyh-Cheng Chen, and Chi-Chen Lee. Comparative Analysis of Energy-Saving Techniques in 3GPP and 3GPP2 Systems. IEEE Transactions on Vehicular Technology, Volume: 58, Issue: 1, Page(s): 432-448, 2009.

[16] Jui-Hung Yeh, Chi-Chen Lee, and Jyh-Cheng Chen. Performance analysis of energy consumption in 3GPP networks. IEEE Wireless Telecommunications Symposium, Page(s): 67-72, 2004.

[17] Qian Feng, Zhzoguang Wang, Alexandre Gerber, Z.Norley Mao, Subhabrata Sen, and Oliver Spatscheck. Characterizing radio resource allocation for 3G networks. Proceeding of the 10th ACM SIGCOMM conference on Internet Measurement, Page(s): 137-150, 2010.

[18] Feng Qian, Zhaoguang Wang, Alexandre Gerber, Z. Morley Mao, Subhabrata Sen, and Oliver Spatscheck. TOP: Tail Optimization Protocol For Cellular Radio Resource Allocation. The 18th IEEE International Conference on Network Protocols, Page(s): 285-294, 2010.

[19] Hao Liu, Yaoxue Zhang, and Yuezhi Zhou. TailTheft: Leveraging the Wasted Time for Saving Energy in Cellular Communications. Proceedings of the 6th ACM International Workshop on MobiArch, Page(s): 31-36, 2011.

[20] Di Zhang, Yaoxue Zhang, Yuezhi Zhou, and Hao Liu. Leveraging the Tail Time for Saving Energy in Cellular Networks. IEEE Transactions on Mobile Computing, Volume: 13, Issue: 7, Page(s): 1536-1549, 2014.

[21] Marcel C. Rosu, C. Michael Olsen, Chandra Narayanaswami, and Lu Luo. PAWP: A Power Aware Web Proxy for Wireless LAN Clients. Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications, Page(s): 206-215, 2004.

[22] Miguel Jimeno and Ken Christensen ad Bruce Nordman. A Network Connection Proxy to Enable Hosts to Sleep and Save Energy. IEEE International Conference of Performance, Computing and Communications, Page(s): 101-110, 2008.

[23] Bo Zhao, Byung Chul Tak, and Guohong Cao. Reducing the Delay and Power Consumption of Web Browsing on Smartphones in 3G Networks. The 31st IEEE International Conference on Distributed Computing Systems, Page(s): 413-422, 2011.

[24] Cisco. Cisco Visual Networking Index: Forecast and Methodology, 2013-2018. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html, referred to 4.12.2015.

[25] Janessa Rivera. Gartner Says Worldwide PC Shipments in the First Quarter of 2013 Drop to Lowest Levels Since Second Quarter of 2009. http://www.gartner.com/newsroom/id/2420816, referred to 4.12.2015.

[26] Christy Pettey. Gartner Says Worldwide PC Shipments in the Third Quarter of 2014 Declined 0.5 Percent. http://www.gartner.com/newsroom/id/2869019, referred to 4.12.2015.

[27] mobiThinking. Global mobile statistics 2014. http://mobiforge.com/research-analysis/global-mobile-statistics-2014-home-all-latest-stats-mobile-web-apps-marketing-advertising-subscriber, referred to 4.12.2015.

[28] Ericsson. Ericsson Mobility Report. http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf, referred to 4.12.2015.

[29] Facebook. Facebook Q1 2014 Earnings. http://investor.fb.com/results.cfm, referred to 4.12.2015.

[30] Amit Kumar, Yunfei Liu, and Jyotsna Sengupta Divya. Evolution of Mobile Wireless Communication Networks: 1G to 4G. The International Journal of Electronics & Communication Technology, Volume: 1, Issue: 1, Dec 2010.

[31] Tinatin Mshvidobadze. Evolution mobile wireless communication and LTE networks. The 6th IEEE International Conference on Application of Information and Communication Technologies, Page(s): 1-7, Oct 2012.

[32] Rysavy Research. Mobile Broadband Explosion. http://www.academia.edu/4372400/4G_Americas_Mobile_Broadband_Explosion_August_2013_FINAL, 2013.

[33] Amel Haji, Asma Ben Letaifa, and Sami Tabbane. Integration of WLAN, UMTS and WiMAX in 4G. The 16th IEEE International Conference on Electronics, Circuits, and Systems, Page(s): 307-310, Dec 2009.

[34] Axis-Communication. The Mobile Internet and Wireless Networking: Opportunities, Challenges and Solutions. http://www.axis.com/files/whitepaper/mobile_access.pdf, referred to 6.2.2015.

[35] Xuan Feng, Qiang Tan, and Zongchuang Liang. The development of satellite mobile communication system. The 6th IEEE International ICST Conference on Communications and Networking, Page(s): 1110-1114, 2011.

[36] Mirette Sadek and Sonia Aissa. Personal satellite communication: technologies and challenges. IEEE Wireless Communications, Volume: 19, Issue: 6, 2012.

[37] Mary Meeker and Liang Wu. 2013 Internet Trends. http://www.kpcb.com/insights/2013-internet-trends, referred to 15.6.2015.

[38] Mary Meeker and Matt Murphy. Top 10 Mobile Internet Trends. http://www.kpcb.com/insights/top-10-mobile-internet-trends, referred to 15.6.2015.

[39] Tim Berners-Lee. The WorldWideWeb browser. http://www.w3.org/People/Berners-Lee/WorldWideWeb.html, referred to 5.6.2015.

[40] Tim O'Reilly. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. http://oreilly.com/web2/archive/what-is-web-20.html, referred to 2.6.2015.

[41] San Murugesan. Understanding Web 2.0. IEEE IT Professional, Volume: 9, Issue: 4, Page(s): 34-41, July-Aug 2007.

[42] Ian Chard. Share, Collaborate, Exploit Defining Mobile Web 2.0. http://www.juniperresearch.com/shop/viewwhitepaper.php?whitepaper=63, referred to 17.4.2015.

[43] Toshihiko Yamakam. Mobile Web 2.0: Lessons from Web 2.0 and Past Mobile Internet Development. International Conference on Multimedia and Ubiquitous Engineering, 2007.

[44] Tim Berners-Lee, Jim Handler, and Ora Lassila. The Semantic Web. Scientific American, Volume: 284, Issue: 5, Page(s): 35-43, 2001.

[45] Ora Lassila and James Hendler. Embracing "Web 3.0". IEEE Internet Computing, Volume:11, Issue: 3, Page(s): 90-93, 2007.

[46] Jim Handler. Web 3.0 Emerging. IEEE Computer, Volume:42, Issue:1, Page(s): 111-113, 2009.

[47] Kim W. Tracy. Mobile application development experiences on Apple's iOS and Android OS. IEEE Potentials, Volume: 31, Issue: 4, Page(s): 30-34, Aug 2012.

[48] Kristofer Kimbler. App Store Strategies for Service Providers. The 14th IEEE International Conference on Intelligence in Next Generation Networks, Page(s): 1-5, 2010.

[49] Marcelo Nogueira Cortimiglia, Antonio Ghezzi, and Filippo Renga. Mobile Applications and Their Delivery Platforms. IT Professional, Volume:13, Issue: 5, Page(s): 51-56, Sept-Oct 2011.

[50] Neolane. The Age of Apps: Evolution of the Mobile Application. http://blog.neolane.com/mobile-marketing-2/mobile-application/, referred to 12.6.2015.

[51] Accenture. Mobile Web Watch 2012. http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Mobile-Web-Watch-Internet-Usage-Survey-2012.pdf, referred to 2.6.2015.

[52] Christy Pettey and Laurence Goasduff. Gartner Identifies 10 Consumer Mobile Applications to Watch in 2012. http://www.gartner.com/newsroom/id/1544815, referred to 22.6.2015.

[53] Christof Weinhardt, Arun Anandasivam, Benjamin Blau, Nikolay Borissov, Thomas Mein, Wibke Michalk, and Jochen Stößer. Cloud Computing - A Classification, Business Models, and Research Directions. Business & Information Systems Engineering, Volume: 1, Issue 5, Page(s): 391-399, 2009.

[54] Joel Gibson, Darren Eveleigh, Robin Rondeau, and Qing Tan. Benefits and Challenges of Three Cloud Computing Service Models. 2012 Fourth International Conference on Computational Aspects of Social Networks, Page(s): 198-205, 21-23 Nov 012.

[55] Gartner. Gartner Says By 2016, 40 Percent of Mobile Application Development Projects Will Leverage Cloud Mobile Back-End Services. http://www.gartner.com/newsroom/id/2463615, referred to 4.6.2015.

[56] Vandana Milind Rohokale, Neeli Rashmi Prasad, and Ramjee Prasad. A cooperative Internet of Things (IoT) for rural healthcare monitoring and control. The 2nd IEEE International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Page(s): 1-6, 2011.

[57] Cisco. Evolution of the Mobile Network. http://www.cisco.com/en/US/solutions/collateral/ns341/ns973/white_paper_c11-624446.html, referred to 4.12.2015.

[58] Bai-Ruei Huangy, Chang Hong Liny, and Chia-Han Lee. Mobile Augmented Reality Based on Cloud Computing. 2012 International Conference on Anti-Counterfeiting, Security and Identification, Page(s): 1-5, Aug, 2012.

[59] Ericsson. Ericsson Mobility Report. http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-june-2013.pdf, referred to 8.21.2015.

[60] Harish Viswanathan. Getting Ready for M2M Traffic Growth. http://www2.alcatel-lucent.com/techzine/getting-ready-for-m2m-traffic-growth/, referred to 3.7.2015.

[61] Ericsson. Ericsson Energy and Carbon Report. http://www.ericsson.com/res/docs/2013/ericsson-energy-and-carbon-report.pdf, referred to 2.10.2015.

[62] Jens Malmodin, Permilla Bergmark, and Dag Ludnén. The future carbon footprint of the ICT and E&M sectors. Proceedings of the First International Conference on Information and Communication Technologies for Sustainability, Feb 2013.

[63] Jens Malmodin, Åsa Moberg, Dag Lundén, Göran Finnveden, and Nina Lövehagen. Greenhouse Gas Emissions and Operational Electricity Use in the ICT and Entertainment & Media Sectors. Journal of Industrial Ecology, Volume14, Issue: 5, Page(s): 770-790, Oct 2010.

[64] Mario Pickavet, Willem Vereecken, Sofie Demeyer, Pieter Audenaert, Brecht Vermeulen, Chris Develder, Didier Colle, Bart Dhoedt, and Piet Demeester. Worldwide energy needs for ICT: The rise of power-aware networking. The 2nd International Symposium on Advanced Networks and Telecommunication Systems, Page(s): 1-3, Dec. 2008.

[65] Commission of the European Communities. An Energy Policy for Europe. http://www.europarl.europa.eu/meetdocs/2004_2009/documents/com/com_com(2007)0001_/com_com(2007)0001_en.pdf, referred to 10.24.2015.

[66] Rich Lechner. An Inefficient Truth. IBM Global Action Plan, Aug, 2008.

[67] Commission of the European Communities. Limiting Global Climate Change to 2 degrees Celsius - The way ahead for 2020 and beyond. http://eur-lex.europa.eu/LexUriServ/site/en/com/2007/com2007_0002en01.pdf, referred to 10.18.2015.

[68] Vineetha Paruchuri. Greener ICT: Feasibility of Successful Technologies from Energy Sector. The 13th IEEE International Conference on Advanced Communication Technology, Page(s): 1398-1403, Fed 2011.

[69] Albrecht Fehske, Gerhard Fettweis, Jens Malmodin, and Gergely Biczók. The Global Footprint of Mobile Communications: The Ecological and Economic Perspective. IEEE Communications Magazine, Volume: 49, Issue: 8, Aug 2011.

[70] Centre of Energy-efficient Telecommunications. The Power of Wireless Cloud: An analysis of the energy consumption of wireless cloud. http://www.ceet.unimelb.edu.au/pdfs/ceet_white_paper_wireless_cloud.pdf, referred to 16.10.2015.

[71] Denzil Ferreira, Eija Ferreira, Jorge Goncalves, Vassilis Kostakos, and Anind K. Dey. Revisiting Human-Battery Interaction with an Interactive Battery Interface. Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Page(s): 563-572, Sept 2013.

[72] Khai N. Truong, Julie A. Kientz, Timothy Sohn, Alyssa Rosenzweig, Amanda Fonville, and Tim Smith. The Design and Evaluation of a Task-Centered Battery Interface. Proceedings of the 12th ACM International Conference on Ubiquitous Computing, Page(s) 341-350, 2010.

[73] Shunliu Deng, Yin Zhang, Alexandra H. Brozena, Maricris Lodriguito Mayes, Parag Banerjee, Wen-An Chiou, Gary W. Rubloff, George C. Schatz, and YuHuang Wang. Confined propagation of covalent chemical reactions on single-walled carbon nanotubes. Nature Coomunications 2, Jul 2011.

[74] Hermann Eul. The future of mobile broadband internet powered by the semiconductor industry. IEEE International SOC Conference, 2009.

[75] Kyunghan Lee, Joohyun Lee, Yung Yi, Injong Rhee, and Song Chong. Mobile Data Offloading: How Much Can WiFi Deliver? IEEE/ACM Transactions on Networking, Volume:21, Issue: 2, Page(s): 536 - 550, 2012.

[76] Aruna Balasubramanian, Ratul Mahajan, and Arun Venkataramani. Augmenting Mobile 3G Using WiFi. Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, Page(s): 209-222, 2010.

[77] Lide Zhang, Birjodh Tiwana, Robert P Dick, Zhiyun Qian, Mao Z Morley, Zhaoguang Wang, and Lei Yang. Accurate Online Power Estimation And Automatic Battery Behavior Based Power Model Generation for Smartphones. IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, Page(s): 105-114, 2010.

[78] Frank Zhao, Weijie Jia, Bin Wang, Guobao Xi, Mijun Wen, and Shuaiyang Lai. Smartphone Solutions White Paper. www.huawei.com/ilink/en/download/HW_193034, referred to 11.10.2015.

[79] Ziqi Zhang, Zhuyan Zhao, Hao Guan, Deshan Miao, and Zhenhui Tan. Study of Signaling Overhead Caused by Keep-Alive Messages in LTE Network. The 78th IEEE Vehicular Technology Conference, Page(s): 1-5, 2013.

[80] Henry Haverinen, Jonne Siren, and Pasi Eronen. Energy Consumption of Always-On Applications in WCDMA Networks. The 65th IEEE Vehicular Technology Conference, Page(s): 964-968, 2007.

[81] Kok-Kiong Yap, Te-Yuan Huang, Masayoshi Kobayashi, Yiannis Yiakoumis, Nick McKeown, Sachin Katti, and Guru Parulkar. Making Use of All the Networks Around Us: A Case Study in Android. Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design, Page(s): 19-24, 2012.

[82] Ning Ding, Daniel Wagner, Xiaomeng Chen, Abhinav Pathak, Y. Charlie Hu, and Andrew Rice. Characterizing and Modeling the Impact of Wireless Signal Strength on Smartphone Battery Drain. ACM SIGMETRICS Performance Evaluation Review - Performance evaluation review, Volume: 41, Issue: 1, Page(s): 29-40, 2013.

[83] Moo-Ryong Ra, Jeongyeup Paek, Abhishek B. Sharma, Ramesh Govindan, Martin H. Krieger, and Michael J. Neely. Energy-Delay Tradeoffs in Smartphone Applications. Proceedings of the 8th International Conference on Mobile systems, Applications, and Services, Page(s): 255-270, 2010.

[84] Ruohong Peng and Xi Wang. ICT Solutions Calculation Model for CO2 Emission Reduction and Prediction on its Emission Reduction Potential. The IEEE International Conference on Management and Service Science, Page(s): 1-5, Sept, 2009.

[85] Aaron Carroll and Cernot Heiser. An Analysis of Power Consumption in a Smartphone. Proceedings of the 2010 USENIX conference on USENIX annual technical conference, Page(s) 21-21, 2010.

[86] Maruti Gupta, Ai Taha Koc, and Rath Vannithamby. Analyzing mobile applications and power consumption on smartphone over LTE network. The 2011 International Conference on Energy Aware Computing, Page(s): 1-4, Nov 30 2011-Dec 2 2011.

[87] Stratos Keranidis, Giannis Kazdaridis, Virgilios Passas, Thanasis Korakis, Iordanis Koutsopoulos, and Leandros Tassiulas. Online Energy Consumption Monitoring of Wireless Testbed Infrastructure through the NITOS EMF Framework. Proceedings of the 8th ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, Page(s): 89-90, 2013.

[88] Gian Paolo Perrucci, Frank H.P. Fitzek, and Jórg Widmer. Survey on Energy Consumption Entities on the Smartphone Platform. The 73rd IEEE Vehicular Technology Conference, Page(s): 1-6, 2011.

[89] Chanmin Yoon, Dongwon Kim, Wonwoo Jung, Chulkoo Kang, and Hojung Cha. AppScope: Application Energy Metering Framework for Android Smartphones using Kernel Activity Monitoring. Proceedings of the 2012 USENIX Conference on Annual Technical Conference, Page(s) 36-36, 2012.

[90] Abhinav Pathak, Y. Charlie Hu, Ming Zhang, Paramvir Bahl, and Yi.Min Wang. Fine-grained power modeling for smartphones using system call tracing. Proceedings of the 6th ACM European Conference Computer Systems, Page(s): 153-168, 2011.

[91] Rajesh Palit, Ajit Singh, and Kshirasagar Naik. Modeling the Energy Cost of Applications on Portable Wireless Devices. Proceedings of the 11th AMC International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Page(s): 346-353, Oct, 2008.

[92] Mian Dong and Lin Zhong. Self-constructive high-rate system energy modeling for battery-powered mobile systems. Proceedings of the 9th International Conference on Mobile systems, applications, and services, Page(s) 335-348, 2011.

[93] Adam Oliner, Anand Padmanabha Iyer, Ion Stoica, Eemil Lagerspetz, and Sasu Tarkoma. Carat: Collaborative Energy Diagnosis for Mobile Devices. Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. Article No. 10, 2013.

[94] The 3rd Generation Partnership Project (3GPP). IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput. IEEE Std 802.11n-2009, Page(s) C1-502, 2009.

[95] The 3rd Generation Partnership Project (3GPP). IEEE Standard for Information Technology - Telecommunications and In- formation Exchange

Between Systems - Local and Metropolitan Area Net- works - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11- 2007 (Revision of IEEE Std 802.11-1999), Page(s) C1-1184, Jun, 2007.

[96] The 3rd Generation Partnership Project (3GPP). Radio resource control (rrc) protocol specification. 3GPP TS 25.331. http://www.3gpp.org/DynaReport/25331.htm, referred to 13.11.2015.

[97] Pekka H.J. Perala, Antonio Barbuzzi, Gennaro Boggia, and Kosta Pentikousis. Theory and practice of RRC state transitions in UMTS networks. Proceeding of 5th IEEE Broadband Wireless Access Workshop colocated with IEEE Globecom, BW- WAWS, 2009.

[98] The 3rd Generation Partnership Project (3GPP). Configuration of Fast Dormancy in Release 8. 3GPP discussion and decision notes RP-090960. , 2009.

[99] Kenneth Barr and Krste Asanovic. Energy-aware lossless data compression. ACM Transactions on Computer Systems, Volume: 24, Issue: 3, Page(s) 250-291, 2006.

[100] Gen Hattori, Keiichiro Hoashi, Kazunori Matsumoto, and Fumiaki Sugaya. Robust Web Page Segmentation for Mobile Terminal Using Content-Distances and Page Layout Information. Proceedings of the 16th ACM International Conference on World Wide Web, Page(s) 361-370, 2007.

[101] Michael Nebeling, Maximilian Speicher, and Moira C. Norrie. CrowdAdapt: Enabling Crowdsourced Web Page Adaptation for Individual Viewing Conditions and Preferences. Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive Computing systems, Page(s): 23-32, 2013.

[102] dotMobi. .mobi. http://dotmobi.mobi/, referred to 21.10.2015.

[103] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. The Internet Engineering Task Force, RFC 2616, Jun 1999.

[104] Opera Software ASA. Opera Mini. http://www.opera.com/mobile/features/, referred to 15.10.2015.

[105] J. Mogul, B. Krishnamurthy, F. Douglis, A. Feldmann, Y. Goland, A. van Hoff, and D. Hellerstein. Delta encoding in HTTP. The Internet Engineering Task Force, RFC 3229, Jun 2002.

[106] Feng Qian, Kee Shen Quah, Junxian Huang, Jeffrey Erman, Alexandre Gerber, Z. Morley Mao, Subhabrata Sen, and Oliver Spatscheck. Web Caching on Smartphones: Ideal vs. Reality. Proceedings of the 10th ACM International Conference on Mobile Systems, Applications, and Services, Page(s): 127-140, 2012.

[107] Zhen Wang, Felix Xiaozhu Lin, Lin Zhong, and Mansoor Chishtie. How effective is mobile browser cache? Proceedings of the 3rd ACM Workshop on Wireless of the Students, by the Students, for the Students, Page(s): 17-20, 2011.

[108] Kaimin Zhang, Lu Wang, Aimin Pan, and Bin B. Zhu. Smart caching for web browsers. Proceedings of the 19th ACM International Conference on World Wide Web, Page(s): 491-500, 2010.

[109] Josep Domenech, Ana Pont-Sanjuan, Julio Sahuquillo, and Jose A. Gil. Evaluation, Analysis and Adaptation of Web Prefetching Techniques in Current Web. Web-based Support Systems Springer, Page(s): 239-271, 2010.

[110] Yin-Fu Huang and Jhao-Min Hsu. Mining web logs to improve hit ratios of prefetching and caching. Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, Page(s): 577-580, 2005.

[111] Jukka Manner Le Wang. A server, a method for operating a server and a system. http://lib.tkk.fi/Reports/2012/urn100594.pdf, 2012.

[112] I. Fette and A. Melnikov. The WebSocket Protocol. The Internet Engineering Task Force, RFC 6455, Dec 2011.

[113] Ian Hickson. Web Storage. http://www.w3.org/TR/webstorage/, referred to 11.11.2015.

[114] S. Farrell and H. Tschofenig. Pervasive Monitoring Is an Attack. The Internet Engineering Task Force, RFC 7258, May 2014.

[115] Chris Grier, Shuo Tang, and Samuel T. King. Secure web browsing with the OP web browser. IEEE Symposium on Security and Privacy, Page(s): 402-416, 2008.

[116] S. Loreto, J. Mattsson, R. Skog, H. SpaakG. Bourg, D. Druta, and M. Hafeez. Explicitly Authenticated Proxy in HTTP/2.0. The The Internet Engineering Task Force Internet-Draft draft-loreto-httpbis-explicitly-auth-proxy-01, July 2014.

[117] The 3rd Generation Partnership Project (3GPP). TR 36.331, E-UTRA; Radio Resource Control (RRC); Protocol specification (Rel. 11). , March 2013.

[118] Junxian Huang. Performance and Power Characterization of Cellular Networks and Mobile Application Optimizations. http://deepblue.lib.umich.edu/bitstream/handle/2027.42/99905/hjx_1.pdf?sequence=1, 2013.

[119] Li-Ping Tung, Li-Chun Wang, Cheng-Wen Hsueh, and Chung-Ju Chang. Analysis of DRX power saving with RRC states transition in LTE networks. IEEE European Conference on Networks and Communications, Page(s): 301-305, 2015.

[120] Inc Docker. Docker-An open platform for distributed applications for developers and sysadmins. https://www.docker.com/, referred to 11.10.2015.

[121] mozilla.org. Firefox OS. https://www.mozilla.org/en-US/firefox/os/, referred to 21.10.2015.

Aalto-DD 40/2016

The transformation from telephony to mobile Internet has fundamentally changed the way we interact with the world by delivering ubiquitous Internet access and reasonable cost of connectivity. The mobile networks and Internet services are supportive of each other and together drive a fast development of new services and the whole ecosystem. As a result, the number of mobile subscribers has skyrocketed to a magnitude of billions, and the volume of mobile traffic has boomed up to a scale no-one has seen before with exponential growth predictions. However, the opportunities and problems are both rising. Therefore, to enable sustainable growth of the mobile Internet and continued mobile service adaption, this thesis proposes solutions to ensure that the reduction of overall environmental presence and the level of QoE are mutually addressed by providing energy-efficient data transmission to mobile devices.

ISBN 978-952-60-6685-1 (printed)
ISBN 978-952-60-6686-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

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

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS