Department of Communications and Networking

Techno-economic analysis of Internet content delivery

Nan Zhang

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 100/2016

Techno-economic analysis of Internet content delivery

Nan Zhang

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall S1 of the school on 22 June 2016 at 12.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Network Economics

Supervising professor
Prof. Heikki Hämmäinen

Thesis advisor
Ph.D. Timo Smura

Preliminary examiners
Prof. David Hausheer, Technische Universität Darmstadt, Germany
Prof. Jarmo Harju, Tampere University of Technology, Finland

Opponent
Prof. Peter Reichl, University of Vienna, Austria

Aalto University publication series
DOCTORAL DISSERTATIONS 100/2016

ISBN 978-952-60-6826-8 (printed)
ISBN 978-952-60-6827-5 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6827-5

Unigrafia Oy
Helsinki 2016

Finland

441 697
Printed matter

A! **Aalto University**

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

Abstract

Author
Nan Zhang

Name of the doctoral dissertation
Techno-economic analysis of Internet content delivery

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 100/2016

Field of research Network Economics

Manuscript submitted 8 October 2015 **Date of the defence** 22 June 2016

Permission to publish granted (date) 18 December 2015 **Language** English

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

Abstract

The Internet traffic volume, especially of heavy content, has increased fivefold in the past five years and is estimated to grow threefold in the next five years. At the same time, traffic volume originating from cellular data is growing three times faster than fixed IP traffic. Caching is utilized to efficiently cope with the growing traffic volume, and new in-network caching solutions enabled by, for example, information-centric networking (ICN) or software-defined networking (SDN) are being researched by the Internet community. For the in-network caching solution to be successfully adopted, it needs to efficiently solve the technical traffic optimization problem, but also meet the demand of the relevant market actors.

This work complements the technical development of in-network caching by analyzing its economic feasibility from the perspectives of the different market actors. First, this research analyzes the changing Internet content delivery ecosystem by identifying the relevant market actors and their respective relationships as well as constructs different scenarios for the future Internet content delivery market. The second part of the research quantitatively models the cost efficiency of SDN-enabled in-network caching compared with the current caching solutions.

Several methods and theories are utilized in the research process. Value network analysis is used to study the relationships between the relevant actors, whereas scenario planning is used to analyze alternative industry scenarios. The net benefit of SDN-enabled in-network caching is quantified with techno-economic modeling. In addition, data collection is based on interviews, brainstorming sessions and literature as well as network simulations.

The qualitative analysis shows that the success of content delivery networks (CDNs) is highly dependent on the existence of a two-sided platform, which induces cross-side network effects and provides quality guarantees, centralized billing and ease of use to the content providers. Distributed in-network caching solutions, such as ICN, need to solve the coordination problems in order to be widely deployed by the Internet service providers, mobile network operators and content providers. On the other hand, the SDN solution has a centralized controller platform, which potentially provides similar two-sided market benefits to the actors as the CDN solution. In addition, SDN-enabled in-network caching is shown to be cost efficient compared with current caching technologies in mobile networks.

Keywords Content delivery, Internet, mobile data, techno-economic analysis

ISBN (printed) 978-952-60-6826-8 **ISBN (pdf)** 978-952-60-6827-5

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

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

Pages 197 **urn** http://urn.fi/URN:ISBN:978-952-60-6827-5

Aalto-yliopisto

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tiivistelmä

Tekijä
Nan Zhang

Väitöskirjan nimi
Internetin sisällönjakelun teknis-taloudellinen analyysi

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Tietoliikenne- ja tietoverkkotekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 100/2016

Tutkimusala Tietoverkkotalous

Käsikirjoituksen pvm 08.10.2015 **Väitöspäivä** 22.06.2016

Julkaisuluvan myöntämispäivä 18.12.2015 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Internet-liikenne, erityisesti raskas sisältö, on viisinkertaistunut viimeisen viiden vuoden aikana ja sen odotetaan kolminkertaistuvan seuraavan viiden vuoden aikana. Samaan aikaan, mobiililiikenne on kasvanut kolme kertaa nopeammin kuin kiinteän verkon liikenne. Välimuistiteknologia on yksi ratkaisu kasvavasta liikennemäärästä selviytymiseen ja uusia verkkovälimuistiteknologioita (engl. in-network caching) on kehitteillä käyttäen, esimerkiksi, tietokeskeisiä verkkoja (engl. information-centric networking, ICN) tai ohjelmistopohjaisia verkkoja (engl. Software-defined networking, SDN). Onnistunut käyttöönotto edellyttää, että verkkovälimuistiteknologiat ratkaisevat liikenteen optimointiongelman ja vastaavat markkinatoimijoiden tarpeisiin.

Tämä väitöskirja täydentää verkkovälimuistiteknologioiden teknistä tutkimusta analysoimalla niiden taloudellista kannattavuutta eri markkinatoimijoiden näkökulmasta. Muuttuvaa sisällönjakelun ekosysteemiä analysoidaan tunnistamalla olennaiset markkinatoimijat ja heidän keskinäiset suhteensa, jonka jälkeen muodostetaan mahdollisia tulevaisuuden skenaarioita. Tutkimuksen toinen osa vertaa kvantitatiivisesti SDN-pohjaisen verkkovälimuistiratkaisun kustannustehokkuutta nykyisiin välimuistiratkaisuihin.

Työssä käytetään useita menetelmiä. Markkinatoimijoiden keskinäiset suhteet on selvitetty arvoverkkoanalyysin avulla ja skenaariosuunnittelumenetelmällä analysoidaan tulevaisuuden skenaarioita. SDN-verkkovälimuistiratkaisun nettohyötyä kvantifioidaan teknis-taloudellisen mallinnuksen avulla. Lisäksi tiedonkeruussa on hyödynnetty haastatteluja, aivoriihiä, kirjallisuutta ja verkon simulointia.

Kvalitatiivisen analyysin tulokset osoittavat, että sisällönjakeluverkkojen (engl. content delivery networks, CDN) menestys perustuu osittain kaksipuolisten markkinoiden verkkovaikutuksiin. Lisäksi sisällöntarjoajien näkökulmasta CDN tuo lisäarvoa tarjoamalla laatutakuuta, keskitettyä laskutusta ja käytön helppoutta. Hajautettujen verkkovälimuistiratkaisujen, kuten ICN:n, täytyy ratkaista nämä koordinaatio-ongelmat, jotta Internet-operaattorit, mobiilioperaattorit ja sisällöntarjoajat ottaisivat teknologian laajalti käyttöönsä. Toisaalta, SDN-ratkaisulla on keskitetty ohjainrakenne, joka mahdollistaa kaksipuolisten markkinoiden tuomat edut eri markkinatoimijoille. Lisäksi SDN-pohjainen verkkovälimuistiteknologia on todettu kustannustehokkaaksi verrattuna nykyisiin välimuistiteknologioihin mobiiliverkossa.

Avainsanat Sisällönjakelu, Internet, mobiilidata, teknis-taloudellinen analyysi

ISBN (painettu) 978-952-60-6826-8 **ISBN (pdf)** 978-952-60-6827-5

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

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2016

Sivumäärä 197 **urn** http://urn.fi/URN:ISBN:978-952-60-6827-5

Preface

"In life, it's not where you go –
It's who you travel with."

– Charles M. Schulz

This thesis would not have been possible without the support and contribution of many amazing people. First and foremost, I would like to express my gratitude to Professor Heikki Hämmäinen for supervising my thesis and providing the necessary resources for finalizing it. Heikki's ability to see the bigger picture and to think out of the box has greatly contributed to finding the right path both for the thesis and for my future career. I would also like to thank Timo Smura for instructing my thesis. Timo was a perfect balance to Heikki with his down to earth advices and practical help, especially in the final phases of the thesis process.

I have had the honor to work and co-author the research papers with several excellent researchers. Firstly, I would like to thank Tapio Levä, who also instructed my M.Sc. thesis, for all the support at the beginning of my Ph.D. journey and for showing me the little tricks that made the research process less intimidating. My gratitude also goes to Jose Costa-Requena, whose ability to juggle endless numbers of projects with a broad smile is an inspiration for everyone around him; let's continue the good collaboration. I am also grateful to Hannu Flinck, Björn Grönvall and Maël Kimmerlin for the enlightening and interesting technical discussions on information-centric networking and software-defined networking. Furthermore, I want to thank Ioanna Papafili, Jörn Künsemöller, João Soares, Kimmo Berg, Manos Dramitinos and Miroslav Kantor for the fruitful collaborations.

The Netbizz team: Henna Suomi, Antti Riikonen, Tapio Soikkeli, Arturo Basaure, Benjamin Finley, Michael Katsigiannis, Alexandr Vesselkov, Jaume Benseny, Juuso Karikoski, Thomas Casey, Kalevi Kilkki, Pekka Kekolahti and Juuso Töyli. Thank you all for making the process so much more enjoyable with the interesting lunch discussions and delicious BBQs. I would also like to thank the entire Comnet staff for providing such an inspiring working environment.

My gratitude also goes to the collegues at SAIL and SIGMONA projects for providing such a multi-cultural and expert network, which has contributed extensively to my thesis. A special thank you goes to Jukka Salo, who has been a constant friendly face since my first project meeting until the very last moments of my Ph.D. journey. In addition, I want to thank all the interviewed experts throughout the years for all their valuable insights and input.

I am grateful for Professor Jarmo Harju and Professor David Hausheer for the detailed pre-examination comments, which helped in imporoving the quality of my thesis. I would also like to thank Professor Peter Reichl for the honor of accepting to be my opponent.

The Electrical Engineering School at Aalto Univeristy has been my home throughout the B.Sc., M.Sc. and Ph.D. degrees. I am honoured to have been part of it, especially the ELEC doctoral school. Furthermore, I want to thank Tekniikan Edistämissäätiö and the Nokia Foundation for the financial support during the research. I am also grateful to the Future Internet Graduate School and EIT Digital Doctoral School, through which I met many interesting and talented young researchers. A special thanks goes to the doctoral students at EIT Digital for making the journey so much fun.

Last but not least, I wish to thank my family and friends for their support and understanding during the research process. Especially, I want to thank Ram Sankar for designing the cover of the thesis. Mom and Dad, I am inspired by your courage and proud of your achievements, despite the unfavorable starting point. Without your hard work, I would not be where I am. Finally, thank you Tomas for believing in me and encouraging me to stay true to myself.

Helsinki, May 8th, 2016

Nan Zhang

Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
4G	Fourth generation (mobile telecommunications technology)
5G	Fifth generation (mobile telecommunications technology)
CAPEX	Capital expenditure
CCN	Content-centric networking
CDN	Content delivery network
CP	Content provider
DB	Database
DONA	Data oriented network architecture
DPI	Deep packet inspection
eNB	Evolved NodeB
EPC	Evolved packet core
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved universal terrestrial radio access
FE	Front-end
GPRS	General packet radio service
GSM	Global system for mobile
GTP	GPRS tunneling protocol
HSS	Home subscriber server
HTTP	Hypertext transfer protocol
HTTPS	Hypertext transfer protocol secure
ICN	Information-centric networking
ICT	Information and communications technology
IDPS	Intrusion detection and prevention system
IETF	Internet Engineering Task Force

IP	Internet protocol
ISP	Internet service provider
LTE	Long-term evolution
MME	Mobility management entity
MNO	Mobile network operator
MVNO	Mobile virtual network operator
NAT	Network address translation
NDN	Named data networking
NetInf	Networking of information
NFV	Network function virtualization
OPEX	Operational expenditure
P2P	Peer-to-peer network
PCRF	Policy and charging rules function
Pub/sub	Publish/subscribe paradigm
QoE	Quality of experience
QoS	Quality of service
SDN	Software-defined networking
SDMN	Software-defined mobile network
SFC	Service function chaining
S/P-GW	Serving/packet data network gateway
TCP	Transmission control protocol
TV	Television
VoD	Video-on-demand
*-C	Control plane functions
*-U	User plane elements

List of figures

List of tables

List of publications

This doctoral dissertation consists of a summary and of the following publications, which are referred to in the text by their Roman numerals.

I. N. Zhang, T. Levä and H. Hämmäinen, "Value Networks and Two-Sided Markets of Internet Content Delivery," Telecommunications Policy, vol. 38, no. 5-6, pp. 460-472, 2014.

II. N. Zhang, T. Smura, B. Grönvall and H. Hämmäinen, "Scenario Analysis for Commercial Internet Content Delivery," Info: The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media, vol. 16, no. 3, pp. 54-71, 2014.

III. N. Zhang, T. Levä, H. Hämmäinen, "SDMN: Industry Architecture Evolution Paths," in Liyanage, M., Gurtov, A. and Ylianttila, M. (eds.), Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture, Wiley Publishers, 2015.

IV. M. Dramitinos, N. Zhang, M. Kantor, J. Costa-Requena and I. Papafili, "Video Delivery over Next Generation Cellular Networks," in Proceedings of 9th International Conference on Network and Service Management, Zurich, Switzerland, October 18, 2013.

V. N. Zhang, H. Hämmäinen, Heikki and H. Flinck, "Cost Efficiency of SDN-enabled Service Function Chaining," Submitted to a journal.

VI. N. Zhang and H. Hämmäinen, "Cost Efficiency of SDN in LTE-based Mobile Networks: Case Finland," in: Proceedings of Workshop on Software-Defined Networking and Network Function Virtualization for Flexible Network Management, Cottbus, Germany, March 12, 2015.

VII. N. Zhang, M. Kimmerlin, J. Costa-Requena and H. Hämmäinen, "Cost Efficiency of Mobile in-Network Caching," International Journal on Network Management Special Issue on SDN and NFV for Flexible Network Management, vol. 26, no. 1, pp. 44-55, 2016.

Author's contribution

I. The idea for this article was formed by Zhang and Levä. The interviews were conducted by Zhang, while Levä assisted in improving the analysis and Hämmäinen provided comments. Zhang wrote the whole article, while iteratively reviewing and editing the article with Levä and Hämmäinen.

II. The idea for this article was formed by Zhang and Hämmäinen. Zhang organized and facilitated the workshops as well as wrote the bulk of the text. The scenarios were constructed jointly by Zhang and Hämmäinen. Smura helped in improving the quality of the method and analysis descriptions (Section 2). Grönvall contributed to the technical implications of the article (Section 5.2). Zhang, Smura and Hämmäinen reviewed and edited the paper iteratively.

III. The idea for this article was formed by Zhang and Hämmäinen. The interviews were conducted and the whole article was written by Zhang. Levä assisted in improving the analysis. Zhang, Levä and Hämmäinen reviewed and edited the article iteratively.

IV. The idea for this article was formed by Zhang and Papafili. Zhang contributed to defining the scenario and the two use cases, identified the current market trends, described the industry architecture notation and provided the template for the notation (Sections I, IV.B and V.A). The main business model analysis was done by Dramitinos. Kantor and Costa-Requena provided the technical knowledge of LTE and SDN. Papafili reviewed and edited the article.

V. The idea for this article was formed by Zhang and Hämmäinen. The interviews were conducted and the bulk of the text was written by Zhang. Hämmäinen assisted in forming the cost analysis and reviewed the article. Flinck provided the technical knowledge of SDN and service function chaining and wrote Section 3.

VI. The idea for this article was formed by Zhang and Hämmäinen. The interviews were conducted and the whole article was written by Zhang. Hämmäinen assisted in forming the cost model. Zhang and Hämmäinen reviewed and edited the article iteratively.

VII. The idea for this article was formed by Zhang and Costa-Requena. Zhang and Hämmäinen contributed to forming the cost analysis and Zhang wrote the bulk of the text. Kimmerlin constructed the performance simulation and wrote Section 4. Costa-Requena provided the technical knowledge of SDN and caching. Zhang, Kimmerlin, Hämmäinen and Costa-Requena reviewed and edited the article together.

1 . Introduction

1.1 Background and motivation

The amount of heavy content, especially video, is increasing in the Internet. According to Cisco (2015), 80% of all consumer Internet traffic will be video in 2019. Global IP traffic is estimated to grow at a compound annual rate of 23% between 2014 and 2019. At the same time, mobile data volume is estimated to grow 61% annually between 2014 and 2019. As a consequence, both Internet service providers (ISPs) and mobile network operators (MNOs) face heavy investments in network capacity. One solution to efficiently cope with the increasing traffic volume is to employ caching, which enhances the end-user perceived quality of experience (QoE), improves network scalability and reduces the traffic volume leaving the ISP or MNO networks. Other prominent traffic optimization solutions include multipath technologies (Wischik et al., 2008) that typically require the involvement of the end users and are, thus, outside the scope of this research.

Caching first emerged in the form of web caching (Barish and Obraczka, 2000), which temporarily stores web pages in web proxies to reduce load on the origin servers and to improve the response time. Web caching preserves the client-server model, whereas in peer-to-peer (P2P) networks (Schollmeier, 2002), established for file sharing, each peer acts both as a client and a cache server. The commercial content providers (CPs) are increasingly outsourcing their content delivery to third-party content delivery networks (CDNs) (Dilley et al., 2002), which place content servers in different ISP or MNO networks to better serve the end users. For example, already in 2014, 57% of all Internet video traffic passed through CDNs (Cisco, 2015). Recent developments (e.g. Krishnan et al., 2000; Chen et al., 2002; Xu et al., 2002; Tang and Chanson, 2002; Jia et al., 2003; Sourlas et al., 2011) focus on cache placement at different parts of the network, such as router and access networks, which is called in-network caching. In-network caching can be enabled by different technologies, such as information-centric networking (ICN) (Jacobson et al., 2012) or software-defined networking (SDN) (Raghavan et al., 2012)

Technically, ICN (Jacobson et al., 2012) introduces flexible routing by content names instead of locations. In addition, the economic incentives of caching with ICN are discussed in Agyapong and Sirbu (2012) and Wang et al. (2014). However, clean-slate ICN implementations require changes in the network topology and the routing principles, and the overlay implementations complicate the network structure as well as add overhead to the network (Ahlgren et al., 2012). At the same time, ICN caching is limited to fixed networks and provides minor benefits in mobile networks, because mobile data is typically enclosed in the GPRS tunneling protocol (GTP) within the mobile network.

Increasing digitalization, on the other hand, shifts the network equipment market from hardware business to software business in a way similar to the handheld device and computer markets. For example, network function virtualization (NFV) and SDN in the context of mobile networking have been widely discussed in recent literature (e.g. Aleksic and Miladinovic, 2014; Salsano et al., 2014; Taleb, 2014; Haleplidis et al., 2015a; Taleb et al., 2015).

Thus, SDN has been proposed as a solution to enable in-network caching in a mobile environment by removing the GTP tunneling from the mobile networks (Costa-Requena, 2014). SDN benefits arise from the decoupling of the control and user planes (Raghavan et al., 2012), and include increased flexibility and dynamicity of the network (Kreutz et al., 2015), as well as faster service and network software update cycles to the ISPs and MNOs (Pentikousis et al., 2013). In addition, SDN promotes new business models, such as virtual CDNs (Veitch et al., 2015), and reduces cost through dynamic service chains (Quinn and Guichard, 2014), where less popular content can bypass the caches to prevent cache throttling. On the other hand, SDN increases signaling traffic in the network as well as adds new points of failure by centralizing the control plane into the data centers. Thus, the net benefits of both ICN and SDN should be quantified.

1.2 Research questions and objectives

The main research question of this dissertation is as follows:

Which Internet content delivery technologies are techno-economically the most feasible?

To answer the question, the Internet content delivery market dynamics and ecosystem should be mapped. In addition, the researched technologies should be compared with the incumbent caching technologies, such as CDNs and other data center caching solutions. Thus, the main objectives of the research are:

1) to identify the key roles and actors for Internet content delivery market and their respective business relationships,
2) to form feasible industry scenarios by identifying the main trends and uncertainties of the Internet content delivery ecosystem, and
3) to model the cost efficiency of the new technologies compared with the current technologies.

1.3 Research scope and definitions

The analysis on the existing Internet content delivery solutions limit to data center caching (e.g. such as CDNs and ISP data center caching) and caching in end-user devices (e.g. P2P) due to their dominant role in the current Internet content delivery market. In addition, in-network caching is discussed by using ICN and

SDN technologies as example cases. The analysis on ICN focuses on caching in fixed IP networks, whereas SDN-enabled caching is utilized in the backhaul of mobile networks, i.e. long-term evolution (LTE) networks in this work.

The scope of this research is limited to human generated content and does not discuss machine-to-machine traffic, because caching benefits are seen mostly with heavy content, such as video. In addition, due to the availability of data, the quantitative analysis mostly focuses on the Finnish market.

The actors that are considered in this research and are important for the Internet content delivery market are defined as follows:

Content provider: Operates a platform that aggregates commercially produced content, such as Netflix, or user-generated content, such as YouTube. Has direct relationship with the end users.

End user: Requests, consumes and potentially pays for the content.

Internet service provider: An actor providing Internet connectivity to both end users and content providers. Can be divided into Internet access providers (access-ISP) and transit providers (transit-ISP).

Mobile network operator: Provides Internet access via the cellular network, i.e. LTE network in this research.

CDN provider: Operates CDN services and interacts with both content providers and ISPs.

1.4 Outline of the thesis

For an efficient presentation of the findings, the structure is divided into two logical parts. The first part qualitatively analyzes the Internet content delivery market from the perspectives of the main actors and forms alternative industry scenarios (Objectives 1 and 2). A quantitative cost modeling approach is taken in the second part of the thesis to analyze the cost efficiency of SDN-enabled in-network caching compared with current caching alternatives (Objective 3). The structure of the thesis is illustrated in Figure 1.

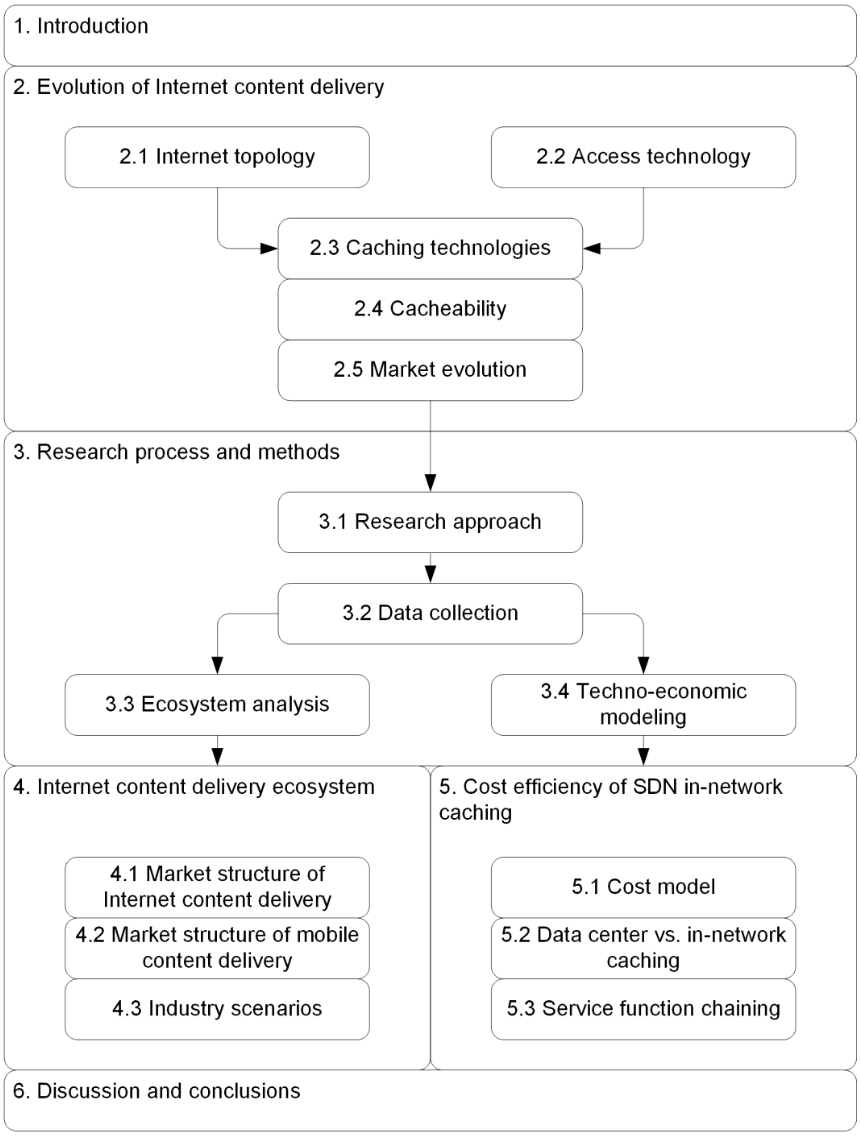

Figure 1. Structure of the thesis.

2 . Evolution of Internet content delivery

This research studies techno-economic efficiency of different caching technologies in the Internet content delivery market. The following subsections highlight the recent evolution trends from the perspectives of the Internet topology, the market and the different technologies.

2.1 Internet topology

The internet topology is changing due to the increase in traffic volumes and emerging market forces. Figure 2 shows the traditional Internet logical topology, where few transit-ISPs form the backbone that provide interconnectivity to lower tier ISPs (Labovitz et al., 2010). As content's importance and traffic volume increase, large content providers such as Google and Netflix are forming their own global networks that run parallel to the traditional backbone networks. In addition, content sources are consolidating and traffic is increasingly flowing directly between large content providers, content delivery networks and end-user networks (Labovitz et al., 2010). Figure 3 illustrates the emerging Internet logical topology.

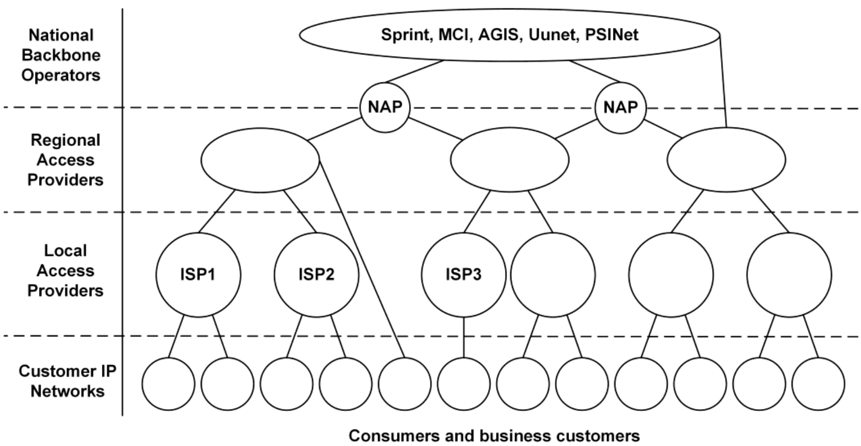

Figure 2. Traditional Internet logical topology (Labovitz, Iekel-Johnson, McPherson, Oberheide and Jahanian, 2010).

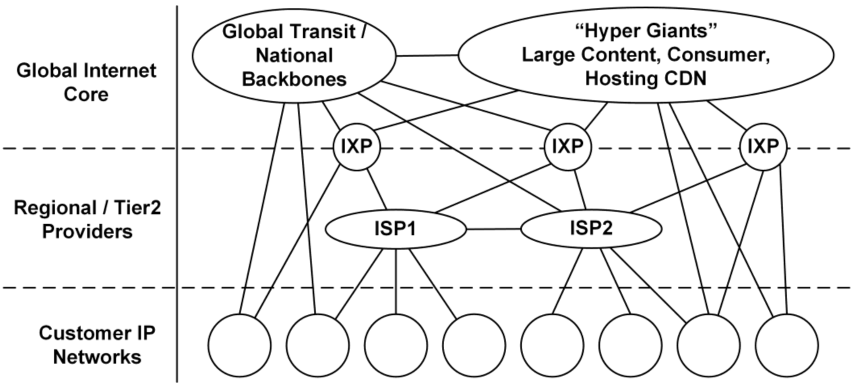

Figure 3. Emerging Internet logical topology (Labovitz, Iekel-Johnson, McPherson, Oberheide and Jahanian, 2010).

For full Internet connectivity, lower tier ISPs can either buy transit from the transit-ISPs or peer with each other through peering agreements (Norton, 2011). When an ISP buys transit, it can access the whole Internet through the same transit provider and the transit traffic is typically charged based on usage by using the 95th percentile measurement method[1]. In addition, the transit agreement may contain details on service levels and bigger transit buyers may get volume discounts by committing to certain levels of transit volume usage.

Peering agreements, on the other hand, have traditionally been settlement free and the peering ISPs gain access only to each other's networks rather than the whole Internet. However, as Internet traffic volumes grow, peering is becoming more popular, especially among ISPs of the same tier and size (Labovitz et al., 2010). On the other hand, if the peering ISPs do not attain equal value from the agreement, paid peering can be adopted (Norton, 2011). For example, some ISPs see more download traffic (i.e. eyeball-ISP), while others have more upload traffic (i.e. content-ISP), which makes the peering relationship asymmetric.

In addition to transit and peering agreements, content providers may offer side payments to ISPs that are not directly connected to the content providers, but whose networks are heavily flooded with the content provider's traffic. The side payments are commonly paid, despite violating the net neutrality principles, to ensure the service quality and potentially to increase the entry barriers for smaller competiting content providers (Tuffin, 2015).

2.2 Access technologies

Traditionally, heavy content such as video is consumed over fixed-access networks. However, with the emergence of 4G and 5G technologies, the mobile network bandwidths are more equivalent to that of the fixed networks. As a consequence, mobile data traffic is estimated to grow three times faster than fixed IP

[1] Every five minutes, the load on the link is measured. The top 5% of data samples taken during the billing period are ignored. The next highest measurement, i.e. 95th percentile, becomes the billable rate (Norton, 2011).

traffic between 2014 and 2019 (Cisco, 2015). Figure 4 shows the evolution of 3GPP mobile access technologies: from circuit switched GSM to first packet switched networks and to evolved packet networks (Nohrborg, 2015).

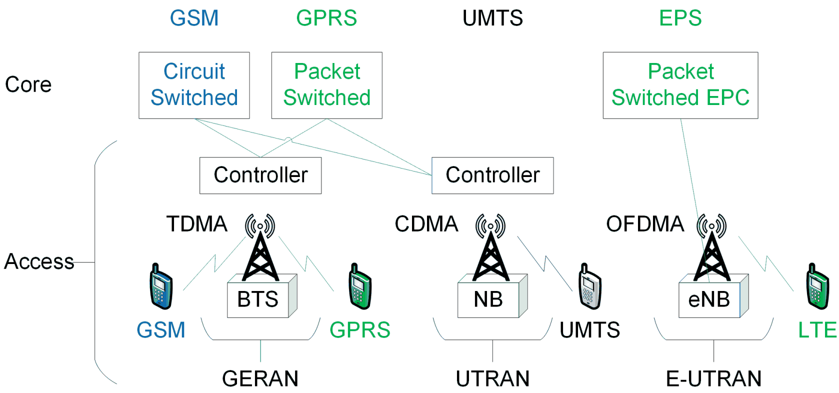

Figure 4. Evolution of 3GPP mobile access technologies (Nohrborg, 2015).

Figure 5 shows the LTE architecture specified by 3GPP (3GPP, 2015) with its main elements and interfaces, specifically the evolved universal terrestrial radio access network (E-UTRAN) and the evolved packet core (EPC). The end-user traffic is sent to the LTE radio base station, called evolved NodeB (eNB), through the radio interface. The eNB asks for the subscriber and authentication information from the mobility management entity (MME) and home subscriber server (HSS). The traffic is then forwarded by the network routers and switches to the serving/packet data network gateway (S/P-GW), where routing decisions are made. This research takes a look at an architectural variant, where the S/P-GWs are physically collocated, as is shown in Figure 5. P-GW is attached to the external interfaces, through which the traffic leaves the mobile core network and enters the public IP network. Additionally, P-GW acts as a firewall and S-GW serves as the mobility anchor during eNB handovers. The policy and charging rules function (PCRF) keeps track of the network usage and handles the billing for each account. In addition, traditional mobility management typically uses the GPRS tunneling protocol (GTP) to ensure end-user mobility and seamless connectivity by creating an end-to-end tunnel between the radio network (eNB) and the core network (P-GW). GTP mainly has two functions: 1) manages user session information and adjusts quality of service, and 2) encapsulates and transports the user data packets through the mobile network between the eNBs and P-GW.

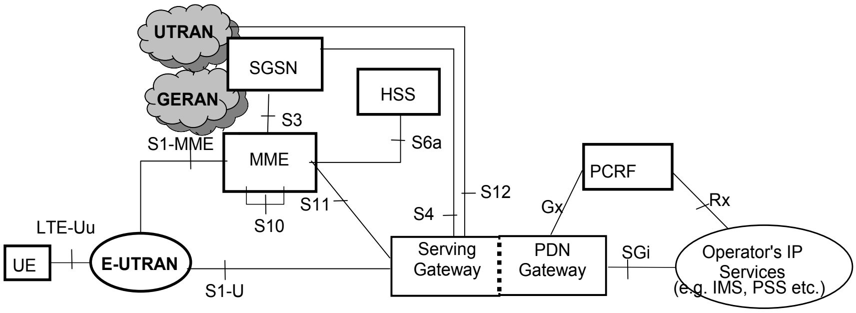

Figure 5. LTE architecture with EPC and E-UTRAN as specified by 3GPP (3GPP, 2015).

2.3 Caching technologies

Web caching (Barish and Obraczka, 2000) is one of the first caching solutions that improves the web browsing (i.e. HTTP traffic) experience by temporarily storing popular web sites in the web browser caches or proxy servers near the end users. A taxonomy of web caching schemes exist, such as proxy caching (Fielding et al., 2014), adaptive web caching (Michel et al., 1998) and push caching (Gwertzman and Seltzer, 1995). In addition, a variety of web caching protocols have been designed, such as the Internet cache protocol (Wessels and Claffy, 1997) or web cache control protocol (Cooper et al., 2001). As the heavy and dynamic content volume continues to grow, new caching solutions that are not limited to HTTP traffic are needed. Several of the new caching solutions, such as content delivery networks (CDNs) and ISP data center caching, utilizes the web caching protocols defined in RFC 3040 (Cooper et al., 2001).

The currently dominant caching technologies include CDNs (Dilley et al., 2002), ISP data center caching and in certain use cases also the peer-to-peer (P2P) networks (Camarillo, 2009; Schollmeier, 2002). Recent developments in caching technologies include different in-network caching solutions enabled by, for example, information-centric networking (ICN) (Jacobson et al., 2012) or software-defined networking (SDN) (Raghavan et al., 2012). The relevant caching solutions are listed in Table 1, together with the descriptions of their basic characteristics.

Table 1. Characteristics of alternative Internet content delivery technologies[2].

Architecture	CDN	ISP data center caching	P2P	ICN	SDN
Cache location	Server-side nodes	Server-side nodes	Client-side nodes	In-network nodes	In-network nodes
Routing decision	By host names	By host names	By host names	By content names	By host names
Level of standardization	Medium	Medium	Low	High	Medium[3]
Content awareness	Medium	Medium	Low	Medium	High

[2] Adapted from Publication II and modified. Reprinted with permission from Emerald Group Publishing Limited.

[3] At the time of writing, SDN's standardization efforts are still ongoing.

2.3.1 Content delivery networks

The CDN model is an overlay to the basic Internet (Vakali and Pallis, 2003), which divides the end-to-end connection into two separate connections: 1) between the content provider and the CDN provider, and 2) between the CDN provider and the end users. The CDN elements communicate with standard interaction protocols defined in RFC 3040 (Cooper et al., 2001), such as the web cache control protocol, cache array routing protocol and hypertext caching protocol (Pathan and Buyya, 2008). On the contrary to web caches, where only highly requested content is temporarily stored, CDN data centers store content specified by the network administrators.

The CDN data centers are geographically distributed and typically operated by three types of CDN providers: 1) The pure-play CDNs, such as Akamai (Dilley et al., 2002), whose content servers are either located in their own data centers or collocated within the ISP or MNO data centers (Frank et al., 2013), 2) Licensed CDNs (Edgecast, 2015), where an ISP or MNO builds its own CDN service and sells it directly to content providers, and 3) content provider CDNs, where the content provider builds its own CDN that could be collocated in the ISP data centers (Netflix, 2015). In addition, CDNs can form federations among themselves through, e.g., the CDN interconnection initiative (Peterson and Davie, 2014).

In addition to lowering latency and reducing origin server load, content providers also benefit from a caching agreement with the CDN providers. The content providers can choose either to cache the content when first requested or push certain content (e.g. heavy video) beforehand to the CDN servers during off peak hours. The caching agreement also provides content access control and usage statistics and enables caching of encrypted traffic.

2.3.2 ISP data center caching

ISP data center caching can be considered as a subset of CDNs, where the same interaction protocols from RFC 3040 (Cooper et al., 2001) are used, but content is cached in ISP or MNO data centers. However, ISP data center caching can also be done without a contract with the content provider and the ISP or MNO can choose what content to cache. In the non-contracted case, encrypted traffic may cause the cacheability of content to drop, as explained in Section 2.4.

2.3.3 Peer-to-peer networks

P2P networks represent another type of overlay to the basic Internet (Camarillo, 2009; Schollmeier, 2002), which consist of distributed content storages (i.e. end-user devices) connected by the network. In the P2P network, any end user can act as both the client requesting content and the server serving a cached copy of the requested content. P2P communication protocols are less standardized and different networks use different technologies (Risson and Moors, 2007). For example, the resource location and discovery base protocol is designed to support the P2P session initiation protocol (Jennings et al., 2014), Bit-

Torrent (Cohen, 2008) and Gnutella (Klingberg and Manfredi, 2002) are utilized for file sharing among end users, and distributed application networks use structured distributed hash table based protocols such as Chord (Stoica et al., 2003) and Pastry (Rowstron and Druschel, 2001).

A P2P network can be either controlled, where the end users are authenticated and authorized, or uncontrolled, where all users can access the network resources. In this research, the scope is limited to content provider controlled P2P networks, where the end users authenticate themselves with the content provider before accessing the content. In controlled P2P networks, the content exists permanently in the content provider's servers and are cached temporarily in the end-user devices, which is similar to a hybrid CDN-P2P solution (Haßlinger and Hartleb, 2011).

2.3.4 In-network caching

In-network caching in this research refers to caching anywhere in the network, e.g. from base stations through routers/switches to gateways in data centers. Literature on in-network caching (e.g. Katsaros et al., 2011; Psaras et al., 2012) typically exclude data center caching, but data center caching is a dominant technology in the current Internet content delivery market and not likely to be removed due to the addition of caches into the network. Two technologies that enable the placement of caches into the network are ICN and SDN.

Information-centric networking

A prominent in-network caching technology is ICN or content-centric networking (CCN). The basic idea of ICN is that the network routers and nodes are equipped with cache storage that cache content (Ahlgren et al. 2012) and the content moves freely in the network. Routing in the ICN concept is based on content names instead of content locations (Jacobson et al., 2012). In addition, the naming scheme in ICN is different to the existing content naming and offers, for example, data integrity, owner continuity and owner identification (Ahlgren et al., 2012). In ICN, when a content traverses from the origin server to the end user that requested the content, the network elements in between cache the content. If other end users request the same content, any network elements that sees the request may reply with the content. As a consequence, the importance and load of origin servers diminish.

Several ICN implementations exist: for example, data oriented network architecture (DONA) (Koponen et al., 2007), named data networking (NDN) (Zhang et al., 2010) with its CCNx (Mosko et al., 2016) and CCN-lite (Beck et al., 2015) protocols, the publish/subscribe (pub/sub) paradigm (Fotiou, Trossen and Polyzos, 2012) prototypes Blackadde and Blackhawk (Pursuit, 2013), and networking of information's (NetInf) (Dannewitz et al., 2013) OpenNetInf prototype (Dannewitz and Herlich, 2011). Though the different ICN implementations share similar assumptions and architectural properties, they differ in details. For example, unique naming is a key component of all ICN implementations, but the pub/sub is more dependent on lower level host addresses due to the rendezvous routing scheme, which matches the content providers with the content requesters (Fotiou et al., 2012). Ahlgren et al. (2012) gives a more detailed comparison of the functionalities of DONA, NDN, pub/sub and NetInf.

Within ICN research, several caching architectures have been proposed. For example, ProbCache (Psaras, Chai and Pavlou, 2012) follows a probabilistic algorithm for caching in fixed-access networks and reduces origin server hits by 15% compared with the ICN caching proposed by Jacobson et al. (2012). Katsaros, Xylomenos and Polyzos (2011) introduces a MultiCache overlay caching system to the fixed-access Internet, which shows an inter-domain traffic load reduction of 53% and inter-domain traffic load reduction of 61% compared with BitTorrent P2P networks. In addition, Wang et al. (2014) compares ICN caching in the EPC data centers and base stations of LTE networks to no caching (80% load reduction), to caching only in EPC without ICN (65% load reduction), and to caching in both EPC and base stations without ICN (20% load reduction). Other ICN cache management schemes include Hu et al.'s (2015) proposal that considers cache placement, replacement and location, an in-network storage capacity coordinator that optimizes the overall network performance and cost (Li et al., 2015), and a distributed autonomic management architecture for content replication (Sourlas et al., 2013).

The ICN model discussed in this research does not reflect any single ICN implementation, but is based on the common conceptual assumptions of ICN. Similarly to the P2P network, caching in the ICN model can be either controlled or uncontrolled. In this research, ICN in-network caching is assumed to be ISP controlled technical caching, where no caching agreements exist between the content provider and the ISP. In addition, ICN caching works best in a fixed IP network setting, where mobility and GTP tunnels do not raise issues. The ICN solutions for caching in mobile networks typically consider only the two ends of the GTP tunnel: the data center and the base stations. In-network caching that covers the whole mobile network is presented next.

Software-defined networking

The above caching technologies could also be used in mobile networks. However, mobile networks are designed to support end-user mobility with GTP, which creates an end-to-end tunnel between the base station and the gateway that makes caching on the path difficult. Currently, caches are located at either end of the tunnel: in the base stations or in the EPC data centers, as is shown in the upper part of Figure 6. The lower part of Figure 6 illustrates an in-network caching solution proposed by Costa-Requena et al. (2014), where SDN is utilized to remove GTP tunneling from the mobile networks (Costa-Requena, 2014) while ensuring mobility to the end users. In addition, the SDN-enabled in-network caching solution dynamically optimizes resources by moving the cached content to more optimal locations based on end-user demand.

SDN (Kreutz et al., 2015) is a networking concept that decouples the network element's control plane functions from its user plane elements[4] (Raghavan et

[4] The user plane still maintains some software functions and is not purely hardware. However, to distinguish from the control plane functions, the user plane functions are grouped under the user plane elements in this work.

al., 2012), which leads to network programmability (Haleplidis et al., 2015c). Thus, in a SDN network, the network management can directly program, control and orchestrate the network resources (ITU-T, 2014). In LTE networks, SDN leads to improved flexibility and agility (Kreutz et al., 2015), as well as faster service roll-out cycles (Pentikousis et al., 2013) and potential for cost reduction (Naudts et al., 2012; Knoll, 2015; Zhang and Hämmäinen, 2015). In addition, SDN can be used for traffic steering, automation and forming dynamic service function (e.g. firewall, caching and deep packet inspection) chains (Quinn and Guichard, 2014; Blendin et al., 2014). The signaling between the control and user planes (Kreutz et al., 2015) can be handled by, for example, the OpenFlow protocol (McKeown et al., 2008; Tourrilhes et al., 2014), the forwarding and control element separation protocol (Doria et al., 2010; Haleplidis et al., 2015b), and the network configuration protocol (Enns et al., 2011).

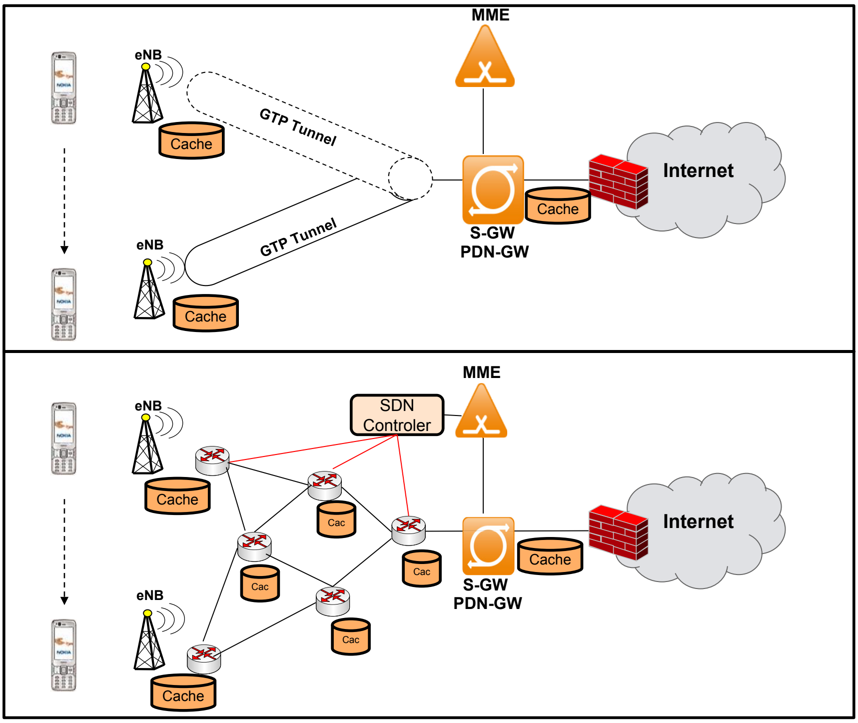

Figure 6. Caching in LTE with and without SDN (Costa-Requena, 2014).

Several SDN-enabled mobile network architectures have been proposed, especially for the core and backhaul networks. For example, Nguyen and Kim (2015) proposes an OpenFlow EPC architecture that introduces a mobile controller for managing the signaling between different network elements. Basta et al. (2014) measures the performance of virtualized S/P-GW, SDN-optimized S/P-GW and a combination of both virtualized and SDN-optimized S/P-GW, the results of which show that with limited data center capacity, the combined deployment achieves the least delay and network load overhead. Costa-Requena et al. (2015c) uses SDN to optimize the backhaul transport by replacing mobile specific tunneling with MPLS or Carrier Grade Ethernet, which has been tested in an ETSI proof-of-concept (ETSI, 2015) together with a virtualized EPC. In addition, several full mobile network architectures based on SDN and virtualization principles have been proposed, for example, by Giraldo et al. (2014), Wang et al. (2015) and Costa-Requena et al. (2015b).

Centralization and virtualization are often mentioned together with SDN. Centralization benefits are assumed to be fully utilized in the current LTE networks, e.g. EPC elements are already running in centralized data centers. Virtualization happens only when the centralized elements no longer run on dedicated hardware, but rather use general purpose hardware. In addition, NFV further increases the flexibility of the data centers, because the control plane functions are re-optimized by dividing them into functional groups (Open Networking Foundation, 2014).

2.4 Cacheability

An important factor of caching is the cacheability of content. Caching can be purely technical, i.e. transparent caching, where MNOs decide the caching location to optimize its own network resources; or contracted, where content providers buy caching services from the caching provider. Literature on caching are mainly on technical transparent caching and technical cacheability, which mostly consider the technical attributes of content. For example, Chi et al. (2006) and Ramanan et al. (2013) deem a content non-cacheable if the cache control field of an HTTP header is set to no-cache, the content length is set to zero or the content has expired. The results of their quantitative studies show that 79% of all monitored objects are cacheable (Chi et al., 2006) and 73% of all traced HTTP traffic volume is cacheable.

In addition, encrypted traffic, such as HTTPS, and user-generated content, such as YouTube, set certain limitations on technical caching. For example, Naylor et al. (2014) report a cache hit ratio drop from 16.8% to 13.2% between 2012 and 2014 in a transparent in-network caching network due to the increased usage of encrypted traffic and personalization of content. Ager et al. (2010) study the cacheability of user-generated content, the results of which show a cacheability of 71% from all HTTP requests and only 28% from the total HTTP traffic volume. Studies on the caching benefits of YouTube show cache hit rates of 25%-35% (Zink, Suh, Gu and Kurose, 2009) and that 40% of all requests can be served from caches (Braun et al., 2012).

Elkadi (2012) identifies several factors that influence the cacheability of content from also the economic and contractual perspectives. For example, economic cacheability is constrained by the popularity of the content and cost of caching, and contractual cacheability can be influenced by the content provider's control on distribution or service level. On the other hand, contracted caching may lift the caching constraints set by the content providers to allow their caching partners to cache all technically cacheable content. For example, Google (Google, 2015) does not allow third-party caching of their content by setting the HTTP response headers to private, but a caching agreement could make the content cacheable by Google's partners.

2.5 Market evolution

Together with the technological change comes the changes in market dynamics and end-user behavior in terms of what content is consumed, how content is accessed and what are the payment preferences. For example, the content providers' revenue model is shifting from free content to freemium and paid content, where the freemium model offers the basic service for free and charges for additional features or better service quality (Pulkkanen and Seppänen, 2012). The increased popularity of Netflix with its 60 million global subscribers (Hastings and Wells, 2015) and Spotify with its paying-to-free end-subscriber ratio growing from 15% in 2012 (Wagner et al., 2013) to 20% in 2015 (Spotify, 2015) demonstrate this change. However, the nature of digitalized content, being easily replicable, durable and the value of the content getting realized only after consumption, tend to lead to the free-rider problem (Rayna, 2008). Thus, proper control mechanisms, such as encryption and free previews of only parts of the content, are needed for the success of paid content.

The type of content consumed is also changing. For example, 80% of global consumer traffic is estimated to be video by 2019 (Cisco, 2015) and the popularity of user-generated content is increasing (Chaet al., 2007). In addition, despite globally available content, much of the video content, especially user-generated, are popular only within certain geographical areas. For example, Brodersen et al. (2012) observe that 50% of YouTube videos have over 70% of their views in a single region.

The viewing behavior of the end users depends on the type of content. For example, shorter videos, such as user-generated content, news clips or short entertainment clips, are viewed twice as much during weekdays than during weekends (Miranda et al., 2013). On the other hand, the viewing of TV programs, movies and other longer videos peaks during Fridays, Saturdays and Sundays (Abrahamsson and Nordmark, 2012). In addition, the shorter videos tend to have most of its views within the first ten days since its publication (Miranda et al., 2013), whereas the popularity of movies and TV programs declines more slowly within several weeks (Abrahamsson and Nordmark, 2012).

3 . Research process and methods

This research is future-oriented by nature, as the investigated technologies are not widely deployed in the current Internet and have gained attention only during the past few years. Thus, a combination of different research methods are utilized. The following sub-sections first give a general view on the research approaches adopted and then describe the main theories and methods used in each of the three research phases: 1) data collection, 2) ecosystem analysis and 3) techno-economic modeling.

3.1 Research approach

This research is also multi-disciplinary by nature as it combines theories, models and methods from the engineering and economics fields. The research is technology-oriented with the main focus on the economic evaluation of Internet content delivery technologies, which falls under the design science activities (i.e. build and evaluate) in information technology research as defined by March and Smith (1995). Similarly, Simon (1996, first published in 1969) defines design sciences as creating innovations that serve human purposes. Järvinen (2004) expands the research of March and Smith (1995) into a taxonomy of six research approaches, as is shown in Figure 7, where the approaches used in this research are highlighted.

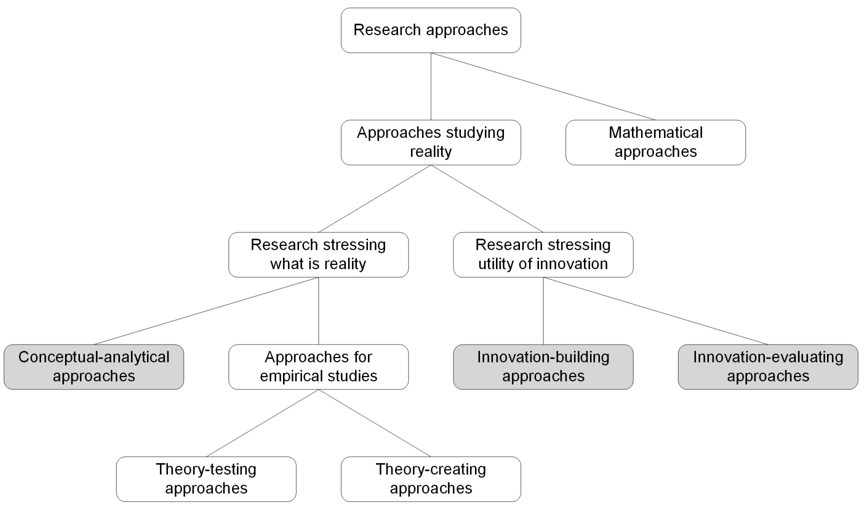

Figure 7. Taxonomy of research approaches (adapted from Järvinen, 2004).

The first part of the research focuses on understanding the Internet content delivery ecosystem, the relevant actors and their respective relationships, as well as the key uncertainties in the market. Thus, the research can be categorized as stressing the reality and utilizing the conceptual-analytical approaches. The second part of the research utilizes techno-economic modeling to evaluate the economic feasibility of new Internet content delivery technologies, which emphasizes the utility of innovation and is essentially an innovation-evaluating research approach. In addition, during the process of creating the techno-economic model and the network simulator, innovation-building approaches are used.

3.2 Data collection

This research utilizes several data collection methods to create, evaluate and quantify the models. Firstly, academic literature is used extensively to identify the relevant building blocks of the models, whereas literature from the industry (e.g. whitepapers, company websites and annual reports) is used to quantify the models. All values used in the models are estimations of the real world. Literature review has been an important source for all seven publications and is the main source of data for Publication IV.

Semi-structured expert interviews are used for a deeper understanding of the technical solutions, evaluating the constructed models and quantifying the models. The interviewees include technical and business experts from mobile network operators, Internet service providers, network infrastructure providers, content providers, and data center providers. All interviews have been conducted face-to-face, though follow-up discussions continued via e-mails. Input from interviews plays an essential role in Publications I, III, V, VI and VII.

Publication II utilizes brainstorming to identify the key trends and uncertainties, which are then used to construct the alternative industry scenarios. Several brainstorming sessions are organized with experts in the field of Internet content delivery, such as network operators, content providers, network infrastructure providers, mobile device vendors, and end users. During the session, the key trends and uncertainties are discussed from the political, economic, social and technological perspectives. In addition, the identified key uncertainties are ranked based on their importance and uncertainty.

Lastly, Publication VII utilizes data from network simulations as input for the cost modeling. The simulator is software based and written in Python. The simulated topology and input parameters are gathered from discussions with mobile network operators and literature on user movements and the content distribution. The simulation setup is described in Section 5.2.1.

3.3 Ecosystem analysis

Value network analysis is used to graphically illustrate the actors and value exchanges between them, which lays the foundation for further analysis. Value chains (Porter, 1985) and value networks (Stabell and Fjeldstad, 1998) have been widely used, but Allee (2000a) is the first to divide the value exchanges into different categories. Allee's Three Currencies of Value include 1) goods, services and revenue, 2) knowledge, and 3) intangible benefits. The first currency includes the actual goods or services and monetary payments flowing between actors. The possession of strategic information, planning and process knowledge and employee competences are defined as knowledge. The intangible benefits include, for example, customer loyalty, sense of community, and image enhancement (Allee, 2000b; 2008). Zhao (2008) adopts a modified categorization of the value exchanges to better fit the networking context: 1) services and goods, 2) monetary benefits, and 3) intangible benefits that include knowledge. The value network notation used in Publication I follows the basic idea of Allee's (2000a) configuration with a few modifications based on Zhao's (2008) categorization and on Faratin's (2007) notation to better suit the context of the study. The currency names used are thus: 1) traffic/content transfer, 2) monetary transfer and 3) intangible benefits. The intangible benefits are further divided into three categories based on Allee's (2000b, 2008) definition of intangibles: brand recognition, information and loyalty.

Industry architectures go further into the value network analysis by defining both the roles required to provide the service and the actors responsible for these roles. A role can be defined as a set of activities and technical components, the responsibilities of which cannot be divided between separate actors, whereas an actor (e.g., an organization or an individual) can take one or many roles (Casey et al., 2010). The industry architecture notation by Casey et al. (2010) illustrates both the technical architecture and the industry architecture as different layers of a single figure. The technical architecture describes the technical components, i.e. collections and realizations of technical functionalities within a technical system, and the interfaces between them. Industry architecture provides a description of the roles and actors within an industry and the relationships among them. The industry architecture analysis is used as the sole method in Publication III and as an illustrative tool in Publications II and IV.

Two-sided market theory is used to explain and analyze the behavior of actors in the Internet content delivery market that exhibit two-sidedness. Rochet and Tirole (2006) define two-sided markets as markets with two distinct sides that are interlinked (through a platform) and where not only the overall price level matters, but also the price structure between the two sides. The demand asymmetries between interlinked markets typically lead to skewed pricing with one side charged less than the other (i.e. subsidized), which in turn leads to higher quantities demanded on the subsidized side. Due to cross-side network externalities (Parker and Van Alstyne, 2005), the increased consumption on the subsidized side increases the demand on the non-subsidized side and introduces the possibility to charge even more on the non-subsidized side. The resulting combined revenue for the platform from both sides is bigger than in the non-skewed pricing case. The platform typically subsidizes the side that brings more value to the platform or has relatively more elastic demand (Armstrong, 2006). The research results of Publication I relies heavily on analyzing the two-sided markets of Internet content delivery.

Scenario planning methods have been commonly used for long range business planning and decision making under uncertainty, such as for war game simulations in 1950s (Schoemaker, 1993), preparation for the 1973 oil crisis (Wack, 1985) and more recently for managing the uncertainties related to emerging technologies in the rapidly evolving ICT industry (Shoemaker and Mavaddat, 2000; Karlson et al., 2003; Heikkinen and Hämmäinen, 2008; Smura and Sorri, 2009). As a result of its versatility, scenario planning methodology has developed in several different directions. Bradfield et al. (2005) provide an overview, classification, and comparison of the main schools of scenario planning methodologies. Publication II adopts the scenario planning method of Schoemaker (Schoemaker, 1991; Schoemaker, 1993; Schoemaker, 1995; Schoemaker and Mavaddat, 2000), as it has been found to be suitable for understanding and bounding the uncertainties related to emerging ICT technologies. The scenario planning process consists of ten steps, which are listed in Table 2. Steps 1 and 2 are carried out by the researchers alone and inputs for steps 3-5 are gathered from brainstorming sessions as described in Section 3.2. Based on the collected data, the scenario construction and analysis (Steps 6-8) are carried out by the researchers. In Step 8, industry architecture notation is used to visualize and communicate the differences between the scenarios. The scope of the study (Publication II) is limited to the qualitative analysis part and Steps 9-10 are not included in the analysis.

Table 2. Scenario planning process (adapted from Schoemaker and Mavaddat, 2000).

#	Step
1	Define the issues you wish to understand better in terms of time frame, scope, and decision variables.
2	Identify the major actors who would have an interest in these issues, and their current roles, interests, and power positions.
3	Identify and study the main forces that are shaping the future within the scope, covering the social, technological, economic, environmental, and political domains.
4	Identify trends (forces, whose outcome is agreed on by experts) or predetermined elements that will affect the issues of interest from the list of main forces.
5	Identify key uncertainties (forces deemed important, whose outcomes are not very predictable) from the list of main forces. Examine how they interrelate.
6	Select the two most important key uncertainties, and cross their outcomes in a matrix. Add suitable outcomes from other key uncertainties, as well as trends and predetermined elements to all scenarios.
7	Assess the internal consistency and plausibility of the initial scenarios, revise.
8	Assess how the key actors might behave in the revised scenarios.
9	See if certain interactions can be formalized in a quantitative model.
10	Reassess the uncertainty ranges of the main variables of interest, and express more quantitatively how each variable looks under different scenarios.

3.4 Techno-economic modeling

Techno-economic modeling is a future oriented method to analyze and evaluate the economic feasibility of technical systems by utilizing forecasting, network design and investment analysis methods (Smura, 2012). The term techno-economic was first introduced in telecommunications by the research program

RACE during 1985-1995 (Konidaris, 1991). Smura (2012) gives a detailed description of the development history of techno-economics. Previously techno-economic modeling has been used, for example, to analyze the virtual network operators (Smura et al., 2007), MNOs' alternatives in 3G (Harno et al., 2009), next generation access networks (Kantor et al., 2010), virtualization of mobile networks (Naudts et al., 2012) and the life-cycle-cost of SDN/NFV in mobile networks (Knoll, 2015). In Publications V, VI and VII, techno-economic modeling is used to analyze the cost efficiency of introducing SDN into the LTE network from the perspectives of caching and service chaining, leaving the revenue considerations for future work.

Two generic approaches exist for cost modeling: top-down and bottom-up. The top-down approach uses the existing network infrastructure, costs and usage levels to calculate costs per service, whereas the bottom-up approach uses demand forecasts to model the required costs for meeting the demand (Smura, 2012). The cost modeling in Publications V, VI and VII utilizes a combination of the two approaches, i.e. the reference values from the existing LTE network are obtained top-down, which are then used as inputs in the bottom-up constructed SDN-LTE network model.

Cost modeling can be divided into capital expenditure (CAPEX) and operational expenditure (OPEX). CAPEX includes the network investments and the initial deployment costs. OPEX typically includes network related OPEX, interconnection and roaming costs, sales and marketing costs, billing and subscriber management costs, and general and administration costs. Traditionally, OPEX has been calculated as a proportion of CAPEX, but Verbrugge et al. (2006) propose a method for calculating both CAPEX and OPEX as separate entities. This research considers network related OPEX and defines it as the costs related to operations, administration and maintenance, which typically use the number of elements as input, as well as the interconnection and roaming costs, which depend on the traffic distribution between the networks.

Due to the immaturity of the technologies and the future-oriented nature of the research, uncertainties in the input estimations call for sensitivity analysis. The purpose of sensitivity analysis is to study the impact on the outputs of the model, when the input assumptions are changed. In sensitivity analysis, the input variables can be changed one at a time, while keeping all other inputs constant. Alternatively, many or all variables can be changed simultaneously, enabling the formation of different what-if and worst/best case scenarios. Both types of sensitivity analysis is utilized in this research.

A summary of the used methods and data sources is shown in Table 3. In addition, the actor perspective taken in each publication is also shown in Table 3.

Table 3. Summary of the research process.

Publication	Objective	Research methods	Data source	Perspective
Publication I: "Value Networks and Two-Sided Markets of Internet Content Delivery"	To identify the key actors of Internet content delivery market and to analyze and compare different Internet content delivery technologies	Value network analysis, two-sided market analysis	Literature, interviews	Market, all actors
Publication II: "Scenario Analysis for Commercial Internet Content Delivery"	To identify and analyze key uncertainties and construct alternative industry scenarios for Internet content delivery	Scenario planning, industry architecture analysis	Literature, brainstorms	Market, all actors
Publication III: "SDMN: Industry Architecture Evolution Paths"	To identify the key roles and actors and analyze potential future industry architectures for software-defined mobile networks	Industry architecture analysis	Literature, interviews	Market, all actors
Publication IV: "Video Delivery over Next Generation Cellular Networks"	To analyze the actors and their business interests and relationships in two video-on-demand use cases over virtualized LTE networks	Industry architecture analysis, business model framework	Literature	Market, all actors
Publication V: "Cost Efficiency of SDN-enabled Service Function Chaining"	To analyze the cost efficiency of SDN enabled service function chaining	Techno-economic modeling	Literature, interviews	Mobile network operator
Publication VI: "Cost Efficiency of SDN in LTE-based Mobile Networks: Case Finland"	To model the Finnish LTE topology and to construct a cost model to analyze the cost efficiency of SDN-LTE	Techno-economic modeling	Literature, interviews	Mobile network operator
Publication VII: "Cost Efficiency of Mobile in-Network Caching"	To analyze the cost efficiency of SDN enabled in-network caching by utilizing network simulation results	Techno-economic modeling, network simulation	Literature, interviews, simulation results	Mobile network operator

4 . Internet content delivery ecosystem

The first part of the research qualitatively analyzes the Internet content delivery ecosystem and industry from the perspectives of the key actors and the main technologies. Firstly, in-network caching's value networks and market dynamics are compared with the existing Internet content delivery models to evaluate the attractiveness of in-network caching to the different actors (Publication I). Secondly, potential role and actor configurations are mapped onto the existing and SDN-enabled LTE architectures (Publication III), together with a Video-on-Demand (VoD) use case (Publication IV). Lastly, industry scenarios of Internet content delivery are discussed together with the key trends and uncertainties of the market (Publication II). The following subsections present the purpose and main results for each of the studies.

4.1 Market structure of Internet content delivery

To compare the market structure of different Internet content delivery solutions, i.e. basic client-server Internet (Civanlar and Haskell, 1999), content delivery networks (CDN) (Vakali and Pallis, 2003), peer-to-peer (P2P) networks (Camarillo, 2009; Risson and Moors, 2007) and information-centric networking (ICN) (Jacobson et al., 2012; Ahlgren et al., 2012), value network analysis and two-sided market theory are used. The value networks are constructed based on the IETF Internet architectural principles and guidelines (Carpenter, 1996; Bush and Meyer, 2002). The results show that Internet Service Providers' (ISPs') interest in deploying ICN (i.e. in-network caching) is crucial due to their ownership of the cache locations, and content providers (CPs) have a strong incentive to pay for providing better quality of experience (QoE) to the end users. However, ICN faces strong competition from CDNs and need to solve the coordination challenges related to content access control, contracting, cost allocation and content usage statistics. ICN is chosen as the example for in-network caching in this analysis, but the results can be extended to other in-network caching solutions as well.

4.1.1 Value networks

Expert interviews with ISPs, CPs and data center providers were conducted in 2010 to identify the key actors in the market and the relationships between the actors, i.e. value networks. In addition to the actors defined in Section 1.3, the content producer as well as the advertisers[5] and sponsors[6] are also considered due to their important role as revenue sources to content providers.

The Internet content delivery value networks can be divided into two layers: 1) the content service layer, where the product is digital content, such as video clips or pieces of music, and 2) the Internet interconnection layer, where the product is transportation of content over the Internet. In the content service layer, the payments reflect the value of the content and can vary between different pieces of content, whereas in the Internet interconnection layer, the level of monetary transfer reflects the value of bit transfer and is indifferent to the content.

Figure 8 shows a simplified value network of the content service layer, which stays the same regardless of the underlying Internet content delivery model. The content service layer's value network can also be considered as the CP's internal value network as it shows the origin of the content (i.e. the content producer), the different ways the content can be stored and the different revenue sources of the CP (i.e. sponsors, advertisers and payments from the end users).

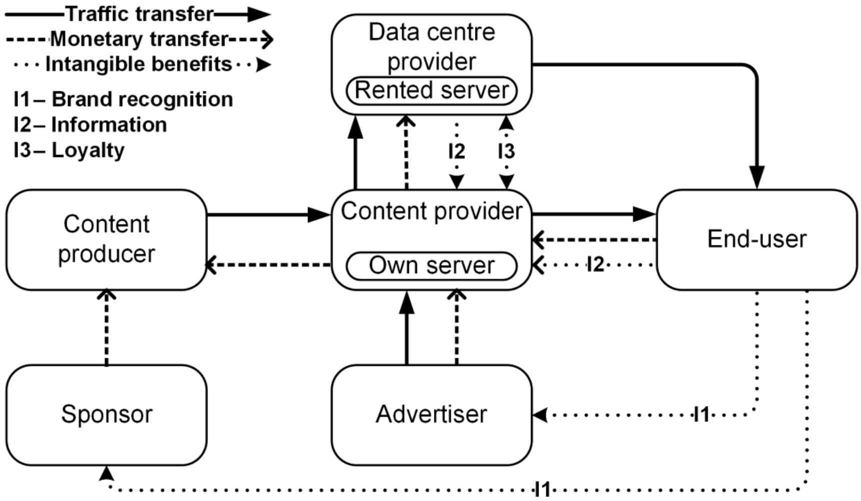

Figure 8. Simplified value network of the content service layer.[7]

The value networks of the Internet interconnection layer are separated for each of the four content delivery models. The client-server model without caching (Figure 9a) is chosen as the base case, and the differences to that model are highlighted by using darker arrows in the CDN model (Figure 9b), the P2P model (Figure 9c) and the ICN model (Figure 9d). The thickness of the arrows represent the relative traffic and monetary volume between actors, and the size differences in actors' boxes show the relative sizes of the same actors (e.g. access-ISP 2 is bigger than access-ISP 1).

[5] Advertisers insert advertisements into final products in the distribution process.

[6] Sponsors input their brand names during the content making process.

[7] Adapted from Publication I and modified. Reprinted with permission from Elsevier Ltd.

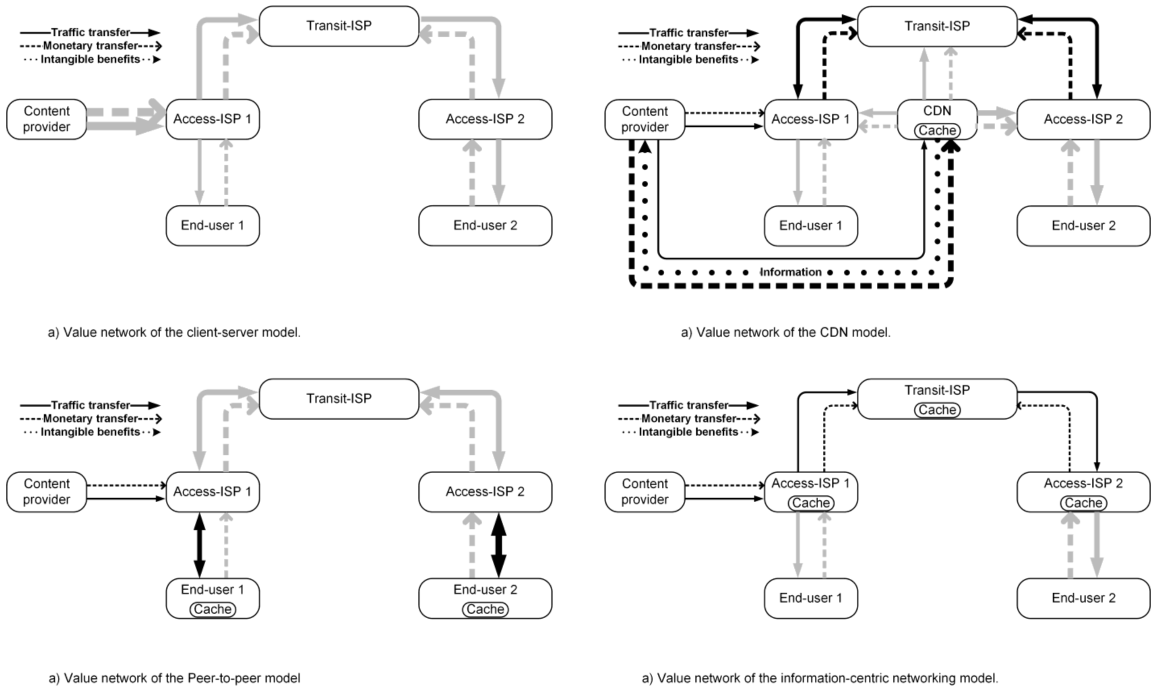

Figure 9. Value networks of a) the client-server, b) CDN, c) peer-to-peer and d) information-centric networking models. The differences of each value network compared with the client-server model are shown with darker arrows.[8]

[8] Adapted from Publication I and modified. Reprinted with permission from Elsevier Ltd.

From the value networks, it can be seen that the CDN and ICN models save transit traffic costs for the access-ISPs. From the CPs perspective, caching at all locations reduce the link size between CP and access-ISP 1, which leads to cost reduction for CPs. On the other hand, caching at end-user devices (i.e. P2P) increases the traffic in the access networks, but does not bring additional revenue for access-ISPs due to the assumed flat-rate data pricing.

4.1.2 Two-sided markets

The traffic and monetary flows of the value networks illustrate several two-sided markets, listed in Table 4. The two-sided markets of the content service layer have been analyzed in previous works by, for example, Hagiu and Yoffie (2009) and Rochet and Tirole (2003). Thus, the focus in this research is on the Internet interconnection layer, where access-ISPs, transit-ISPs and CDNs can act as two-sided platforms.

The side that receives subsidies from the platform is also shown in Table 4. The clearest two-sided market is the CDN market, where the ISPs are subsidized, because the CPs are more willing to pay for improving the QoE of the end users. Due to the cross-side network effects of two-sided markets, the big CDN providers will grow even bigger and few CDNs will dominate the whole market. As a consequence, end users have access to a bigger content selection and content providers can serve more end users. However, multi-homing CPs and ISPs, as well as the CDN-interconnect initiative (Peterson and Davie, 2014) may diminish the cross-side network effects.

In the access-ISP market, the end users are subsidized, because their demand tend to be more price elastic compared to single-homing CPs. However, more commonly CPs are multi-homing, which diminishes the market power of a single access-ISP and, thus, CP’s demand is becoming more price elastic.

In the transit-ISP market, the eyeball-ISPs are subsidized due to the asymmetries in cost from the inbound and outbound traffic of the transit-ISP. The terminating access-ISP (i.e. eyeball-ISP) incurs more costs to the transit-ISP due to hot-potato routing, in which the originating transit-ISP has an incentive to pass on the off-net traffic as soon as possible to the terminating transit-ISP (Laffont et al., 2003; Clark et al., 2011). At the same time, the eyeball-ISP and the content-ISP are assumed to pay the same transit unit price despite the eyeball-ISP causing more costs to the transit-ISP (Norton, 2011).

The ICN model does not present any significant two-sidedness and, thus, the actors involved do not benefit directly from two-sided pricing as in the CDN market. In addition, the actor that can monetize from the in-network caches remains uncertain and ISPs might opt for building their own CDNs rather than installing in-network caches. Table 5 compares the four models with parameters such as cost for access-ISPs and CPs, CP’s control over its content and QoE, where darker cells are more preferable than lighter cells from the perspective of the actor in question.

Table 4. Two-sided markets in Internet content delivery.[9]

Layer	Platform	Side 1 (subsidized side)	Side 2	Market Name
Content Service Layer	**Content provider**	End users	Advertisers	Advertising Market
	Content producer	End users	Sponsors	Sponsorship Market
	Content provider	End users	Content producers	Content Sharing Market
Internet Interconnection Layer	**Internet access provider**	End users	Content providers	Access-ISP Market
	Internet backbone provider	Eyeball-ISPs	Content-ISPs	Transit-ISP Market
	CDN provider	ISPs	Content providers	CDN Market

Table 5. Comparison of different Internet content delivery models based on several parameters.[10]

Parameter	Client-Server Model	CDN Model	P2P Model	ICN Model
Cost for access-ISP	high	low	high	low
Cost for CP	high	high	low	low
CP's control over caching	high	high	low	low
CP's control over content distribu-tion	high	high	high	low
Level of guaran-teed QoS	medium	high	low	low
Latency for end user	high	low	medium	low
Scalability	low	medium	high	high

[9] Adapted from Publication I and modified. Reprinted with permission from Elsevier Ltd.

[10] Adapted from Publication I and modified. Reprinted with permission from Elsevier Ltd.

4.2 Market structure of mobile content delivery

For the MNO's perspective, SDN enables in-network caching in the mobile networks. Thus, SDN's influence on mobile Internet service provisioning and the industry structure is analyzed in this subsection. In addition, new business opportunities are discussed with two video-on-demand (VoD) use cases that utilize SDN-LTE. The technology underlying the constructed industry architectures are based on 3GPP network architectures and the evolved universal terrestrial radio access network specifications (3GPP, 2015a; 3GPP, 2015b).

Two deployment approaches for SDN-LTE are discussed: 1) evolutionary SDN-LTE and 2) revolutionary SDN-LTE. In the evolutionary SDN-LTE, the control plane is decoupled from the user plane (Raghavan et al., 2012; Kreutz et al., 2015), but the user plane elements are assumed to stay in the same physical locations, though their control plane functions are centralized into data centers. In the revolutionary SDN-LTE, network functions virtualization (NFV) is utilized to re-optimize the control plane functions (Open Networking Foundation, 2014).

From the economic perspective, evolutionary SDN-LTE could bring incremental improvements to the operational efficiency of the existing industry architectures. On the other hand, revolutionary SDN-LTE could disrupt the current market situation through new industry architectures, where the value is redistributed among the existing and potential new actors in the market.

Table 6 describes the key roles of SDN-LTE that can be mapped to the decoupled technical elements of a LTE network. Two of the roles, network usage and interconnection provisioning, are always controlled by the same actors, as illustrated in the generic role configuration (Figure 10). The next subsections present how the rest of the roles can be managed by different actors.

Table 6. Key roles of SDN-LTE.[11]

Roles	Description
Network usage	Accessing the network with a mobile device.
Radio network forwarding	Receiving the user traffic in eNBs and forwarding it to the evolved packet core.
Radio network management	Management and operation of the base stations and radio frequencies.
Core network forwarding	Traffic forwarding in the evolved packet core network.
Core network routing	Traffic routing in the evolved packet core network.
Public network forwarding	Traffic forwarding and filtering (i.e. firewall functionality) between the public network and the core network.
Connectivity management	Management of connectivity 1) between the public network and the core network and 2) in the evolved packet core, including situations of inter-eNB handover. Can be divided into 1) public network connectivity management and 2) mobile network connectivity management, respectively.
Mobility management	Management of control plane signaling between the eNB and other network elements like HSS.
Subscriber management	Management of the user- and subscription related information, including user authentication, access authorization and home network information.
Policy and charging	Brokering quality of service and charging policy on a per-flow basis.
Interconnection provisioning	Providing the interconnection to public IP networks and other mobile networks through transit, peering and roaming agreements.

[11] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

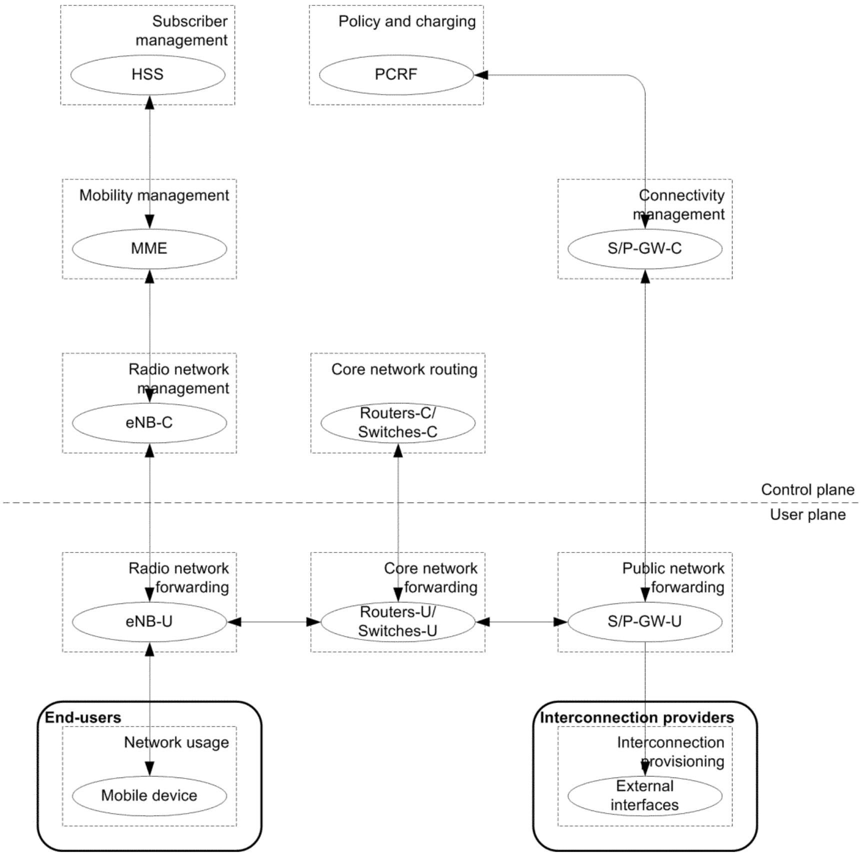

Figure 10. Generic role configuration of SDN-LTE.[12]

4.2.1 Evolutionary SDN-LTE

The three evolutionary SDN-LTE industry architectures constructed are technically feasible with the existing deployments of the MNOs in the Finnish market. Thus, the focus of the evolutionary SDN-LTE analysis is on how SDN could increase flexibility and improve operational efficiency of the mobile Internet service provisioning.

Due to their ownership of the current network infrastructure and the business relationship with the end users, MNOs are natural drivers for the SDN-LTE deployment. Thus, Figure 11 shows a monolithic MNO that runs the control plane functions on its own mobile cloud platform. The same MNO also negotiates with the interconnection providers for the roaming, transit and peering agreements. The main benefit for the MNO in this industry architecture is the potential cost saving from using general purpose hardware in both the user plane elements and the control plane cloud platform. The cost efficiency of owning and operating the network in Figure 11 is discussed in Section 5.

[12] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

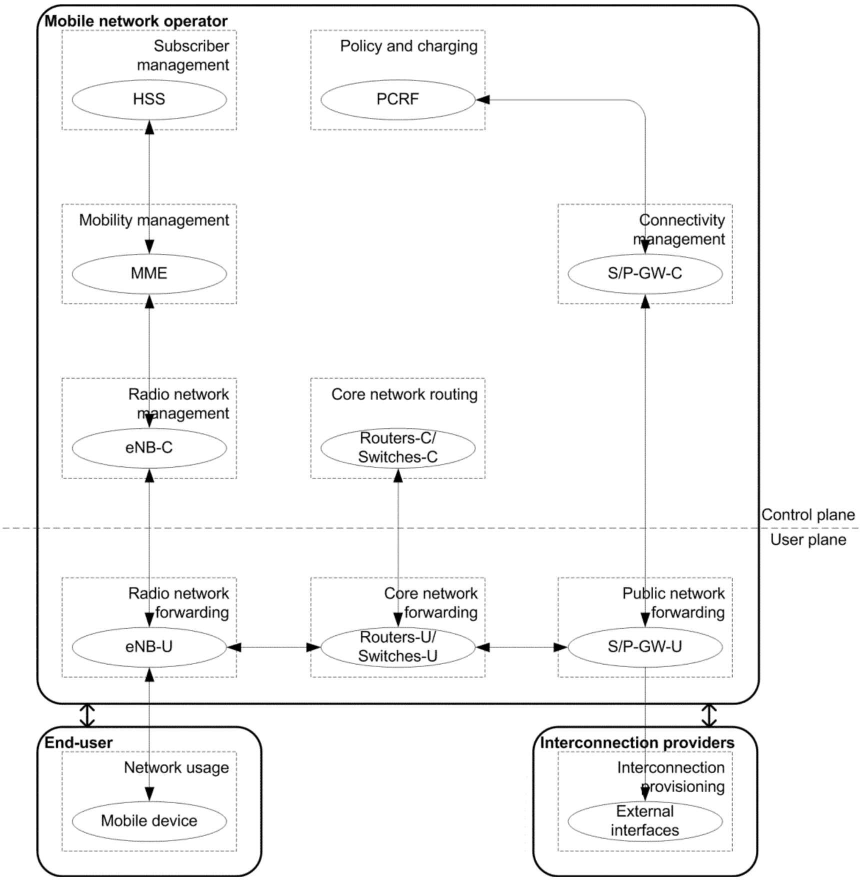

Figure 11. Evolutionary SDN-LTE with monolithic MNO.[13]

The second evolutionary SDN-LTE industry architecture (Figure 12) illustrates a situation, where a mobile virtual network operator (MVNO) manages a front-end HSS (marked as FE-HSS in Figure 12) and PCRF. The MNO retains control of the critical functions of the network, such as MME, S/P-GW and HSS database (marked as HSS-DB in Figure 12). In addition, the MVNO could be leasing the cloud platform that the HSS and PCRF are running on. This industry architecture could be especially beneficial to the MNO if it can charge the MVNO a higher fee for the more agile network infrastructure.

[13] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

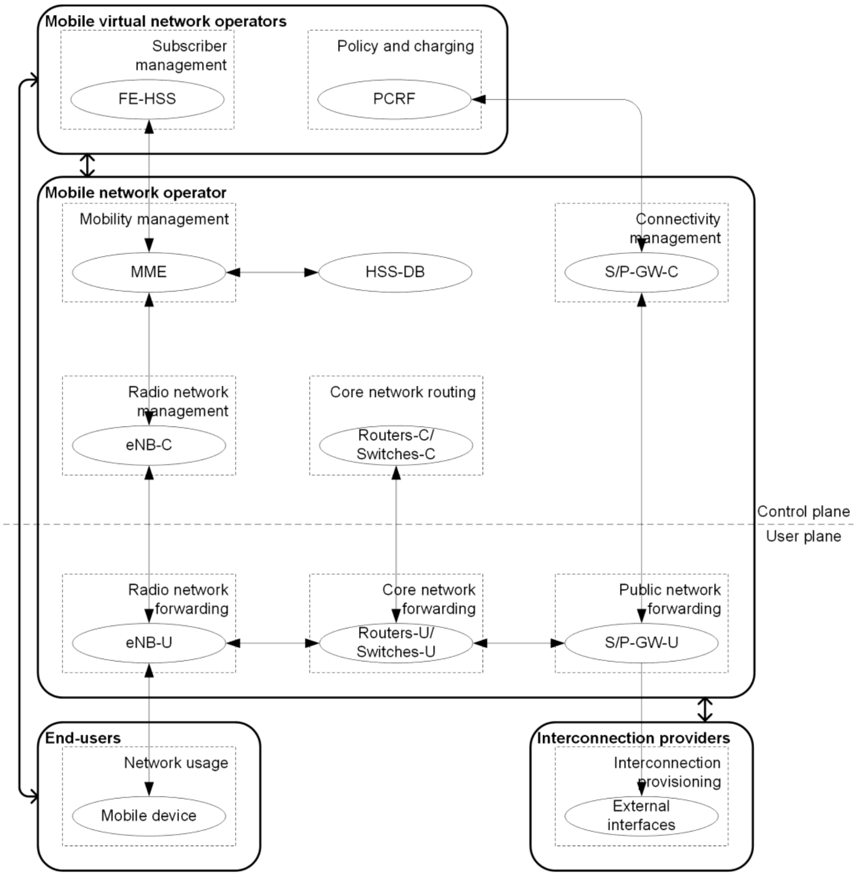

Figure 12. Evolutionary SDN-LTE with outsourced subscriber management.[14]

In Figure 13, the traditional monolithic MNO outsources also the core network connectivity to a connectivity provider, but the connectivity management and radio network functions are still in the hands of the MNO. The subscriber management and policy and charging are managed by the MVNO in the same way as in Figure 12. Interconnection negotiations and business agreements with the transit, peering and roaming partners are still taken care by the MNO. In this industry architecture, the MNO loses some control of the transport network, but also saves cost by not operating the transport infrastructure.

[14] Adapted from Publication III and modified. Reprinted with permission from John Wiley & Sons, Ltd.

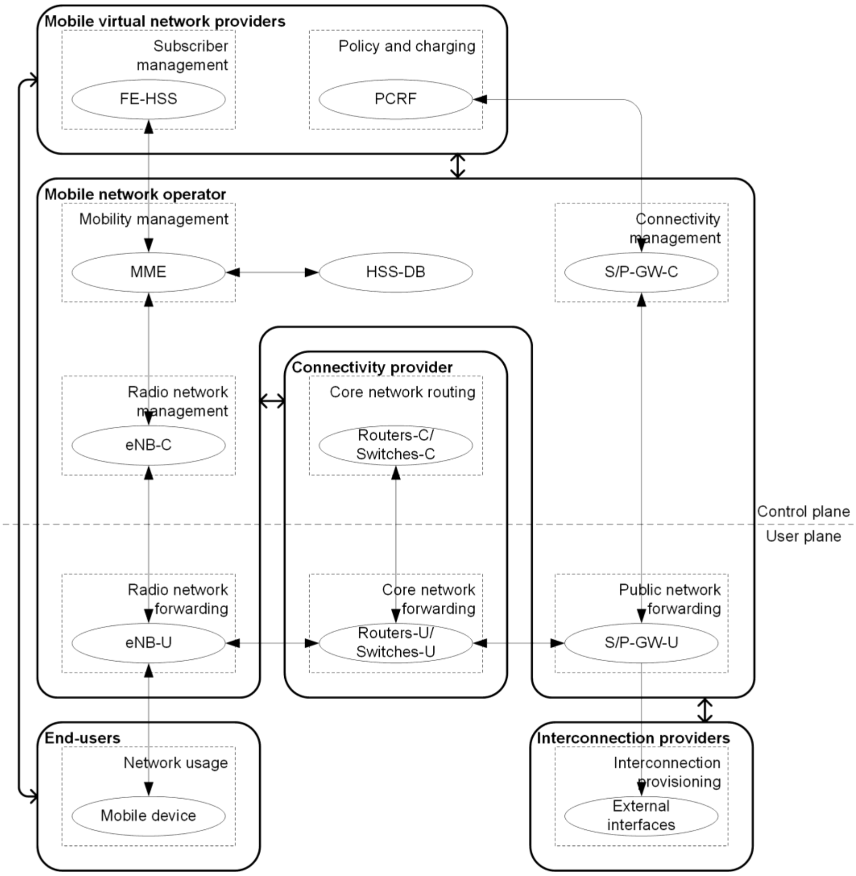

Figure 13. Evolutionary SDN-LTE with outsourced connectivity.[15]

4.2.2 Revolutionary SDN-LTE

The scenarios of evolutionary SDN-LTE have their continuations in the revolutionary phase, but the focus of the revolutionary industry architectures is on the changes in the market structure, when SDN-LTE enables outsourcing of the different network elements. For example, MNO's control over the network is reduced. At the same time, MNOs maintain the control of the radio frequencies and the base station infrastructure, thus, their position in the market is still strong. The three revolutionary scenarios discussed are: 1) full MVNO, 2) outsourced interconnectivity, and 3) outsourced mobility management.

Figure 14 illustrates an industry architecture that is enabled by the decoupling of the control and user planes, where a full MVNO manages the whole mobile core network and leases just the frequency and connectivity from the MNO. In addition, the interconnection agreements are also controlled by the MVNO. This scenario needs a global MVNO, who can accommodate all the network elements, whose users are highly mobile, and who wants to control its own roaming agreements instead of relying on the MNO's roaming partners. On the other hand,

[15] Adapted from Publication III and modified. Reprinted with permission from John Wiley & Sons, Ltd.

the MNO loses control over the network operations and the management of the baseband processing pooling sites (i.e. eNB-C), which means access management and resource optimization in the radio and core networks might be less efficient.

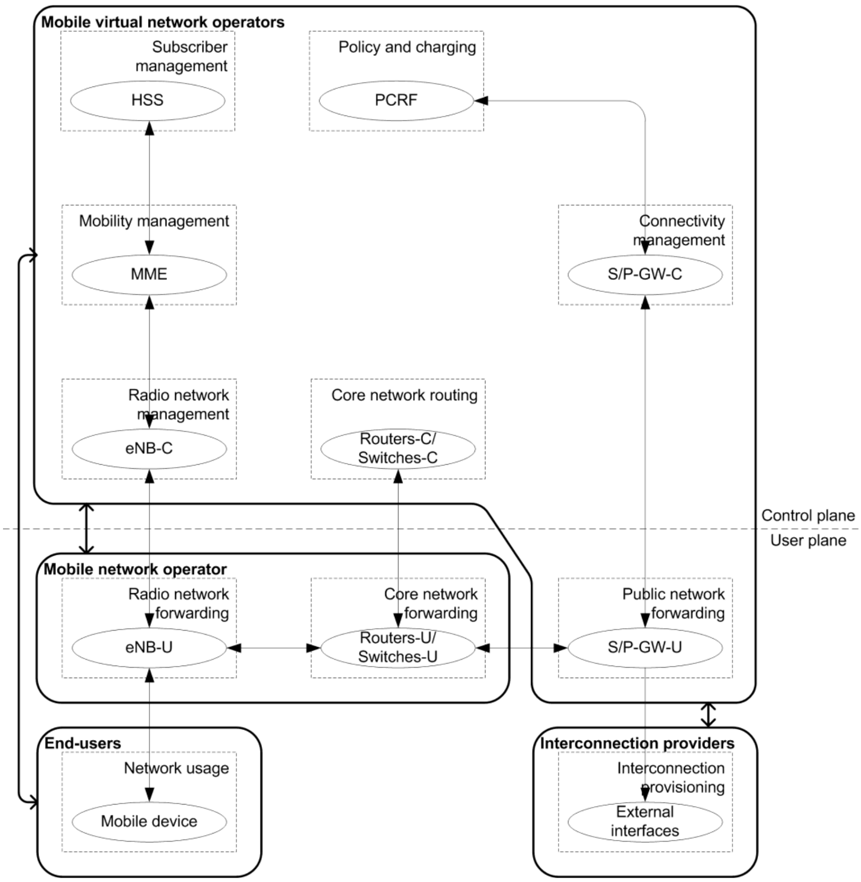

Figure 14. Revolutionary SDN-LTE with full mobile virtual network operator.[16]

The second and third revolutionary SDN-LTE industry architectures are constructed for a technical architecture, where the S/P-GW elements are removed from the network, as is shown in Figure 15 and Figure 16. The removal of the S/P-GW technical components (Costa-Requena et al., 2015a) enables a more flexible division of the control plane functions. The functionalities of the S/P-GWs are divided between new collections of functionalities marked as MME+, HSS+ and routers-C+/switches-C+.

In Figure 15, the roles related to S/P-GW are now managed by the MNO (mobile network connectivity management role) and the connectivity provider (public network connectivity management role). As a consequence, mobility management could be done in a simpler and lighter way, and local breakout could be handled more efficiently, especially if the MVNO is big enough and has

[16] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

a trustworthy reputation. In addition, the management of the interconnection agreements is outsourced to a connectivity provider, who gains economies of scale benefits from serving several MNOs. A bigger connectivity provider might be able to negotiate better prices from the interconnection providers and have higher chances of peering with other big connectivity providers instead of buying transit from them. However, roaming brings significant revenues for MNOs, who might not want to give up control of the roaming agreements to connectivity providers.

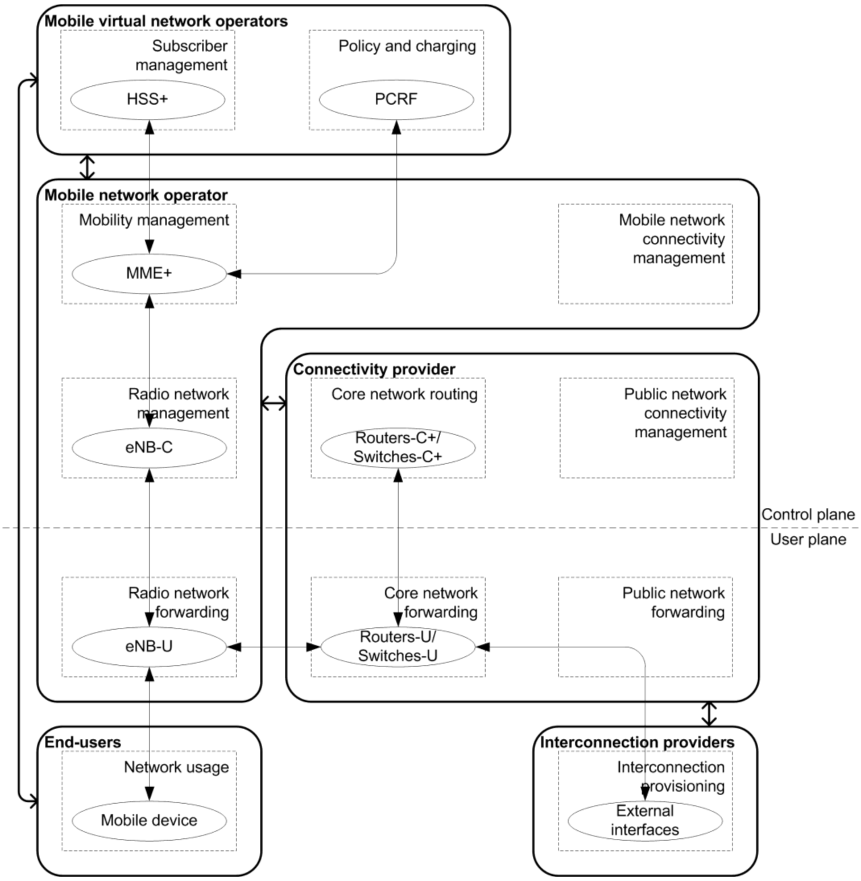

Figure 15. Revolutionary SDN-LTE with outsourced interconnection.[17]

In Figure 16, the MNO outsources also the mobility management and radio network management to a mobility provider. Potential mobility providers include network equipment vendors, such as Ericsson and Nokia, who could offer the service to run the mobility management and radio network management on their cloud platforms instead of selling the network infrastructure to MNOs. The mobility provider in this industry architecture could also serve several MNOs and gain economies of scale benefits, which could potentially reduce the operational expenditure of the MNO. MNOs still own the frequency licenses and

[17] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

would control the radio related resources. Figure 16 proposes a separate MVNO, who manages the subscribers, services and charging. However, these roles can also be taken by the MNO itself, who might not wish to lose control of the business interfaces to the end users in a network, where everything else is outsourced.

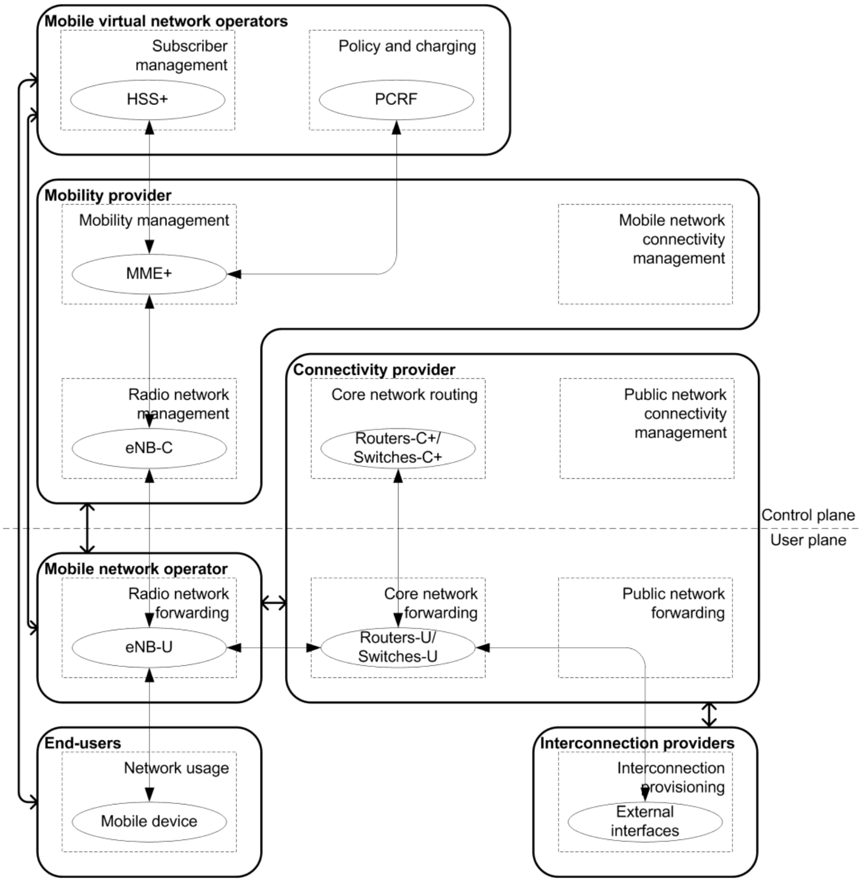

Figure 16. Revolutionary SDN-LTE with outsourced mobility management.[18]

4.2.3 Video delivery use cases

This section discusses two VoD use cases utilizing the evolutionary SDN-LTE architecture with outsourced connectivity. In the first use case, MNO operates its own VoD service from its own video platform, as illustrated in Figure 17. Access control and billing are provided by end-user subscriptions, i.e. SIM cards. Content can be produced by a third-party content provider or by the MNO itself, but content ownership is omitted from the figure for simplicity. The MNO may also utilize caching for better performance and lower latency for the end users. The MNO buys or rents the router network and full Internet connectivity from the connectivity provider. The MNO in the VoD use cases includes both the

[18] Adapted from Publication III. Reprinted with permission from John Wiley & Sons, Ltd.

MNO and the MVNO variants, since this distinction has limited impact on the VoD service.

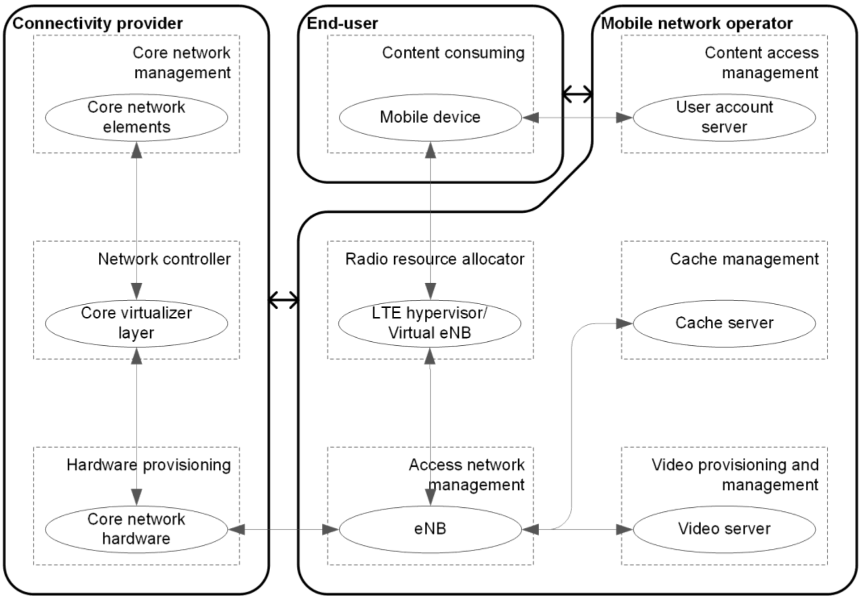

Figure 17. MNO operated video-on-demand service.[19]

Figure 18 illustrates the second use case, where the VoD service is operated over a third-party video platform, e.g. a CDN or a cloud provider that is located outside the MNO's LTE network. The video platform in this use case may also represent multiple instances of CDN or cloud providers, who interact and have business agreements with each other. For example, a federation of cloud operators or CDNs, which essentially serves as a large virtual CDN and has the potential of attracting large content providers, can be considered in this case. A sufficiently large MNO in this use case could establish a direct peering link with the CDN or cloud provider, which would allow bypassing the connectivity provider and the possibility to control the quality of the video service. The efficient provisioning of the service to the end users requires business agreements between the aforementioned actors, which could include service level agreements. This use case divides the responsibilities of the VoD service among different actors and cost savings could be obtained due to economies of scale.

[19] Adapted from Publication IV and modified. Reprinted with permission from IEEE.

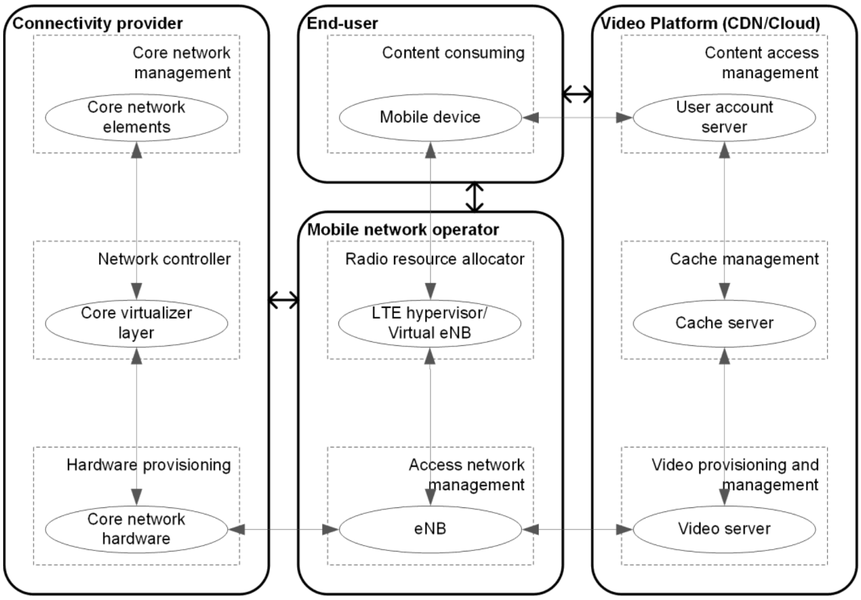

Figure 18. Third-party operated video-on-demand service.[20]

4.3 Industry scenarios

Lastly, the key uncertainties are identified and analyzed, and alternative industry scenarios for Internet content delivery during the next 10 years[21] are constructed. The focus of this work is on commercial and human usable heavy content, such as video. The considered caching technologies are CDNs, P2P networks and in-network caching (using ICN as an example). In addition, a socio-economic perspective is taken, though the various technical forces are also considered. For data collection and evaluation of the data, two workshops were organized between fall 2011 and spring 2012 with senior experts in the field of Internet content delivery.

Four industry scenarios are constructed based on the identified two most important uncertainties: the content provider's revenue model and ISP's role[22] in content provision. Based on the results, the relative positions and roles of different industry actors and content delivery technologies in each of the scenarios are discussed.

4.3.1 Key trends and uncertainties

The two workshops yielded 94 different forces affecting the evolution of the Internet content delivery market in the next ten years. The identified forces were prioritized by the experts during the sessions. Ten key trends and eight key uncertainties were prioritized by the researcers after grouping similar forces. The

[20] Adapted from Publication IV and modified. Reprinted with permission from IEEE.

[21] The study was done during 2011-2012.

[22] ISP, as opposed to the definition in Section 1.3, in this study is used to represent both fixed and mobile connectivity providers.

identified key trends (Table 7) mainly revolve around four themes: the growing content volume, the increasing demand for connectivity, the increasing importance of mobility, and personalization of content.

Table 7. Key trends by themes.[23]

Growing content volume
More mini content producers and user-generated content will emerge into the market.
More than 50 billion devices will be connected in 10 years.
More metadata will be created and big data harvesting is becoming an important economic driver.
Increasing demand for connectivity
ISPs have more pressure to offer access.
Connectivity is becoming a commodity.
Increasing importance of mobility
End-user content demand and increased QoE expectations are to be met at any time and place.
Higher data rates, higher processing power and high definition screens in mobile devices are becoming more common.
Personalization of content
Context, e.g. location, aware services will increase.
Personalization of content is becoming more important.
Personal identity and profile data increasingly used as currency to pay for services.

Eight uncertainties in their order of importance are listed in Table 8, together with their mutual correlations. Pairs of uncertainties that show correlation with each other are marked with + and uncorrelated pairs with 0. The uncertainties can be divided into four main themes: 1) how content is accessed, 2) who controls the content, 3) who controls the network, and 4) how caching is done. The first two uncertainties are chosen for constructing the scenarios. These two uncertainties are evaluated to be sufficiently uncorrelated and to span the best matching matrix of four distinctive scenarios for the future of Internet content delivery.

The most important uncertainty is ISP's role in the content delivery process: will mobile ISPs go for strong ISP bundling or no ISP bundling? To avoid being a mere bit pipe, the ISPs need new revenue sources. For example, at the time of writing Publication II[24], Orange in Europe and Comcast in the USA were offering bundles that include a certain amount of films or TV channels, the Internet broadband and a mobile connection.

On the other side of the market is the content provider and its revenue source. In the past, advertising revenue has dominated the Internet content provider market. However, several payment-based content providers have emerged into the market, e.g., Spotify and Netflix. In addition, increasingly the payment model is freemium (Wagner et al., 2013), where the basic service (with lower quality, advertising and potentially limited access) is offered for free and end users can pay for the premium service. How will the dynamics evolve: will paid

[23] Adapted from Publication II. Reprinted with permission from Emerald Group Publishing Limited.

[24] Since 2012, the market has evolved towards stronger ISP and MNO bundling. Many more ISPs and MNOs have their content platforms, through which the ISPs' or MNOs' subscribers can access movies and TV channels for a fixed fee.

content, i.e. consumer revenue model, become more dominant than the advertiser revenue model?[25]

Table 8. Key uncertainties and their correlations.[26]

Id	Key uncertainty	Possible outcomes	Correlation with other uncertainties							
			U1	U2	U3	U4	U5	U6	U7	U8
U1	Mobile ISP bundling	1) strong ISP bundling 2) no ISP bundling	1	0	+	+	+	0	0	+
U2	Content provider revenue model	1) advertiser revenue 2) consumer revenue		1	+	0	0	0	+	+
U3	Content delivery control	1) content provider 2) network			1	+	+	+	0	+
U4	Content provision aggregation	1) aggregated 2) fragmented				1	+	+	+	0
U5	Cache ownership	1) independent 2a) bundled to network provisioning 2b) bundled to content provisioning					1	0	+	+
U6	Content exclusivity	1) exclusive 2) not exclusive						1	+	0
U7	Dominating cacheable content	1) global 2) local							1	+
U8	Driver for cache location optimization	1) low cost 2) fast response time								1

4.3.2 Resulting scenarios

Figure 19 illustrates the four scenarios constructed by crossing the two most important uncertainties in a matrix. The y-axis represents the two outcomes of mobile ISP bundling (U1), i.e. whether the ISP provides access and content as a service bundle or not. The x-axis represents the content provider's revenue model (U2), where the content provider can either charge end users for the content or use the advertising based revenue model. Scenario naming is based on the prevailing end-user type in each scenario, including 1) Comfort Buyer, 2) Indifferent Saver, 3) Quality Buyer, and 4) Demanding Saver. The prevailing end-user type directly influences the type of actor that is likely to have a dominant position in the Internet content delivery market and the power to make decisions on the used caching system. The dominant actors in each of the scenarios are 1) a payment-based ISP, 2) an advertising ISP, 3) a payment-based content provider, and 4) an advertising content provider.

[25] With hindsight, the uncertainty is becoming less uncertain. Freemium seems to be the dominating payment model in Internet content delivery and an increasing number of the end users are paying for the premium service (Wagner et al., 2013; Spotify, 2015).

[26] Adapted from Publication II. Reprinted with permission from Emerald Group Publishing Limited.

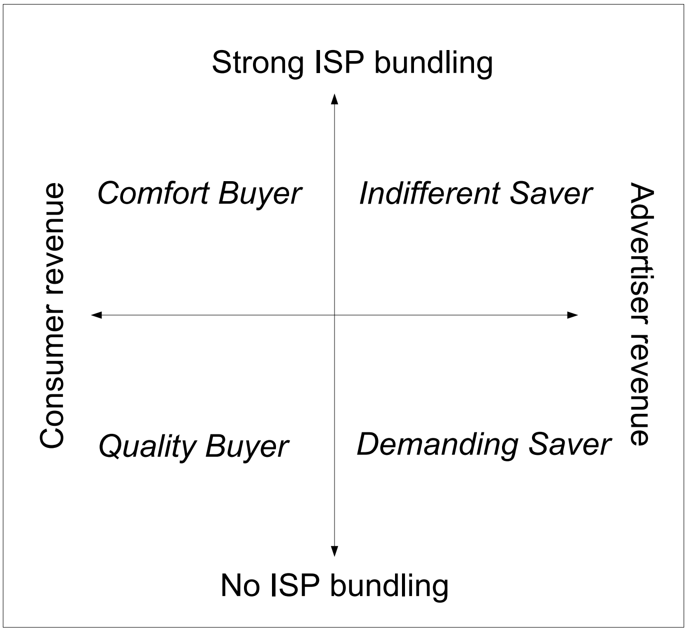

Figure 19. Scenario matrix.[27]

In the Comfort Buyer scenario, the end users are willing to pay for their comfort, i.e. they prefer bundled services and willingly pay for the content. Thus, the payment-based ISPs are dominating the Internet content delivery process by controlling the content delivery platform as well as the content itself. In other words, the ISP takes control of the content provision role and provides service bundles to end users. In addition, unlike the advertiser revenue model, the consumer revenue model does not necessarily require a large subscriber base. Thus, the first scenario is suitable also for smaller, local ISPs. The standardized SIM technology can be used to bill both the Internet access and the content consumption, which makes the payment process easy for the end users. In addition, caching and content provision contracts between the content owners and the ISPs enable the collection and analysis of usage statistics.

The end users are relatively indifferent about the service quality in the Indifferent Saver scenario and are not willing to pay for the content. However, the end users prefer to have bundled services due to the ease of usage. This means that the content is still provided by the ISP, but the ISP receives content revenue from advertisers. Because advertisers attempt to reach a wide audience, the ISP in this scenario tend to be a global actor that is not limited by the population of one country. Due to the large subscriber base, an advertising ISP can experience economies of scale from utilizing its own caching systems for internal traffic optimization and the role of standardization is diminished. In addition, larger ISPs have more bargaining power over content owners, who have limited control over their content in this scenario. On the other hand, advertising ISPs are in a good

[27] Adapted from Publication II. Reprinted with permission from Emerald Group Publishing Limited.

position to collect extensive usage statistics and possibly make them available also to content owners and advertisers.

In the Quality Buyer scenario, the end users know clearly what they want and make an effort in choosing services and paying for the content. Thus, payment-based content providers, such as Netflix or HBO, have the power to decide which Internet content delivery system to use. The payment based content providers experience strong network effects, resulting in a large content provider platform. However, if the network effects are not sufficiently high and the content provider is small, building its own billing and caching systems may not be feasible. Smaller content providers may use third parties for handling the payments, including credit card companies, banks and online payment providers. From the end-user perspective, the used payment system is important as credit cards are not available for everyone and third-party billing systems add a few extra steps in the payment process. In addition, the used payment scheme, i.e. pay-per-view or flat rate, also influences the end-user decision. In this scenario, the content provider can easily control the content access and collect usage statistics for meeting the end-user demand better.

The last scenario (Demanding Saver) is dominated by global content providers using advertising as their revenue model, such as Google. The end users know what they want and choose each service separately, but are not willing to pay for the content. Usage statistics play an important role in this scenario and the advertising content provider benefits from building a sophisticated content management system. The large scale of advertising content providers may lead to the technology they use, e.g. caching technology, to become de facto standards. On the other hand, the end-user QoE may be lowered due to the mandatory advertisements.

A summary of the scenario analysis together with a comparison of each scenario against certain general and industry properties are shown in Table 9. The general and industry properties reflect the important factors for each industry actors in the Internet content delivery market and are acquired from expert interviews during the market structure research. The comparison does not aim at ranking the scenarios, as the perception of the evaluation results depends on the actor in question. In addition, Table 9 shows each dominating actors' preference for the three caching technologies, as they can decide the winning caching technology in their corresponding scenario.

The results show that in-network caching, e.g. ICN, is a likely candidate for Comfort Buyer and Indifferent Saver scenarios due to the ISP's strong role in them. On the other hand, ISPs may also choose to build their own CDN service. Thus, CDNs – whether pure-play CDNs, content provider CDNs or ISP CDNs – have a strong position in all four scenarios. P2P technology seems the least preferred by the dominating actors. However, for an advertising content provider, who wishes to save on costs, P2P could be a good option.

Table 9. Summary and comparison of scenarios.[28]

Scenario	**Comfort Buyer**	**Indifferent Saver**	**Quality Buyer**	**Demanding Saver**
Dominating actor	Payment-based ISP	Advertising ISP	Payment-based content provider	Advertising content provider
		General properties		
Economies of scale	Low	Medium	Medium	High
Network effect	Low	Medium	Medium	High
Possibility of own billing mechanism	Medium	High	Low	High
Ease of paying for end users	Medium	Medium	Low	High
Amount of content usage statistics	Low	Medium	Medium	High
Importance of standardization	Medium	High	Low	Medium
		Industry properties		
Content provider control	Medium	Low	High	High
Bank's role	Medium	Low	High	Low
		Technology preferences		
CDN	High	Medium	High	High
P2P	Low	Low	Low	Medium
ICN	Medium	High	Low	Low

[28] Adapted from Publication II. Reprinted with permission from Emerald Group Publishing Limited.

5 . Cost efficiency of SDN in-network caching

The second part of the research utilizes the insights learned from the qualitative research in Section 4 and aims at quantitatively modeling the cost efficiency of in-network caching by using SDN-enabled caching as a use case. The following subsections first present the used cost model and the case description (Publication VI) and then elaborate on the performance of in-network caching as compared with that of data center caching (Publication VII). In addition, the cost efficiency of service function chaining is discussed, because caching is an essential part of the service function chain (Publication V).

5.1 Cost model

A Finnish reference LTE network is utilized to model the cost of building and operating a SDN-LTE network. The constructed cost model serves as the base model for analyzing 1) SDN in-network caching vs. data center caching (Section 5.2) and 2) SDN service chain vs. non-SDN service chain (Section 5.3). The inputs are gathered through semi-structured interviews during 2014 and 2015 with experts from all three Finnish MNOs and global network equipment providers operating in the Finnish market. The interviewees ranged from network directors to senior researchers and business developers. The topics discussed include the current LTE network structure, potential SDN architectures and quantitative data on the current Finnish networks. In addition, annual reports of the Finnish MNOs are utilized. This section describes the anticipated Finnish SDN-LTE topology and the cost model together with the assumptions adopted in constructing the model.

5.1.1 Case Finland

Three physical MNOs operate in Finland, each having approximately a third of the market. For modeling purposes, all mobile data connections are assumed to be LTE, though in reality older generation equipment are still in use in 2015. Thus, the modeled MNO is assumed to have one third of the Finnish market with approximately three million LTE subscriptions in 2015. The usage pattern is assumed to stay the same regardless of SDN and flat-rate pricing is assumed to remain popular among subscribers. Thus, the average revenue per user of approximately 17 euros per month is not expected to change because of SDN.

The anticipated Finnish SDN-LTE topology is illustrated in Figure 20, where the number of each mobile network element, i.e. eNB, gateways, mobility management and switches[29], used in the modeled MNO network is also shown. This number is directly adopted from the current non-SDN LTE topology, which is illustrated in Zhang et al. (2015), though SDN potentially reduces the demand for each element due to better network optimization. The current peering and transit agreements as well as roaming partnerships will also continue, as is shown in Figure 20.

The amount of the eNBs used in the modeled MNO network, 11000, is enough to serve many times more subscriptions in a more densely populated country (Naudts et al., 2012). However, the Finnish networks are coverage limited rather than capacity limited and a higher number of eNBs is needed to serve three million subscriptions in Finland. All eNBs are assumed to be macro sites, because only a small proportion of the Finnish LTE base stations are micro sites, though the amount is increasing in the near future. In addition, all eNB sites are assumed to have optical fiber or radio link connections with equivalent bandwidths to ensure high availability.

The level of centralization with the network elements is assumed to remain the same despite SDN, though the control plane functions are assumed to be centralized into data centers. Thus, the EPC elements, that are already running in three data centers in the non-SDN topology, are assumed to stay in these data centers in SDN-LTE. The user plane elements of the switch network and macro sites are to stay in their current positions despite of SDN. The SFC appliances are assumed to be virtualized onto general purpose hardware in SDN and running in the three data centers.

SDN controllers are new elements in the SDN-LTE topology. One backbone controller is assumed to manage ten backbone switches, whereas one eNB controller and one backhaul controller manages 100 eNBs and backhaul switches, respectively. In addition, a SDN enabled difference is the placement of the caching servers in the backhaul network and the macro sites, which is not possible in the current non-SDN topology. In the current LTE topology, only data center caching is feasible due to the tunneling functionalities of mobile networks.

[29] Both routers and switches are used in the mobile network. However, as the cost modeling focuses on one MNO network, for simplicity, all routers and switches are called switches in the quantitative part of the thesis.

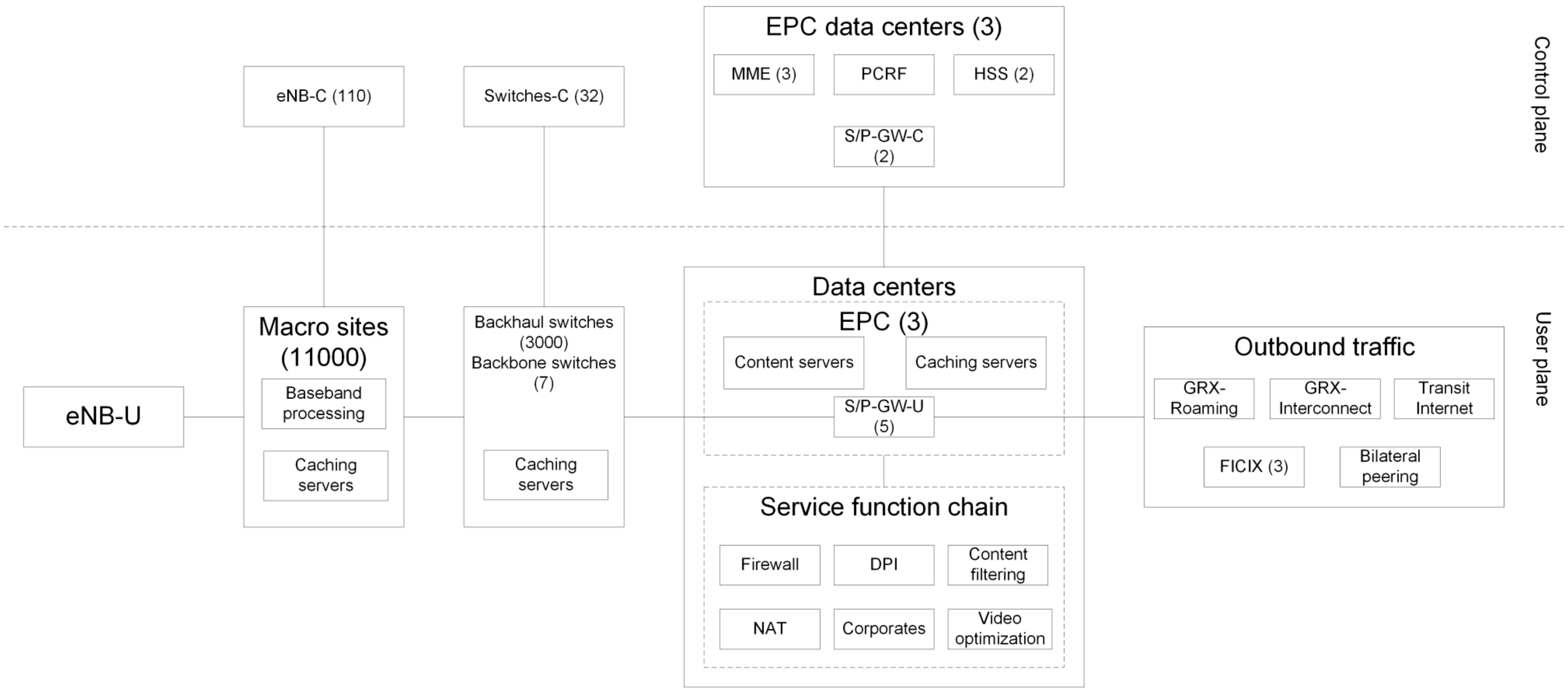

Figure 20. Anticipated Finnish SDN-LTE topology with caching and service function chain.[30]

[30] Adapted from Publication VI and modified. Reprinted with permission from IEEE.

5.1.2 Model assumptions

Figure 21 illustrates the comparative cost model used in this research. The main inputs for the cost model are the cost types influenced by SDN and caching, which are marked by a darker color in Figure 21 and explained in this section. The other cost types, such as marketing, personnel and other operating costs, are assumed to stay constant within the scope of the research and, thus, do not affect the outcome of the comparative cost modeling. However, the unchanging costs are included in the total network CAPEX and total network OPEX. In addition, SDN increases the amount of signaling traffic, but the volumes are considered insignificant compared to the user plane traffic and, thus, not taken into consideration in the cost model.

Network costs are divided into CAPEX and OPEX. Network CAPEX in this model considers the investment expenses of purchasing the network equipment, such as eNBs, gateways, switches and caches, as well as their deployment costs. The deployment expenses include the network configuration, e.g. defining the topology related parameters of the eNBs, and the on-site installation of the equipment. Configuration is done in a more centralized manner in SDN and the deployment costs are assumed to be lower compared with non-SDN LTE's deployment.

The number of each network element is acquired directly from the MNOs. Thus, the network capacities that typically determine the element demand, such as radio coverage, processing power and storage space, are not considered. The site infrastructure and cabling are reused in all scenarios and are, thus, not included in the comparative cost model. In addition, network planning before the purchase itself is not taken into consideration, because SDN does not influence the time taken to plan a network.

The one-time investment of the network CAPEX is annualized for the year 2015 for comparability to the OPEX, which is calculated over one year (2015). The Finnish MNOs on average use 15% of their revenue each year for new investments. Thus, the total CAPEX is scaled to the year 2015 by a factor so that the most costly scenario meets the 15% requirement.

Network OPEX is the cost of providing services to end users by running and managing the purchased network elements. The modeled OPEX includes energy consumption, site visits, network management and connectivity related expenses. Network management costs typically constitute a big part of the total network OPEX, but this model only considers the portion that is influenced by the higher automation level of SDN-LTE, i.e. fault detection and recovery from failure.

Energy costs include the energy consumed by the equipment and data centers, which is calculated from unit equipment energy consumption at full load. In addition, cooling consumes energy, which is taken as a percentage of the total energy consumption.

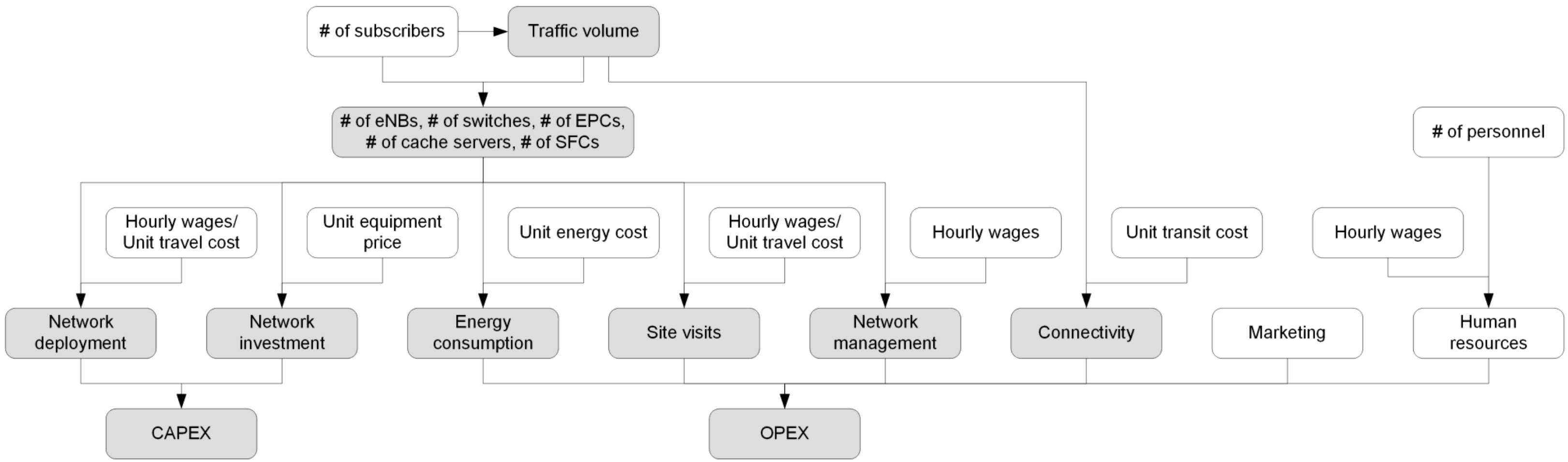

Figure 21. Cost model.[31]

[31] Adapted from Publication VII and modified. Reprinted with permission from John Wiley & Sons, Ltd.

Due to the distributed nature of eNBs, switches and the caches on site, they are usually not manned and require site visits for maintenance and fixing. The expenses include travel to and from the site, time spent on site and the difficulty of the task. As SDN removes intelligence from the eNBs and switches, i.e. they become less complex, site visits are assumed to decrease. On the other hand, caching at macro sites and the switch network increases the complexity of both, which increases the need for site visits.

The non-SDN scenarios are modeled with the estimated hardware performance, traffic volume and cost levels of 2015. The prices for the non-SDN LTE network elements are acquired from the interviews and reflect the current pricing level for Finnish MNOs. Due to the lack of pure SDN equipment, the SDN-LTE element prices are acquired by adjusting the non-SDN LTE prices with the complexity level change, standardization factor and the automation level. The values for the complexity level, standardization factor and automation level changes are presented in Publication VI together with an evaluation of their influence on the cost analysis of SDN-LTE compared with non-SDN LTE.

The SDN-LTE network elements (i.e. user plane) are assumed to be less complex in implementation and operation due to the removal of intelligence. However, requirements of high performance and availability keep local intelligence in certain network elements, such as eNBs and gateways, leading to the overall complexity decrease of network elements to be limited. In addition, data centers are assumed to become more complex due to the centralization of the control plane into them.

For virtualizing the SFC's user plane, general purpose hardware is assumed to be cheaper to purchase than dedicated hardware. However, the performance of general purpose hardware is also worse and more equipment is needed to match the same performance level of dedicated hardware. Thus, virtualization in SFC's user plane is assumed not to change the overall cost of SFC hardware. On the other hand, standardization lowers the price of the equipment due to economies of scale and increased competition. Thus, standardization's effect is taken into consideration for all the network elements, though for the EPC, it is assumed to be small due to the ongoing virtualization efforts in centralized EPCs.

Network load reduction due to in-network caching is determined by simulating the network. In addition, dynamic SFC enabled by SDN reduces traffic volumes passing through each SFC appliance, the percentages of which are acquired through expert interviews. The exact figures and the simulation setup are presented in the following subsections together with the results from the cost modeling.

5.2 Data center vs. in-network caching

The cost model presented in Section 5.1 is utilized to evaluate the cost efficiency of in-network caching compared with data center caching. Simulation results for network and gateway load of both the in-network and data center caching are used in the cost analysis. The results show that in-network caching saves both CAPEX and OPEX compared with data center caching. However, the cost efficiency of in-network caching is highly dependent on the content cacheability due to, for example, encrypted or dynamic content, as is shown with the sensitivity analysis.

5.2.1 Simulation setup

The simulations are run on a custom-made network simulator written in Python. The simulation area is a 4x4 kilometer square (i.e. 16 square kilometers) that has 16 LTE base stations with equal coverage size. The base stations are divided into four clusters, each of which is connected to two aggregation switches for redundancy. In addition, the network has one data center and two external interfaces to the public Internet. Caches of size 1GB are located at each base station and pre-aggregation switch, and the data center site has a cache of 5GB storage. The topology of the simulated network is illustrated in Figure 22 together with the link capacities connecting the sites.

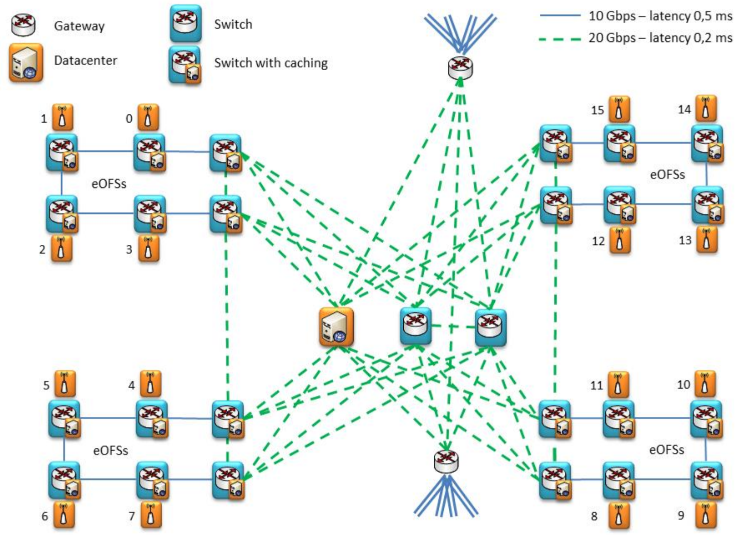

Figure 22. Topology of the simulated network.[32]

The simulation has 1000 end users, who request content from a pool of 100 000 items with a Zipf-like distribution. The end users are connected to their closest base station and have a uniformly distributed download speed of up to 200 Mbit/s. In addition, the end users are moving with a Gauss-Markov distribution between 0 km/h and 80 km/h. Base station handovers happen, when the end user leaves the first base station's coverage and is 100m into the second base station's coverage area. Between every request, the end user waits for a certain time that follows a Poisson distribution with a mean of 2 seconds.

The simulated caching system utilizes a central controller to make decisions on the optimal caching location. The caching location is decided based on the previous requests with an aim to minimize network load on every link. In addition, the simulator periodically relocates the content if more optimal locations

[32] Adapted from Publication VII. Reprinted with persmission from John Wiley & Sons, Ltd.

are found based on recent requests. The relocation interval vary between 15 and 90 seconds in the simulation. The content relocation process in the data center only caching scenario evaluates the optimality of each cached content, and the content items yielding lower load reduction values are removed from the cache.

The network load reduction results are illustrated in Figure 23 and Figure 24, respectively, as a function of how often the content is relocated in the network. The network load reduction of in-network caching compared with data center caching appears to be the highest, when six copies of the same content is allowed to be cached in the network. Due to the greater caching capacity in the in-network caching scenario, less traffic leaves and enters the MNO network, as shown by the gateway load reduction values. In addition, a higher relocation interval influences positively the load reduction in both the network and the gateway.

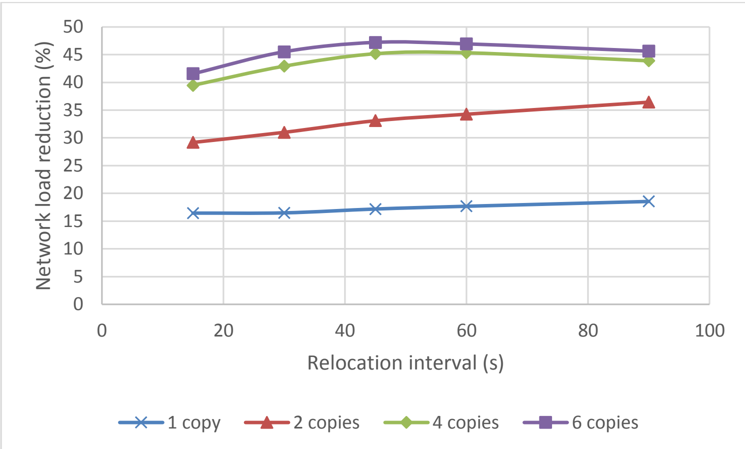

Figure 23. Network load reductions as a function of relocation interval.[33]

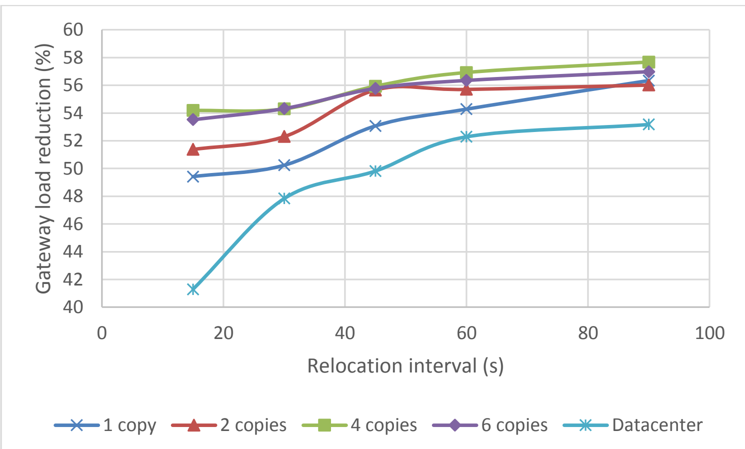

Figure 24. Gateway load reductions as a function of relocation interval.[34]

[33] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

[34] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

5.2.2 Cost modeling results

For determining the overall cost differences from adding cache servers into the network elements, load reduction values from the simulations are utilized. Minimum, median and maximum values of the 6-copies simulation results on gateway and network load reduction are used in the cost analysis. In addition, the load reduction values are scaled with the contracted cacheability of 85%. For example, a 45% simulated network load reduction scaled with the cacheability factor gives 39% actual load reduction, which translates directly into the network equipment demand while meeting the redundancy requirements. Similarly, a 55% gateway load reduction lead to a 47% drop in transit traffic volume.

The CAPEX changes from in-network caching compared with data center caching are shown in Table 10, where the MIN, MED and MAX columns show the results from using the minimum, median and maximum network load reduction values, respectively. The first four entries show the changes in investment costs for each network element including the investments of cache servers for each site. The comparison is done by estimating the overall investment cost of, e.g., 11 000 macro sites in data center caching and the same cost for in-network caching. The installation and configuration of the whole network (i.e. eNBs, switches, EPC and caches) is shown under one row: deployment. Total network CAPEX sums up all the investment and deployment costs of in-network caching and compares it to the overall CAPEX of data center caching.

The results show that the base station and backbone switch sites have lower investment costs when using in-network caching, whereas the addition of cache servers to backhaul sites increases the investment costs of the backhaul sites. This signifies that caching is cost efficient, only if the reduced equipment demand saves more than the additional cost from the cache servers, which is not the case in the backhaul network. However, in-network caching becomes less expensive in the backhaul network, when network load is reduced more. At the same time, the different network load reduction rates do not change the cost saving percentage in the backbone network. This is because the variation in the network load reduction rates is not high enough to either add or remove one more backbone switch.

Table 10. In-network caching CAPEX compared with data center caching CAPEX.[35]

CAPEX changes	MIN	MED	MAX
eNB macro sites	-0.4 %	-1.1 %	-1.4 %
Backbone switches	-27.4 %	-41.8 %	-41.8 %
Backhaul switches	64.4 %	55.7 %	52.3 %
EPC	0.0 %	0.0 %	0.0 %
Deployment	-14.8 %	-16.6 %	-17.3 %
Total network CAPEX	-0.6 %	-1.5 %	-1.9 %

As macro sites comprise over 95% of the investments and 80% of the deployment costs of a LTE network, the overall network CAPEX savings follow closely

[35] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

the cost savings from eNB investments. However, the cost saving from total network CAPEX is diverging from the eNB investment saving with higher network load reduction rates. In addition, EPC's CAPEX does not change, because its costs are based on the number of subscribers rather than the traffic volume.

In-network caching's network OPEX compared with data center caching's network OPEX is shown Table 11, where only the OPEX categories influenced by in-network caching are listed. The values for each row only show the cost differences for the cost category in question. The total network OPEX sums up both the affected and unaffected OPEX categories of in-network caching and compares it to the overall network OPEX of data center caching. Minimum, median and maximum network and gateway load reduction values are used for the MIN, MED and MAX columns.

Table 11. In-network caching OPEX compared with data center caching OPEX.[36]

OPEX changes	MIN	MED	MAX
Connectivity	-15.4 %	-8.4 %	-5.6 %
Energy	31.9 %	28.7 %	27.4 %
Site visit	-16.0 %	-17.8 %	-18.4 %
Network management	96.3 %	91.9 %	90.2 %
Total network OPEX	-0.2 %	-0.5 %	-0.6 %

Both data center caching and in-network caching reduces transit traffic volume, thus, in-network caching saves only between 5-15% on connectivity costs compared with data center caching. Due to reduced demand for network equipment, in-network caching reduces the need for site visits by over 16%. At the same time, energy consumption and network management costs are increasing due to the addition of cache servers into the network. As site visit costs constitute a big part of the modeled OPEX, in-network caching presents overall network OPEX savings. However, the total network OPEX saving from in-network caching is relatively small, because many OPEX categories are not influenced by in-network caching.

5.2.3 Sensitivity analysis

From caching's perspective, cacheability is an important factor in determining the network load and cost efficiency. The cost analysis results presented in Section 5.2.2 uses contracted cacheability, which is approximated to be 85%. However, the technical cacheability without caching agreements is only approximately 70% based on Ramanan et al.'s research [17]. In addition, the increase of encrypted content, such as HTTPS traffic, limits the cacheability and benefits for in-network caching. The MNOs estimate the HTTPS traffic volume to be 40% from the overall traffic volume in 2015, leaving only 60% of the traffic cacheable. Certain content types, such as dynamic content, further reduces cacheability and benefits from caching. Thus, the sensitivity analysis discusses

[36] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

technical cacheability (70%), HTTPS cacheability (60%) and content cacheability (assumed to be 50%). The median network and gateway load reduction values are used in the sensitivity analysis. In addition, all other parameter values are kept constant.

The sensitivity analysis results on CAPEX is illustrated in Figure 25, where cacheability's influence on in-network caching's equipment investment costs and total CAPEX are compared with that of the data center caching. The results show that in-network caching becomes less cost efficient, when MNOs do not have caching contracts with the content providers to ensure high cacheability. Similar patterns can be seen for network OPEX (Figure 26), where in-network caching becomes less cost efficient compared with data center caching, when no caching contract exists. However, in-network caching provides the highest connectivity related savings, when the cacheability is 60%, because peering through Internet exchange points are charged with a block pricing scheme.

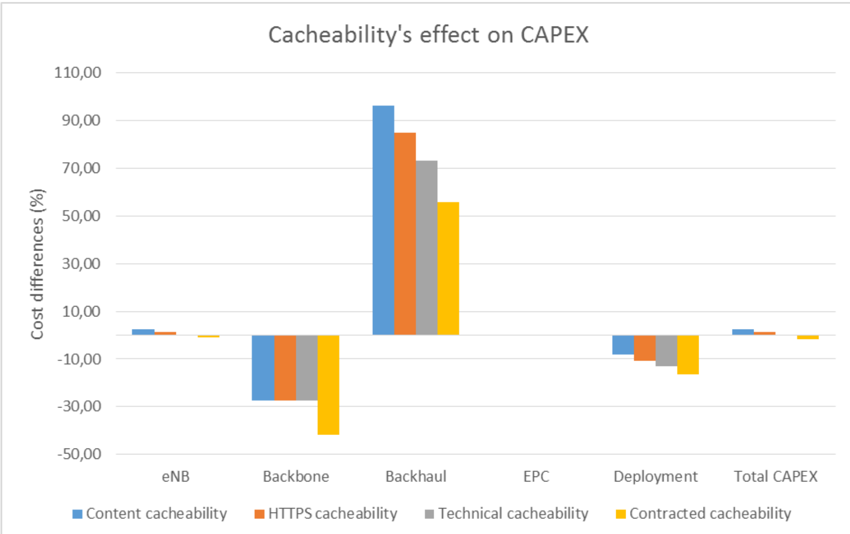

Figure 25. In-network caching CAPEX compared with data center caching CAPEX with different cacheabilities.[37]

[37] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

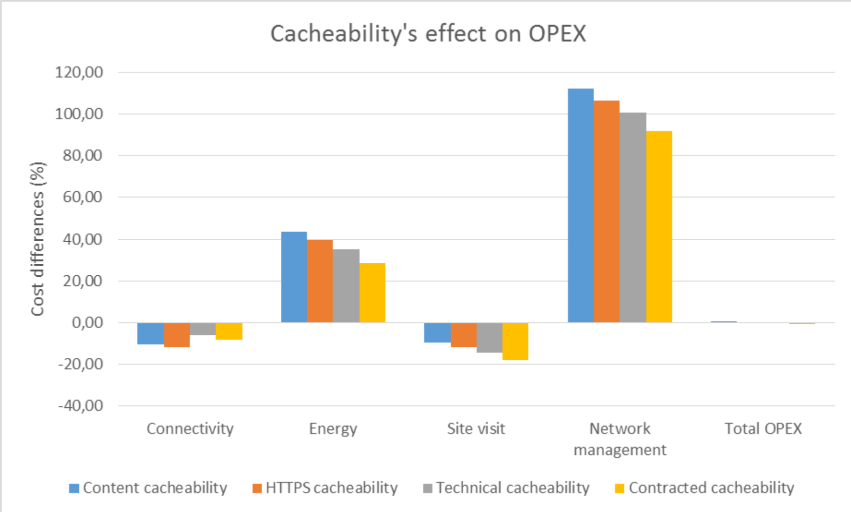

Figure 26. In-network caching OPEX compared with data center caching OPEX with different cacheabilities.[38]

5.3 Service function chaining

The cost efficiency of SDN enabled service function chaining (SFC) is also modelled, as caching and video optimization are important service functions that could benefit from a more dynamic service chain. As SDN dynamically steers traffic in the SFC, the traffic load in each of the SFC appliance is assumed to be lower, as is discussed in the following subsections. The results of the cost modeling show that SDN reduces both the CAPEX and OPEX of SFCs compared with the current hard-wired service functions. SDN's flexibility to add new services is not taken into consideration. Thus, the results are a baseline SFC cost analysis without any speculation on the potential new services.

5.3.1 Dynamic service function chaining

A service function chain is a set of services offered by the ISPs or MNOs that are traditionally physical boxes wired together in a specific order. These services range from a number of performance enhancement proxies (e.g. split TCP proxies, video optimizers, traffic redirection, etc.) to operator application platforms (e.g. IP multimedia subsystem and charging) and operate on the application specific traffic passing through them based on operator defined policies. In addition, service functions can be active (i.e. they change the packets that pass through) or passive (i.e. they only extract information from the packets without changing them).

When the service function boxes are wired together, all traffic passes through all the functions regardless of the need, which leads to overprovisioning of the service capacity. SDN and service function chain encapsulation from the IETF offers dynamic traffic steering on the granularity of each user and removes the

[38] Adapted from Publication VII. Reprinted with permission from John Wiley & Sons, Ltd.

need for reclassification between the functions. Similarly, Quinn and Guichard (2014) propose a service plane that utilizes SDN controllers and network service headers to dynamically add and remove service functions. In addition, service function virtualization is used to more flexibly scale and redimension the service functions. However, the cost modeling in this work does not reflect any single SFC implementation, but rather on a generic dynamic SFC concept that utilizes SDN and virtualization. A simplified figure of a service function chain, as defined by the IETF, is illustrated in Figure 27 (Halpern and Pignataro, 2015).

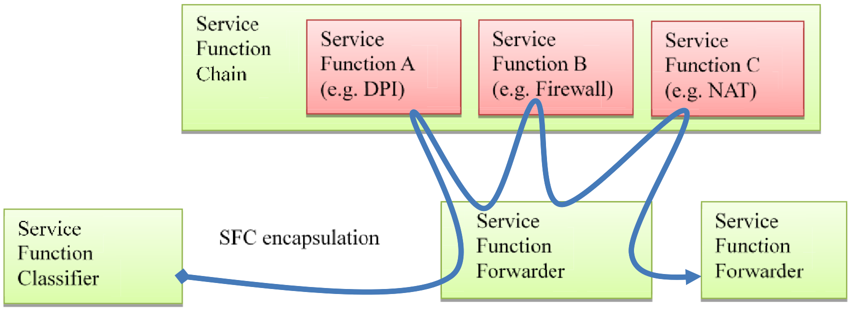

Figure 27. Service function chaining.[39]

A Service Function Classifier matches the incoming packets with the MNO's policies (based on e.g. user profile level, user equipment type and radio access type) and directs the traffic to the corresponding service function chains. The SFC encapsulation that contains a Service Path Identifier of the selected path and the relevant context information about the classification, such as information about the data plan of the user, is also added by the Service Function Classifier. The encapsulated packet is forwarded to a Service Function Forwarder that forwards the packets to a set of service functions that are associated with the Service Path Identifier in the encapsulation header. However, the service function's association with the Service Path Identifier can be dynamically changed on demand. When all the relevant service functions have processed the packet, the Service Function Forwarder forwards the packet to the next Service Function Forwarder, where the need for further processing is determined based on the information in the Service Path Identifier. After all necessary processing has been done, the Service Function Forwarder terminates the chain by removing the SFC encapsulation and forwards the packet to normal routing.

5.3.2 Cost model inputs

The service functions modeled in this research include NAT, firewall functionalities together with the intrusion detection and prevention system (IDPS), deep packet inspection (DPI), video and protocol optimization, content filtering and corporate services. SFC can also include other functions, but based on the interviews, the listed functions are considered the most important ones for MNOs.

[39] Adapted from Publication V.

In addition, DPI is assumed to operate as the Service Function Classifier and the existing server stacks are able to host Service Function Forwarders. The modeled service functions and their characteristics relevant for cost modeling are listed in Table 12.

The first three characteristics in Table 12 show the complexity and technical requirements of each service function, e.g. if the service function actively changes the packets or passively extracts information from them and whether the processing is limited by the number of packets, flows, connections or traffic throughput. These characteristics directly influence the cost of each service function. The SFC's total CAPEX (100%) is distributed in Table 12 among the different service functions, where video optimization and DPI have the biggest shares. The prices and capacities for the non-SDN SFC appliances are acquired from the interviews. Currently, approximately 2-3 boxes of each service function operate in each of the three EPC data centers. Though the SDN-SFC appliances are virtualized, the cost of acquiring the same processing capacity is assumed to be the same as non-SDN, because the lower cost of general purpose hardware is compensated with the higher number of hardware needed. However, cost savings can be achieved from dynamic traffic steering, which reduces the load through the SFC appliances and the capacity requirements.

In the non-SDN SFC, all incoming traffic passes through all the service functions, whereas SDN enables the forwarding of only relevant traffic to each of the service functions. The traffic that would go through each service function is acquired from MNO interviews based on estimations of their current service distribution. Thus, as shown in Table 12, when SDN is used, the traffic through the DPI is reduced by 20%, traffic through the video and protocol optimization is reduced by 40%, content filtering traffic is reduced by 80% and 80% less traffic goes through corporate services. Keeping in mind that DPI acts also as a Service Function Classifier, 100% of all traffic still goes through it, but the application and user specific usage of DPI can be lowered due to SDN; thus, the overall load is lowered by 20%.

Three sets of input values are defined for the uncertain parameters in the SDN-SFC scenario and listed in Table 13. The SDN compared with non-SDN traffic load through each SFC appliance is varied by ±10%. According to AT&T estimates (Thomas, 2015), the reduction in deployment times can be up to 95%, which is used in SDN-SFC 3, whereas the lower boundary of deployment time reduction is set at 50% to better see the effect of deployment time on cost. The reduction in recovery time due to SDN is varied by ±20% to evaluate the uncertainty.

Table 12. Characterization of modeled SFC appliances.[40]

Characteristics	Firewall & IDPS	NAT	DPI	Content filtering	Video & protocol optimization	Corporate services
OSI layer	1, 2, 3, 4, 7	3	2, 3, 4, 5, 6, 7	3, 4, 7	4, 6, 7	2, 3
Active (A) vs. Passive (P)	A	A	P	A	A	A
Processing	Firewall: Packet & Flow IDPS: Packet & Connection	Packet & Flow	Packet & Flow	Packet	Video: Flow Protocol: Packet	Packet & Flow & Connection
SFC cost distribution	20 %	15 %	20 %	10 %	30 %	5 %
SFC load - current	100 %	100 %	100 %	100 %	100 %	100 %
SFC load - SDN	-0 %	-0 %	-20 %	-80 %	-40 %	-80 %

[40] Adapted from Publication V.

Table 13. Input values for basic cost modelling.[41]

Input parameters	SDN-SFC 1	SDN-SFC 2	SDN-SFC 3
DPI load reduction	-10%	-20%	-30%
Content filtering load reduction	-70%	-80%	-90%
Video & protocol optimization load reduction	-30%	-40%	-50%
Corporate services load reduction	-70%	-80%	-90%
Deployment time reduction	-50%	-75%	-95%
Recovery time reduction	-50%	-70%	-90%

5.3.3 Cost modeling results

The cost model compares the CAPEX and OPEX of SDN SFC 1, SDN-SFC 2 and SDN-SFC 3 to that of the non-SDN SFC. The resulting CAPEX and OPEX differences are illustrated in Figure 28. SDN-SFC 2 shows the outcome from the interviewed values, whereas SDN-SFC 1 and SDN-SFC 3 can be considered as sensitivity analysis.

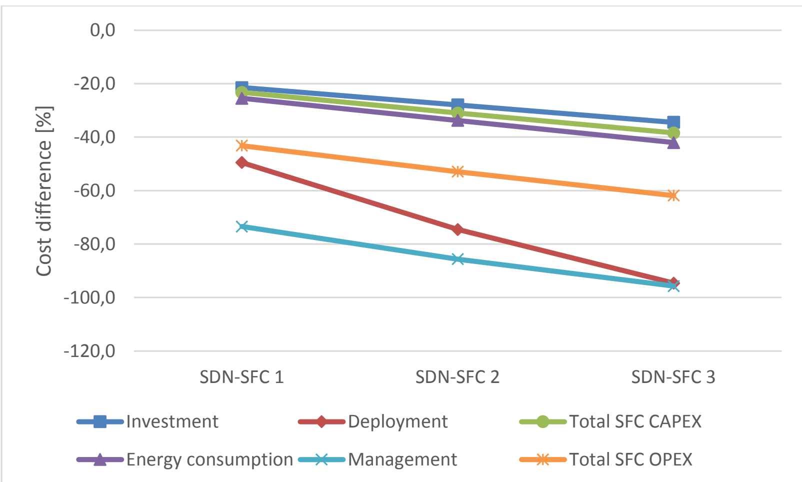

Figure 28. SFC CAPEX and OPEX changes compared with non-SDN SFC.[42]

Based on the results, SDN's impact on CAPEX and OPEX can be seen to be linear. SDN load reductions lower both the investment and deployment costs in SFC CAPEX and saves energy and SFC management costs in SFC OPEX. The reduced deployment time in SDN significantly reduces the deployment costs. In addition, increased automation level and decreased lead time in SDN further

[41] Adapted from Publication V.

[42] Adapted from Publication V.

lowers the management costs. The sensitivity analysis shows that a 10% increase in service function load leads to an approximately 6% increase in purchasing costs.

The SFC cost model is also used to evaluate the cost efficiency of different types of service providers. Four MNOs with different sevice offerings are considered: 1) operates in a country, where DPI is banned by the regulators, 2) targets only corporate customers, 3) offers only the basic connectivity service without value adding services, and 4) specializes in content delivery with a strong emphasis on video and protocol optimization. The SDN induced cost savings are illustrated in Figure 29. Noteworthy is that SDN reduces OPEX significantly regardless of the chosen MNO type.

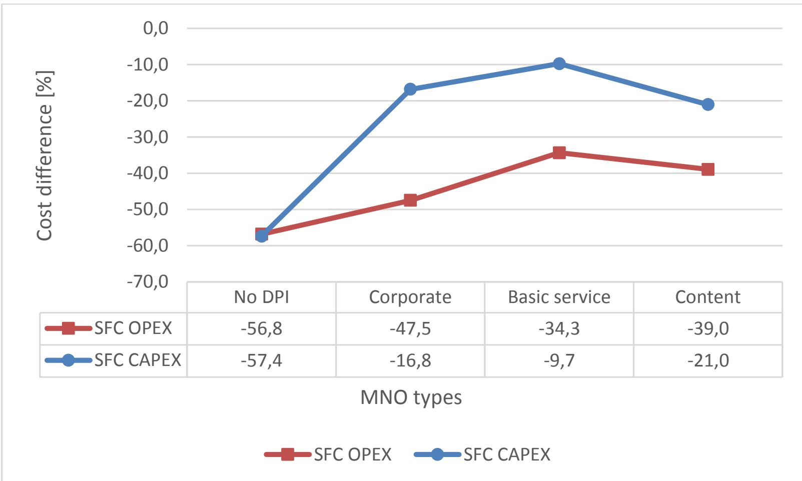

Figure 29. Cost savings for different types of MNOs.

6 . Discussion and conclusions

This research discussed the economic feasibility of alternative Internet content delivery technologies from the perspectives of the relevant actors. The results are especially valuable for the developers of in-network caching solutions as they point out the techno-economic limitations and issues of in-network caching. In addition, the future-oriented market analysis helps the actors involved in Internet content delivery and the regulatory authorities to make informed decisions.

6.1 Contributions and implications

The research aimed to evaluate the economic feasibility of in-network caching enabled by information-centric networking (ICN) or software-defined networking (SDN) as a means of efficient content distribution in the Internet. The results show that ICN and SDN in-network caching solutions can be economically feasible if the new technologies do not change the network topology drastically or the way the actors operate and interact with each other in the market. However, both technologies, especially ICN, can introduce major changes to the network and market dynamics. Thus, further quantitative research is needed, when ICN and SDN are more technically standardized. The qualitative results of this research and the proposed cost model can be employed as the baseline for future research.

This research makes three main contributions that reflect the three research objectives set in Section 1.2:

1) This study identified the main market roles and actors of Internet content delivery, analyzed their respective relationships and examined each of the actors' interest in adopting in-network caching.
2) This study identified the key trends and uncertainties, analyzed alternative future industry scenarios of Internet content delivery and proposed the most likely actor that drives each scenario.
3) This study evaluated the cost efficiency of SDN enabled in-network caching.

Contribution 1: The most important actors of Internet content delivery include the end user, content providers, fixed-access Internet service providers (ISPs), mobile network operators (MNOs), transit-ISPs and third-party content platforms, such as content delivery networks (CDNs). From the end-user perspective, caching improves the perceived user experience in the form of reduced latency. On the other hand, content providers also benefit from caching as caching reduces the traffic volume and congestion around the content providers' servers, leading to cost savings. The ISPs or MNOs, who connect the end users and content providers, see cost savings in transit traffic volumes, when caching is employed either in-network or in CDNs. Thus, content providers and ISPs or MNOs have an incentive to deploy in-network caching based on the actor analysis. However, the decision to invest depends also on other factors, such as the cost of additional cache servers and potential new revenue sources.

The popularity of CDNs is shown to benefit from the possibility of two-sided pricing and cross-side network effects. In addition, the CDN provider acts as a content platform that solves the access control, billing and coordination problems of content providers having to contract with several ISPs for better service levels. A contract-based SDN-enabled in-network caching platform that is controlled by the MNO could achieve the same two-sided market benefits as the CDN providers, and the MNOs have existing sophisticated mechanisms for billing its subscribers regardless of their location. On the other hand, ICN caching presents no significant two-sidedness due to the lack of a platform that can monetize from the network caches. In addition, the fixed-access Internet is less restricted compared to the mobile networks and the ISPs cannot guarantee that only authorized end users can access the content in the network caches and, thus, billing becomes an issue.

As SDN does not aim at changing the network topology and the basic routing principles of the Internet, several potential technical architectures and industry architectures could be proposed. The industry architectures illustrate a range of possibilities, where a traditional monolithic MNO can control the mobile Internet content delivery market, or parts of the network can be outsourced to other existing actors, such as connectivity providers and virtual network operators. In addition, SDN may also give rise to new actors in the market, such as the mobility provider, who would operate the mobility management and the control plane functions of a base station. The prevailing industry architecture for Internet content delivery depends on the MNO's trade-off between cost savings, lost control over the network functions and potential revenue sources (e.g. video-on-demand services) in each industry architecture.

Contribution 2: Tens of important market trends and uncertainties have been identified and the most important trends can be grouped under four major themes: 1) Growing content volume, 2) increasing demand for connectivity, 3) increasing importance of mobility, and 4) personalization of content. To efficiently deliver the increasing content volume, solutions like in-network caching are needed. Combined with the increasing importance of connectivity and mobility, the scale seems to tip towards SDN-enabled in-network caching. On the other hand, personalization of content limits the technical cacheability and business agreements are needed to improve cacheability.

Two most important uncertainties have been singled out to form the alternative industry scenarios. First is the ISP's/MNO's role in the content delivery process: will ISPs/MNOs stay as bit pipes or opt for a stronger presence in the content delivery process, e.g., through bundling. The second uncertainty discusses the content provider's revenue source: will advertising remain the main source of revenue or will paid content become more popular. Since the writing of Publication II, the market has developed towards strong ISP/MNO content bundling and the freemium payment model is gaining popularity. However, no obvious winner in either of the uncertainties exist yet.

The four scenarios are named after the prevailing end-user type in each scenario: 1) Comfort Buyer, 2) Indifferent Saver, 3) Quality Buyer, and 4) Demanding Saver. The prevailing end-user type also directly influences the type of actor that is likely to have a dominant position in that scenario and the power to decide on the used caching system. For example, when the end user prefers comfort over quality or cost, the end user will likely buy a content bundle from the ISP/MNO that offers connectivity, content and device. Thus, the dominant actors in each of the scenarios are 1) payment-based ISP/MNO, 2) advertising ISP/MNO, 3) payment-based content provider, and 4) advertising content provider.

The scenario analysis results show that in-network caching is a likely candidate for Comfort Buyer and Indifferent Saver scenarios due to the ISP's/MNO's strong role in them. On the other hand, ISPs/MNOs may also choose to build their own CDN service. Thus, CDNs – whether pure-play CDNs, content provider CDNs or ISP CDNs – have a strong position in all four scenarios. The scenario analysis results also indicate that the dominant actors prefer P2P technology the least. However, for a cost saving content provider, P2P is a viable option.

Contribution 3: An essential contribution of this research is the mapping of the current Finnish LTE topology, including the approximate numbers of each network elements. In addition, the most important network capital expenditure (CAPEX) and network operational expenditure (OPEX) categories of the Finnish MNOs are identified and a cost model is constructed based on these categories. The investment cost for each network element and the link capacities are also important findings of the research.

Based on the MNO's current topology and cost structure, an anticipated SDN-LTE topology and cost structure is projected. The results of the cost modeling show that SDN-enabled in-network caching, when caching agreements exist between the MNOs and content providers, saves both network CAPEX and network OPEX for the Finnish MNO compared with the currently adopted data center caching, such as CDNs. Assuming that ICN does not change the given LTE topology and caching agreements, and using the load reduction rates from Wang et al. (2014), ICN has similar potential for CAPEX and OPEX savings as SDN-enabled in-network caching. However, the topological assumption could not be made in the ICN case due to the uncertain direction of the technical development and lack of standardization.

This research, therefore, concludes that SDN-enabled in-network caching is cost efficient compared with current LTE and has the potential to enjoy the same two-sided benefits as the CDN solution. On the other hand, ICN with its different caching schemes provides higher technical performance (Wang et al., 2014) compared with the existing caching solutions, but needs to solve the coordination problems of e.g. caching contracts, access control and billing.

6.2 Limitations

This research utilizes qualitative and quantitative methods, such as value network analysis and techno-economic modeling, to model the economic feasibility of in-network caching. Similar to any scientific model, also the value networks and cost models employed in this research are simplifications of a more complex real world. These simplifications are an outcome of the problem and scope definition that are essential for achieving reliable and relevant results. In addition, certain assumptions are made about the technical details, such as the small role of signaling traffic in SDN in relation to the user plane traffic volume, which needs revisiting once the technology is better defined and standardized.

The future-oriented nature of the research limits the availability of quantitative data. For example, the ICN and SDN technologies have not been fully standardized yet, and several definitions and implementations exist. Thus, the reliability of the results are limited, because the technical details are still changing. For SDN related analysis, certain topological assumptions can be made based on the current topology, because the SDN technology does not aim at changing the network topology and structure. However, ICN changes the way routing is done in the Internet, which may require the reconstruction of the network topology. Thus, quantitative analysis of ICN caching could not be completed.

The cost modeling is done for a Finnish LTE topology for an average Finnish MNO due to the limited availability of input data. For example, the Finnish topology is coverage limited rather than capacity limited due to its large land area in relation to the population size. In addition, the cost of labor is very country-specific, and the cost of network equipment depends on the size of the MNO due to potential scale benefits. Thus, the generalizability of the results at best can be at the level of suburban areas in more developed countries.

Interviews with content providers are essential for a better understanding of the Internet content delivery market. However, most of the bigger content providers, such as Google and Netflix, are based outside Finland, which limits the selection of interviewee candidates from the content provider industry. For this research, few interviews with smaller Finnish content providers have been conducted, but the views and opinions are limited by their operation size and the Finnish regulations.

6.3 Future work

This research provides the basic understanding, tools and models for further analysis of the Internet content delivery market. For example, the cost model can be easily applied to other geographical areas and topology types, such as less developed countries and more capacity limited topologies, for a more conclusive analysis of the cost efficiency of the SDN-enabled in-network caching. In addition, the cost model can also be extended to evaluate the cost efficiency of ICN-enabled in-network caching, once the network topology of ICN can be defined. Similarly, the network simulations on in-network caching could be complemented by real network traces for more realistic numerical results.

The quantitative analysis in this research is limited to the perspective of the MNOs due to the growing importance of mobility and their control over the network locations, where in-network caches are to be installed. However, content providers, ISPs and CDNs also play important roles in the Internet content delivery market, and the market can function in an equilibrium, only if all actors are better off. Thus, the net benefit of in-network caching should be quantified also for the other actors.

In addition, the techno-economic modeling in this research limits only to cost analysis, whereas the potential to monetize the in-network caches is important for the investors of the technology. Thus, the revenue side of techno-economic modeling is a natural next step in the feasibility analysis of in-network caching. In addition, new service ideas and revenue models arising from in-network caches should also be included in the revenue modeling.

Finally, SDN enables faster service roll-out, dynamic service chaining and the technical ability of differential pricing to different subscription levels. This means, in theory, that the MNOs could improve the quality of experience of the higher paying subscribers by serving them from the in-network caches. The traffic from the lower subscription levels can then bypass the video optimization service function and be served with a lower priority. However, how much of this can be done within the boundaries of net neutrality principles should be researched.

References

3GPP. (2015a). General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access. 3GPP TS 23.401, release 13.2.0, March 19th, 2015. Accessed on July 23rd, 2015 at: http://www.3gpp.org/DynaReport/23401.htm.

3GPP. (2015b). Network architecture. 3GPP TS 23.002, release 13.2.0, June 21st, 2015. Accessed on September 2nd, 2015 at: http://www.3gpp.org/DynaReport/23002.htm.

Abrahamsson, H. & Nordmark, M. (2012). Program popularity and viewer behaviour in a large TV-on-demand system. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, pp. 199-210, Boston, MA, USA, November 14-16, 2012.

Ager, B., Schneider, F., Kim, J. & Feldmann, A. (2010). Revisiting cacheability in times of user generated content. In *Proceedings of 2010 INFOCOM IEEE Conference on Computer Communications Workshops*, San Diego, CA, USA, March 15-19, 2010.

Agyapong, P. & Sirbu, M. (2012). Economic incentives in information-centric nenworking: Implications for protocol design and public policy. IEEE *Communications Magazine*, 50(12), pp. 18–26.

Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D. & Ohlman, B. (2012). A survey of information-centric networking. *IEEE Communications Magazine*, 50(7), pp. 26–36.

Aleksic, S. & Miladinovic, I. (2014). Network virtualization: Paving the way to carrier clouds. In *Proceedings of 16th International Telecommunications Network Strategy and Planning Symposium*, Funchal, Portugal, September 17-19, 2014.

Allee, V. (2000a). Reconfiguring the value network. Jou*rnal of Business Strategy*, 21(4), pp. 36–39.

Allee, V. (2000b). The value evolution: Addressing larger implications of an intellectual capital and intangibles perspective. *Journal of Intellectual Capital*, 1(1), pp. 17–32.

Allee, V. (2008). Value network analysis and value conversion of tangible and intangible assets. *Journal of Intellectual Capital*, 9(1), pp. 5–24.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, 37(3), pp. 668–691.

Barish, G. & Obraczka, K. (2000). World Wide Web caching: Trends and techniques. *IEEE Communications Magazine*, 38(5), pp. 178–184.

Basta, A., Kellerer, W., Hoffmann, M., Morper, H. & Hoffmann, K. (2014). Applying NFV and SDN to LTE mobile core gateways, the functions placement problem. In *Proceedings of the 4th Workshop on All Things Cellular: Operations, Applications & Challenges*, pp. 33-38, Chicago, IL, USA, August 17-22, 2014.

Beck, L., Scherb, C., Sifalakis, M. & Tschudin, C.F. (2015). CCN-LITE, Version 0.3.0. Updated July, 2015. Accessed on September 1st, 2015 at: https://github.com/cn-uofbasel/ccn-lite/blob/master/README.md.

Blendin, J., Rückert, J., Leymann, N., Schyguda, G. & Hausheer, D. (2014). Position paper: Software-defined network service chaining. In *Proceedings of 2014 European Workshop on Software Defined Network*, pp. 109-114, Budapest, Hungary, September 1-3, 2014.

Bradfield, R., Wright, G., Burt, G., Cairns, G. & Van Der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures*, 37(8), pp. 795-812.

Braun, L., Klein, A., Carle, G., Reiser, H. & Eisl, J. (2012). Analyzing caching benefits for YouTube traffic in edge networks - A measurement-based evaluation. In *Proceedings of the 2012 IEEE Network Operations and Management Symposium*, pp. 311–318, Maui, Hawaii, USA, April, 16-20, 2012.

Brodersen, A., Scellato, S. & Wattenhofer, M. (2012). YouTube around the world: Geographic popularity of videos. In *Proceedings of the 21st International Conference on World Wide Web*, pp. 241–250, Lyon, France, April 16-20, 2012.

Bush, R. & Meyer, D. (2002). Some Internet architectural guidelines and philosophy. RFC 3439 (Informational), December, 2002: https://tools.ietf.org/rfc/rfc3439.txt.

Camarillo, G. (2009). Peer-to-peer (P2P) architecture: Definition, taxonomies, examples, and applicability. RFC 5694 (Informational), November 2009: https://tools.ietf.org/rfc/rfc5694.txt.

Carpenter, B. (1996). Architectural principles of the Internet. RFC 1958 (Informational), June, 1996: https://tools.ietf.org/rfc/rfc1958.txt.

Casey, T., Smura, T. & Sorri, A. (2010). Value network configurations in wireless local area access. In *Proceedings of the 9th Conference on Telecommunications, Media and Internet Techno-Economics*, Ghent, Belgium, June 7-9, 2010.

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y. & Moon, S. (2007). I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 1-14, San Diego, CA, USA, October 24-26, 2007.

Chen, Y., Katz, R. & Kubiatowicz, J. (2002). Dynamic replica placement for scalable content delivery. In *Proceedings of 1st International Workshop on Peer-to-Peer Systems* (IPTPS'02), pp. 306-318, Cambridge, MA, USA, March 7-8, 2002.

Chi, C. H., Liu, L. & Zhang, L. (2006). Quantitative analysis on the cacheability factors of web objects. In *Proceedings of International Computer Software and Applications Conference*, pp. 532–538, Chicago, IL, USA, September 17-21, 2006.

Cisco (2015). Cisco visual networking index: Forecast and methodology, 2014-2019. Updated May 27th, 2015. Accessed on June 14th, 2015 at: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

Civanlar, M.R. & Haskell, B.G. (1999). Client-server architecture using Internet and public switched networks. USA Patent 5995606, November 30th, 1999.

Clark, D., Lehr, W. & Bauer, S. (2011). Interconnection in the Internet: the policy challenge. In *Proceedings of 39th Research Conference on Communication, Information and Internet Policy*, Arlington, VA, USA, September 23-25, 2011.

Cohen, B. (2008). The BitTorrent protocol specification. Updated October 11th, 2013. Accessed on September 1st, 2015 at: http://www.bittorrent.org/beps/bep_0003.html.

Cooper, I., Melve, I. & Tomlinson, G. (2001). Internet web replication and caching taxonomy. RFC 3040 (Informational), January 2001: https://www.ietf.org/rfc/rfc3040.txt.

Costa-Requena, J. (2014). SDN integration in LTE mobile backhaul networks. In *Proceedings of International Conference on Information Networking*, pp. 264-269, Phuket, Thailand, February 12-14, 2014.

Costa-Requena, J., Kimmerlin, M., Manner, J. & Kantola, R. (2014). SDN optimized caching in LTE mobile networks. In *Proceedings of 2014 International Conference on Information and Communication Technology Convergence*, pp. 128-132, Busan, South Korea, October 22-24, 2014.

Costa-Requena, J., Guasch, V.F. & Santos, J.L. (2015a). Software defined networks based 5G backhaul architecture. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, Bali, Indonesia, January 8-10, 2015.

Costa-Requena, J., Santos, J.L., Guasch, V.F., Ahokas, K., Premsankar, G., Luukkainen, S., Ahmad, I., Liyanage, M., Ylianttila, M., Lopez, O.R., Uriarte, M.I. & Montes de Oca, E. (2015b). SDN and NFV integration in generalized mobile network architecture. In *Proceedings of European Conference on Networks and Communications*, pp. 154-158, Paris, France, June 29-July 2, 2015.

Costa-Requena, J., Santos, J.L. & Guasch, V.F. (2015c). Mobile backhaul transport streamlined through SDN. In *Proceedings of 17th International Conference on Transparent Optical Networks*, Budapest, Hungary, July 5-9, 2015.

Dannewitz, C. & Herlich, M. (2011). OpenNetInf documentation, design and implementation. Technical report TR-RI-11-314, University of Paderbon. Accessed on September 1st, 2015 at: https://code.google.com/p/opennetinf/downloads/detail?name=NetInf_documentation_Sep-2011.pdf.

Dannewitz, C., Kutscher, D., Ohlman, B., Farrell, S., Ahlgren, B. & Karl, H. (2013). Network of information (NetInf) – An information-centric networking architecture. *Computer Communications*, 36(7), pp. 721-735.

Dilley, J., Maggs, B., Parikh, J., Prokop, H., Sitaraman, R. & Weihl, B. (2002). Globally distributed content delivery. *IEEE Internet Computing*, 6(5), pp. 50-58.

Doria, A., Hadi Salim, J., Haas, R., Khosravi, H., Wang, W., Dong, L., Gopal, R. & Halpern, J. (2010). Forwarding and control element separation (ForCES) protocol specification. RFC 5810 (Standards Track), March, 2010: https://tools.ietf.org/rfc/rfc5810.txt.

EdgeCast. (2015). Licensed CDN. Accessed on June 16th, 2015 at: http://www.edgecast.com/solutions/licensed-cdn/.

Elkadi, I. (2012). Impact of content caching on competitive dynamics of Internet content delivery ecosystem. Master's thesis, School of Electrical Engineering, Aalto University.

Enns, R. Bjorklund, M., Schoenwaelder, J. & Bierman, A. (2011). Network configuration protocol (NETCONF). RFC 6241 (Standards Track), June, 2011: https://tools.ietf.org/rfc/rfc6241.txt.

ETSI, 2015. Virtual EPC with SDN function in mobile backhaul networks. ETSI NFV Proof-of-Concept, July, 2015: http://nfvwiki.etsi.org/index.php?title=Virtual_EPC_with_SDN_Function_in_Mobile_Backhaul_Networks.

Faratin, P. (2007). Economics of overlay networks: An industrial organization perspective on network economics. In *Proceedings of the Joint Workshop on the Economics of Networked Systems and Incentive-based Computing, in Conjunction with ACM Conference on Electronic Commerce*, San Diego, CA, USA, June 11-15, 2007.

Fielding, R., Nottingham, M. & Reschke, J. (2014). Hypertext transfer protocol (HTTP/1.1): Caching. RFC 7234 (Standards Track), June, 2014: https://tools.ietf.org/rfc/rfc7234.txt.

Fotiou, N., Trossen, D. & Polyzos, G. (2012). Illustrating a publish-subscribe Internet architecture. *Telecommunication Systems*, 51(4), pp. 233-245.

Frank, B., Poese, I., Lin, Y., Smaragdakis, G., Feldmann, A., Maggs, B., Rake, J., Uhlig, S. & Weber, R. (2013). Pushing CDN-ISP collaboration to the limit. *ACM SIGCOMM Computer Communication Review*, 43(3), pp. 34-44.

Giraldo, C., Gil-Castineira, F., Lopez-Bravo, C. & Gonzales-Castano, F.J. (2014). A software-defined mobile network architecture. In *Proceedings of IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 287-291, Larnaca, Cyprus, October 8-10, 2014.

Gutierrez, P.A. & Carapinha, J. (2011). Cloud networking: Implications of agile virtualization on provider relationships. In *Proceedings of Workshop on Challenges and Solutions for Network Virtualization,* Kiel, Germany, March 8-11, 2011.

Gwertzman, J.S. & Seltzer, M. (1995). The case for geographical push-caching. In *Proceedigs of Fifth Workshop on Hot Topics in Operating Systems*, pp. 51-55, Orcas Island, WA, USA, May 4-5, 1995.

Hagiu, A. & Yoffie, D. (2009). What's your Google strategy. *Harvard Business Review*, April, 2009, pp. 74-81.

Haleplidis, E., Salim, J. H., Denazis, S. & Koufopavlou, O. (2015a). Towards a network abstraction model for SDN. *Journal of Network and Systems Management*, 23(2), pp. 309–327.

Haleplidis, E., Salim, J. H., Halpern, J. M., Hares, S., Pentikousis, K., Ogawa, K., Wang, W., Denazis, S. & Koufopavlou, O. (2015b). Network programmability with ForCES. *IEEE Communications Surveys & Tutorials*, 17(3), pp. 1423–1440.

Haleplidis, E., Pentikousis, K., Denazis, L, Salim, J. H., Meyer, D. & Koufopavlou, O. (2015c). Software-defined networking (SDN): Layers and architecture terminology. RFC 7426 (Informational), January, 2015: https://tools.ietf.org/rfc/rfc7426.txt.

Halpern, J. & Pignataro, C. (Eds.) (2015). Service function chaining (SFC) architecture. RFC 7665 (Informational), October, 2015: https://tools.ietf.org/rfc/rfc7665.txt.

Harno, J., Katsianis, D., Smura, T., Eskedal, T.-G., Venturin, R., Pohjola, O. P., Kumar, K. R. R. & Varoutas, D. (2009). Alternatives for mobile operators in the competitive 3G and beyond business. *Telecommunication Systems*, 41(2), pp. 77–95.

Haßlinger, G. & Hartleb, F. (2011). Content delivery and caching from a network provider's perspective. *Computer Networks*, 55(18), pp.3991–4006.

Hastings, R. & Wells, D. (2015). Netflix: Q1 Letter to shareholders, April 15th, 2015. Accessed on June 14th, 2015 at: http://files.shareholder.com/downloads/NFLX/266140950x0x821407/DB785B50-90FE-44DA-9F5B-37DBF0DCD0E1/Q1_15_Earnings_Letter_final_tables.pdf.

Heikkinen, M.V.J. & Hämmäinen, H. (2008). Scenario planning of mobile peer-to-peer service usage. In *Proceedings of 7th International Conference on Mobile Business*, pp. 145-152, Barcelona, Spain, July 7-8, 2008.

Hu, X., Gong, J., Cheng, G. & Fan, C. (2015). Enhancing in-network caching by coupling cache placement, replacement and location. In *Proceedings of 2015 IEEE International Conference on Communications*, pp. 5672-5678, London, UK, June 8-12, 2015.

ITU-T. (2014). Framework of software-defined networking. International Telecommunication Union Telecommunication Standardization Sector's Recommendation Y.3300, 2014.

Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M. F., Briggs, N. H. & Braynard, R. L. (2012). Networking named content. *Communications of the ACM*, 55(1), pp. 117–124.

Jennings, C., Lowekamp, B., Rescorla, E., Baset, S. & Schulzrinne, H. (2014). Resource location and discovery (RELOAD) base protocol. RFC 6940 (Standards Track), January, 2014: https://tools.ietf.org/rfc/rfc6940.txt.

Jia, X., Li, D., Hu, X., Wu, W. & Du, D. (2003). Placement of web-server proxies with consideration of read and update operations on the Internet. *The Computer Journal*, 46(1), pp. 1–14.

Järvinen, P. (2004). On research methods. Opinpajan kirja, Tampere.

Kantor, M., Wajda, K., Lannoo, B., Casier, K., Verbrugge, S., Pickavet, M., Wosinska, L., Chen, J. & Mitcsenkov, A. (2010). General framework for techno-economic analysis of next generation access networks. In *Proceedings of 12th International Conference on Transparent Optical Networks*, Munich, Germany, June 27-July 1, 2010.

Karlson, B., Bria, A., Lönnqvist, P., Norlin, C. & Lind, J. (2003). Wireless foresight: Scenarios of the mobile world in 2015. Chichester, UK: Wiley, 2003.

Katsaros, K., Xylomenos, G. & Polyzos, G. C. (2011). MultiCache: An overlay architecture for information-centric networking. *Computer Networks*, 55(4), pp. 936–947.

Klingberg, T. & Manfredi, R. (2002). Gnutella 0.6. Updated June 2002. Accessed on September 1st, 2015 at: http://rfc-gnutella.sourceforge.net/src/rfc-0_6-draft.html.

Knoll, T.M. (2015). Life-cycle cost modelling for NFV/SDN based mobile networks. In *Proceedings of 12th Conference of Telecommunication, Media and Internet Techno-Economics*, Munich, Germany, November 9-10, 2015.

Konidaris, S. (1991). The RACE programme - Research for advanced communications in Europe. In *Proceedings of Global Telecommunications Conference GLOBECOM*, pp. 1496-1500, Phoenix, AZ, USA, December 2-5, 1991.

Koponen, T., Chawla, M., Chun, B.-G., Ermolinskiy, A., Kim, K. H., Shenker, S. & Stoica, I. (2007). A data-oriented (and beyond) network architecture. *ACM SIGCOMM Computer Communication Review, 37(4)*, pp. 181-192.

Kreutz, D., Ramos, F., Verissimo, P. E., Rothenberg, C. E., Azodolmolky, S. & Uhlig, S. (2015). Software-defined networking: A comprehensive survey. *Proceedings of IEEE*, 103(1), pp. 14–76.

Krishnan, P., Raz, D. & Shavitt, Y. (2000). The cache location problem. *IEEE/ACM Transactions on Networking*, 8(5), pp. 568–582.

Labovitz, C., Iekel-Johnson, S., McPherson, D., Oberheide, J. & Jahanian, F. (2010). Internet inter-domain traffic. *ACM SIGCOMM Computer Communication Review*, 40(4), pp. 75–86.

Laffont, J.-J., Marcus, S., Rey, P. & Tirole, J. (2003). Internet interconnection and the off-net-cost pricing principle. *The RAND Journal of Economics*, 34(2), pp. 370-390.

Li, Y., Xie, H., Wen, Y., Chow, C. & Zhang, Z. (2015). How much to coordinate? Optimizing in-network caching in content-centric networks. *IEEE Transactions on Network and Service Management*, 12(3), 420-434.

March, S. & Smith, G. (1995). Design and natural science research on information technology. *Decision support systems*, 15(1995), pp. 251-266.

McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S. & Turner, J. (2008). OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 38(2), pp. 69–74.

Michel, S., Nguyen, K., Rosenstein, A., Zhang, L., Floyd, S. & Jacobson, V. (1998). Adaptive web caching: towards a new global caching architecture. *Computer Networks and ISDN Systems*, 30(22-23), pp. 2169-2177.

Miranda, L., Santos, R. & Laender, A. (2013). Characterizing video access patterns in mainstream media portals. In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1085–1092, Rio de Janeiro, Brazil, May 13-17, 2013.

Mosko, M., Solis, I. & Wood, C. (2016). CCNx semantics. Internet-Draft: https://tools.ietf.org/html/draft-irtf-icnrg-ccnxsemantics-01, work in progress, exprires in July, 2016.

Naudts, B., Kind, M., Westphal, F.-J., Verbrugge, S., Colle, D. & Pickavet, M. (2012). Techno-economic analysis of software defined networking as architecture for the virtualization of a mobile network. In *Proceedings of 2012 European Workshop on Software Defined Networking*, pp. 67–72, Darmstadt, Germany, October 25-26, 2012.

Naylor, D., Finamore, A., Leontiadis, I., Grunenberger, Y., Mellia, M., Munafò, M., Papagiannaki, K. & Steenkiste, P. (2014). The cost of the "S" in HTTPS. In *Proceedings of the 10th ACM International Conference on Emerging Networking Experiments and Technologies*, pp. 133–140, Sydney, Australia, December 2-5, 2014.

Netflix. (2015). Netflix open connect content delivery for ISPs. Accessed on June 14th, 2015 at: https://openconnect.itp.netflix.com/deliveryOptions/index.html.

Nguyen, V. & Kim, Y. (2015). Proposal and evaluation of SDN-based mobile packet core networks. *EURASIP Journal on Wireless Communications and Networking*, December 2015(172).

Nohrborg, M. (2015). LTE overview. Accessed on July 23rd, 2015 at: http://www.3gpp.org/technologies/keywords-acronyms/98-lte.

Norton, W. (2011). The Internet peering playbook: Connecting to the core of the Internet. DrPeering Press.

Open networking foundation. (2014). OpenFlow-enabled SDN and network functions virtualization, February 17th, 2014. Accessed on June 14th, 2015 at: https://www.opennetworking.org/images/stories/downloads/sdn-resources/solution-briefs/sb-sdn-nvf-solution.pdf.

Parker, G. G. & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 51(10), pp. 1494–1504.

Pathan, M. & Buyya, R. (2008). A Taxonomy of CDNs. In: Buyya, R., Pathan, M. & Vakali, A. (eds.). Content delivery networks, Springer Berlin Heidelberg.

Pentikousis, K., Wang, Y. & Hu, W. (2013). MobileFlow: Toward software-defined mobile networks. *IEEE Communications Magazine*, 51(7), pp. 44–53.

Peterson, L. & Davie, B. (2014). Framework for content distribution network interconnection (CDNI). RFC 7336 (Informational), August, 2014: http://tools.ietf.org/rfc/rfc7336.txt.

Porter, M. E. (1985). Competitive advantage: Creating and sustaining superior performance. The Free Press, New York, NY, USA.

Psaras, I., Chai, W. & Pavlou, G. (2012). Probabilistic in-network caching for information-centric networks. In *Proceedings of the Second ACM SIGCOMM Workshop on Information-Centric Networking*, pp. 55–60, Helsinki, Finland, August 17, 2012.

Pulkkanen, A. & Seppänen, M. (2012). Freemium business models in technology product markets. In *Proceedings of the International Society for Professional Innovation Management Conference*, Barcelona, Spain, June 17-20, 2012.

Pursuit. (2013). Software. Accessed on September 1st, 2015 at: http://www.fp7-pursuit.eu/PursuitWeb/?page_id=12.

Quinn, P. & Guichard, J. (2014). Service function chaining: Creating a service plane via network service headers. *Computer*, 47(11), pp. 38–44.

Raghavan, B., Casado, M., Koponen, T., Ratnasamy, S., Ghodsi, A. & Shenker, S. (2012). Software-defined Internet architecture: Decoupling architecture from infrastructure. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pp. 43–48, Redmond, WA, USA, October 29-30, 2012.

Ramanan, B. A., Drabeck, L. M., Haner, M., Nithi, N., Klein, T. E. & Sawkar, C. (2013). Cacheability analysis of HTTP traffic in an operational LTE network. In *Proceedings of Wireless Telecommunications Symposium*, Phoenix, AZ, USA, April 17-19, 2013.

Rayna, T. (2008). Understanding the challenges of the digital economy: The nature of digital goods. *Communications & Strategies*, 3rd Quarter(71), September 2008, pp. 13–36.

Risson, J. & Moors, T. (2007). Survey of research towards robust peer-to-peer networks: Search methods. RFC 4981 (Informational), September, 2007: http://tools.ietf.org/rfc/rfc4981.txt.

Rochet, J.-C. & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of European Economic Association*, 1(4), pp. 990-1029.

Rochet, J. & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics*, 37(3), pp. 645–667.

Rowstron, A. & Druschel, P. (2001). Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In *Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms*, pp., 329-350, Heidelberg, Germany, November 12-16, 2001.

Salsano, S., Blefari-Melazzi, N., Presti, F. Lo, Siracusano, G. & Ventre, P. L. (2014). Generalized virtual networking: An enabler for service centric networking and network function virtualization. In *Proceedings of 16th International Telecommunications Network Strategy and Planning Symposium*, Funchal, Portugal, September 17-19, 2014.

Schoemaker, P.J.H. (1991). When and how to use scenario planning: A heuristic approach. *Journal of Forecasting*, 10(6), pp. 549-564.

Schoemaker, P.J.H. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal*, 14(3), pp. 193-213.

Schoemaker, P.J.H. (1995). Scenario planning : A tool for strategic thinking. *Sloan Management Review*, 36(2), pp. 25-40.

Schoemaker, P.J.H. & Mavaddat, V.M. (2000). Scenario planning for disruptive technologies. In: Day, G.S. & Schoemaker P.J.H. (eds.). Wharton on managing emerging technologies, pp. 206-241, John Wiley & Sons, Ltd, New York.

Schollmeier, R. (2002). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Proceedings of the First International Conference on Peer-to-Peer Computing*, pp. 101-102, Linkoping, Sweden, August 27-29, 2002.

Simon, H. (1996). The sciences of the artificial, 3rd ed. The MIT Press, Cambridge.

Smura, T. (2012). Techno-economic modelling of wireless network and industry architectures. Doctoral thesis, Aalto University.

Smura, T., Kiiski, A. & Hämmäinen, H. (2007). Virtual operators in the mobile industry: A techno-economic analysis. *NETNOMICS: Economic Research and Electronic Networking*, 8(1-2), pp. 25–48.

Smura, T. & Sorri, A. (2009). Future scenarios for local area access: Industry structure and access fragmentation. In *Proceedings of the Eighth International Conference on Mobile Business*, pp. 57-63, Dalian, China, June 26-28, 2009.

Sourlas, V., Flegkas, P., Paschos, G., Katsaros, D. & Tassiulas, L. (2011). Storage planning and replica assignment in content-centric publish/subscribe networks. *The*

International Journal of Computer and Telecommunications Networking, 55(18), pp. 4021-4032.

Sourlas, V., Gkatzikis, L., Flegkas, P. & Tassiulas, L. (2013). Distributed cache management in information-centric networks. *IEEE Transactions on Network and Service Management*, 10(3), pp. 286-299.

Spotify. (2015). Information. Accessed on June 14th, 2015 at: https://press.spotify.com/fi/information/.

Stabell, C. B. & Fjeldstad, Ø. D. (1998). Configuring value for competitive advantage: On chains, shops, and networks. *Strategic Management Journal*, 19(5), pp. 413–437.

Stoica, I., Morris, R., Liben-Nowell, D., Karger, D.R., Kaashoek, M.F., Dabek, F. & Balakrishnan, H. (2003). Chord: A scalable peer-to-peer lookup protocol for Internet applications. *IEEE/ACM Transactions on Networking*, 11(1), pp. 17–32.

Taleb, T. (2014). Toward carrier cloud: Potential, challenges, and solutions. *IEEE Wireless Communications*, 21(3), pp. 80–91.

Taleb, T., Corici, M., Parada, C., Jamakovic, A., Ruffino, S., Karagiannis, G. & Magedanz, T. (2015). EASE: EPC as a service to ease mobile core network deployment over cloud. *IEEE Network*, 29(2), pp. 78–88.

Tang, X. & Chanson, S.T. (2002). Coordinated en-route web caching. *IEEE Transactions on Computers,* 51(6), pp. 595–607.

Thomas, S. (2015). AT&T: SDN is slashing provisioning cycle times by up to 95%, August 12th, 2015. Accessed on December 21st, 2015 at: http://www.lightreading.com/carrier-sdn/sdn-architectures/atandt-sdn-is-slashing-provisioning-cycle-times-by-up-to-95-/d/d-id/717582.

Tourrilhes, J., Sharma, P., Banerjee, S. & Pettit, J. (2014). SDN and OpenFlow evolution: A standards perspective. *Computer*, 47(11), pp. 22–29.

Tuffin, B. (2015). Side payments as barriers to entry in non-neutral networks. In *Proceedings of the 12th Conference of Telecommunication, Media and Internet Techno-Economics*, Munich, Germany, November 9-10, 2015.

Vakali, A. & Pallis, G. (2003). Content delivery networks: Status and trends. *IEEE Internet Computing*, *7*(6), pp. 68–74.

Veitch, P., McGrath, M. J. & Bayon, V. (2015). An instrumentation and analytics framework for optimal and robust NFV deployment. *IEEE Communications Magazines*, 53(2), pp. 126-133.

Verbrugge, S., Colle, D., Pickavet, M., Demeester, P., Pasqualini, S., Iselt, A., Kirstädter, A., Hülsermann, R., Westphal, F.-J. & Jäger, M. (2006). Methodology and input availability parameters for calculating OpEx and CapEx costs for realistic network scenarios. *Journal of Optical Networking*, 5(6), pp.509–520.

Wagner, T. M., Benlian, A. & Hess, T. (2013). The advertising effect of free - Do free basic versions promote premium versions within the freemium business model of music services? In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 2928–2935, Maui, Hawaii, USA, January 7-10, 2013.

Wack, P. (1985). Scenarios: Uncharted waters ahead. *Harvard Business Review*, 63(5), pp. 73-89.

Wang, H., Chen, S., Xu, H., Ai, M. & Shi, Y. (2015). SoftNet: A software defined decentralized mobile network architecture toward 5G. *IEEE Network*, 29(2), pp. 16-22.

Wang, X., Chen, M., Taleb, T., Ksentini, A. & Leung, V. C. M. (2014). Cache in the air: Exploiting content caching and delivery techniques for 5G systems. *IEEE Communications Magazine*, 52(2), pp. 131–139.

Wessels, D. & Claffy, K. (1997). Internet cache protocol (ICP). RFC 2186 (Informational), September 1997: https://tools.ietf.org/rfc/rfc2186.txt.

Wischik, D., Handley, M. & Braun, M.B. (2008). The resource pooling principle. *ACM SIGCOMM Computer Communication Review*, 38(5), pp.47–52.

Xu, J., Li, B. & Lee, D. (2002). Placement problems for transparent data replication proxy services. *IEEE Journal on Areas in Communications*, 20(7), pp. 1383–1398.

Zhang, N. & Hämmäinen, H. (2015). Cost efficiency of SDN in LTE-based mobile networks: Case Finland. In *Proceedings of Workshop on Software-Defined Networking and Network Function Virtualization for Flexible Network Management*, Cottbus, Germany, March 12, 2015.

Zhang, L., Afanasyev, A. Burke, J., Jacobson, V., Claffy, K. C., Crowley, P., Papadopoulos, C., Wang, L. & Zhang, B. (2010). Named data networking. *ACM SIGCOMM Computer Communication Review*, 44(3), pp. 66-73.

Zhao, H. (2008). Emerging business models of the mobile internet market. Master's thesis, Department of Communications and Networking, Helsinki University of Technology.

Zink, M., Suh, K., Gu, Y. & Kurose, J. (2009). Characteristics of YouTube network traffic at a campus network – Measurements, models, and implications. *Computer Networks*, 53(4), pp. 501–514.

Errata

I. No errata

II. No errata

III. Table 19.1: Radio network routing → Radio network management

IV. No errata

V. No errata

VI. No errata

VII. No errata

Aalto-DD 100/2016

The Internet traffic volume, especially of heavy content and in mobile networks, is increasing heavily and networks are facing scalability challenges. In-network caching, specifically information-centric networking and software-defined networking, is utilized to efficiently solve the technical traffic optimization problem. This thesis complements the technical development of in-network caching by analyzing its economic feasibility from the perspectives of the different market actors. The research makes several contributions by analyzing different industry scenarios for the future Internet content delivery market and evaluating the cost efficiency of in-network caching compared with current caching solutions.

ISBN 978-952-60-6826-8 (printed)
ISBN 978-952-60-6827-5 (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