Towards SDN/NFV-based Mobile Packet Core
Benefits, Challenges, and Potential Solutions

Van-Giang Nguyen
Towards SDN/NFV-based Mobile Packet Core

Benefits, Challenges, and Potential Solutions

Van-Giang Nguyen
Towards SDN/NFV-based Mobile Packet Core: Benefits, Challenges, and Potential Solutions

Van-Giang Nguyen

LICENTIATE THESIS

Karlstad University Studies  |  2018:25

urn:nbn:se:kau:diva-67132

ISSN 1403-8099

ISBN 978-91-7063-858-9 (print)


© The author

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Mathematics and Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2018

WWW.KAU.SE
Towards SDN/NFV-based Mobile Packet Core: Benefits, Challenges, and Potential Solutions

Van Giang Nguyen
Department of Mathematics and Computer Science

Abstract

In mobile networks, the mobile core plays a crucial role in providing connectivity between mobile user devices and external packet data networks such as the Internet. Through the years, along with the dramatical changes in radio access networks, the mobile core has also been evolved from being a circuit-based analog telephony system in its first generation (1G) to become a purely packet-based network called the Evolved Packet Core (EPC) in the current generation (4G). In recent years, the explosion of mobile data traffic and devices and the advent of new services have led to the investigation of the next generation of mobile networks, i.e., 5G. A wide range of technologies has been proposed as candidates for the development of 5G. Among other technology candidates, Software Defined Networking (SDN) and Network Function Virtualization (NFV) have been widely considered to be key enablers for the network architecture of 5G, especially the mobile packet core (MPC) network.

This thesis aims at identifying challenges and problems of introducing SDN and NFV to re-architect the current MPC network architecture towards 5G and addressing some of the challenges. To this end, we conduct a comprehensive literature review of the state-of-the-art work leveraging SDN and NFV to re-design the 4G EPC architecture. Through this survey work, several research questions for future work have been identified and we contribute to address two of them in this thesis. Firstly, since most of the current works focus on unicast services, we propose an SDN/NFV-based MPC architecture for providing multicast and broadcast services. Our numerical results show that the proposed architecture can reduce the total signaling cost compared to the traditional architecture. Secondly, we address the question regarding the scalability of the control plane. We take the Mobility Management Entity (MME) – one of the key control plane entities of EPC – as a case study. In our work, the MME is deployed as a cluster of multiple virtual instances (vMMEs) and a front-end load balancer. We focus on investigating different approaches to achieve better load balancing among these vMMEs, which in turn improves scalability. Our experimental results suggest that carefully selected load balancing algorithms can significantly reduce the control plane latency.

Keywords: mobile packet core, evolved packet core, SDN, NFV, 5G, scalability, load balancing, control plane, open5Gcore.
Acknowledgements

The work presented in this thesis encapsulates the half-way of my 5-year PhD research I have undertaken at Karlstad University. It would not have been accomplished without the assistance of many people. I would like to take this opportunity to extend my sincere gratitude and appreciation to all those who made this thesis possible.

First and foremost, I would like to express my sincere gratitude and special thanks to my main supervisor, Professor Anna Brunström for giving me the opportunity to undertake a PhD here in Karlstad and for keeping me motivated throughout my research. Her guidance, constructive comment, suggestions, continuous support and encouragement have not been invaluable to the birth of this thesis, but also to my general development as a researcher.

I am also pleased to have Associate Professor Karl-Johan Grinnemo and Associate Professor Javid Taheri as my co-supervisors. They have given me insightful feedback, encouragement and support during my research. Special thanks to Karl-Johan Grinnemo for his tireless effort in reviewing, and commenting on my papers in the very early stage of my writing.

I would like to thank all colleagues and friends from the department of computer science at Karlstad university, in particular the Distributed and Systems and Communication Research (DISCO) group for maintaining a motivating and friendly environment for research.

I would like to thank Dr. Do Truong Xuan from Soongsil University, South Korea for his collaboration and help. A special thank goes to my close friends Nguyen Tien Cuong, Vu Dinh Quyen, and Vu Huu Tiep for their support and encouragement.

Finally, I would like to dedicate this thesis to my family and my wife, Ton Thi Kim Loan for their unconditional love, patience, support and encouragement during the ups and downs that I went through.
Contents

List of Appended Papers xii

INTRODUCTORY SUMMARY 1

1 Introduction 3

2 Background 4
   2.1 Mobile Packet Core Architecture 4
   2.2 Software Defined Networking 6
   2.3 Network Function Virtualization 7

3 Research Questions 8

4 Contributions 9

5 Research Methods 10

6 Summary of Appended Papers 11

7 Conclusions and Future Work 14

PAPER I: 5G Mobile Networks: Requirements, Enabling Technologies, and Research Activities 19

1 Introduction 19
   1.1 What is 5G? 20
      1.1.1 From a System Architecture Perspective 20
      1.1.2 From a Spectrum Perspective 20
      1.1.3 From a User and Customer Perspective 21
   1.2 Typical Use Cases 21

2 5G Requirements 22
   2.1 High Data Rate and Ultra Low Latency 23
   2.2 Massive Connectivity and Seamless Mobility 23
   2.3 Reliability and High Availability 24
   2.4 Flexibility and Programmability 24
   2.5 Energy, Cost and Spectrum Efficiency 25
   2.6 Security and Privacy 25

3 5G Enabling Technologies 26
   3.1 5G Radio Access Network 27
      3.1.1 mmWave Communication 27
      3.1.2 Massive MIMO 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3</td>
<td>Ultra-Dense Small Cells</td>
<td>28</td>
</tr>
<tr>
<td>3.1.4</td>
<td>M2M and D2D Communications</td>
<td>29</td>
</tr>
<tr>
<td>3.1.5</td>
<td>Cloud-based Radio Access Network</td>
<td>30</td>
</tr>
<tr>
<td>3.1.6</td>
<td>Mobile Edge and Fog Computing</td>
<td>31</td>
</tr>
<tr>
<td>3.2</td>
<td>5G Mobile Core Network</td>
<td>32</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Software Defined Networking</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Network Function Virtualization</td>
<td>33</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Cloud Computing</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>5G End-to-End System</td>
<td>34</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Network Slicing</td>
<td>34</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Management and Orchestration</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>5G Standardization Activities</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>ITU Activities</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>3GPP Activities</td>
<td>38</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Pre-5G Phase</td>
<td>38</td>
</tr>
<tr>
<td>4.2.2</td>
<td>5G Phase I</td>
<td>38</td>
</tr>
<tr>
<td>4.2.3</td>
<td>5G Phase II</td>
<td>39</td>
</tr>
<tr>
<td>4.3</td>
<td>ETSI Activities</td>
<td>39</td>
</tr>
<tr>
<td>4.4</td>
<td>IEEE Activities</td>
<td>40</td>
</tr>
<tr>
<td>4.5</td>
<td>IETF Activities</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>5G Research Communities</td>
<td>41</td>
</tr>
<tr>
<td>5.1</td>
<td>European 5G Related Activities</td>
<td>41</td>
</tr>
<tr>
<td>5.2</td>
<td>Asian 5G Related Activities</td>
<td>42</td>
</tr>
<tr>
<td>5.2.1</td>
<td>South Korea: 5G Forum</td>
<td>43</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Japan: 5GMF Forum</td>
<td>43</td>
</tr>
<tr>
<td>5.2.3</td>
<td>China: IMT-2020 5G Promotion Group</td>
<td>43</td>
</tr>
<tr>
<td>5.3</td>
<td>American 5G Related Activities</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion</td>
<td>44</td>
</tr>
<tr>
<td>7</td>
<td>Acknowledgment</td>
<td>44</td>
</tr>
</tbody>
</table>

**Paper II:**

SDN/NFV-based Mobile Packet Core Network Architectures: A Survey

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>52</td>
</tr>
<tr>
<td>1.1</td>
<td>Scope and Contributions</td>
<td>55</td>
</tr>
<tr>
<td>1.2</td>
<td>Organization</td>
<td>57</td>
</tr>
</tbody>
</table>
2 Background
2.1 Evolved Packet Core Architecture ............... 58
2.2 Software Defined Networking .................. 59
2.3 Network Function Virtualization ................. 61
2.4 SDN versus NFV .................................. 62

3 SDN/NFV-Based MPC Architectures ............. 63
3.1 NFV-based EPC Architecture .................... 63
3.2 SDN/NFV-based EPC Architecture ............... 67
3.3 Full-SDN MPC Architecture ..................... 68

4 Taxonomy Description ............................. 69
4.1 Architectural Approach: Revolutionary vs Evolutionary 71
4.2 Technology Adoption: SDN vs NFV .............. 71
4.3 Functional Implementation: Splitting vs Merging ...... 72
4.4 Deployment Strategy: Centralized vs Distributed ...... 73
4.5 Classification of Contemporary Work ............ 74

5 Architectural Approaches ....................... 74
5.1 Revolutionary Approaches ...................... 75
5.2 Evolutionary Approaches ...................... 75

6 Technology Adoption ......................... 75
6.1 SDN Only ........................................ 77
   6.1.1 Full SDN adoption ......................... 77
   6.1.2 Partial SDN adoption ...................... 79
   6.1.3 Summary .................................... 80
6.2 NFV Only ........................................ 81
   6.2.1 Full NFV adoption ......................... 81
   6.2.2 Hybrid NFV adoption ...................... 83
   6.2.3 Summary .................................... 84
6.3 SDN and NFV .................................... 84
   6.3.1 Full Virtualization ......................... 84
   6.3.2 Partial Virtualization ...................... 87
   6.3.3 Partial and Full Virtualization .......... 89
   6.3.4 Summary .................................... 90

7 Functional Implementation ................... 90
7.1 1:1 Migration ................................... 92
7.2 Splitting ....................................... 92
7.3 Merging ....................................... 93
7.4 Summary ....................................... 95
8 Deployment Strategy
8.1 Control Plane ........................................ 97
8.1.1 Centralized Control Plane ......................... 97
8.1.2 Distributed Control Plane ......................... 97
8.2 Data/User Plane ....................................... 99
8.2.1 Centralized Data Plane ......................... 99
8.2.2 Distributed Data Plane ......................... 100
8.3 Summary ........................................ 100

9 Research Challenges and Future Directions 101
9.1 Southbound and Northbound Interfaces ............ 102
9.2 Scalable, Reliable, and High Available Design .... 103
9.2.1 Scalability ........................................ 103
9.2.2 Reliability and High Availability ............... 106
9.3 Placement Optimization and Resource Allocation .... 106
9.4 Management and Orchestration ....................... 109
9.5 Network Sharing and Slicing ....................... 111
9.6 Network Performance, Evaluation and Benchmarks ... 112

10 Conclusion 114

11 Acknowledgment 116

Paper III:
SDN-based Mobile Packet Core for Multicast and Broadcast Services 139

1 Introduction 139

2 Related Works 141
2.1 Current Mobile Network Architecture ............... 141
2.2 Proposals for SDN-based Mobile Packet Core .... 141
2.3 Multicast and Broadcast Support Over Current Mobile Networks ......................... 143

3 SDN-based Mobile Packet Core for MBMS Services 144
3.1 Architecture for Multicast and Broadcast Service Over SDN-based 5G Mobile Packet Core ......................... 144
3.2 Procedures ........................................ 146
3.2.1 MBMS Multicast Service Activation ........... 146
3.2.2 Registration Procedure ......................... 147
3.2.3 MBMS Session Start ......................... 148
3.2.4 MBMS Session Stop ......................... 149
3.2.5 MBMS Session Update ......................... 149
4 Performance Analysis 150
  4.1 Signaling Cost for 3GPP MBMS Architecture . . . . . . . . . 151
    4.1.1 Multicast Service . . . . . . . . . . . . . . . . . . . 151
    4.1.2 Broadcast Service . . . . . . . . . . . . . . . . . . . 152
  4.2 Signaling Cost for SDN-based MBMS Architecture . . . . . 152
    4.2.1 Multicast Service . . . . . . . . . . . . . . . . . . . 152
    4.2.2 Broadcast Service . . . . . . . . . . . . . . . . . . . 152

5 Evaluation Results and Discussion 153
  5.1 Numerical Results on Signaling Cost . . . . . . . . . . . . . 153
  5.2 Discussion on CAPEX and OPEX Cost . . . . . . . . . . . . . 154

6 Conclusion 157

PAPER IV:
On Load Balancing for a Virtual and Distributed MME in
the 5G Core 163

1 Introduction 163

2 Related Works 166

3 MME Load balancing 168
  3.1 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . 168
  3.2 Traditional approach . . . . . . . . . . . . . . . . . . . . 169
  3.3 vMME approaches . . . . . . . . . . . . . . . . . . . . . . 169
    3.3.1 RD Algorithm . . . . . . . . . . . . . . . . . . . . . 169
    3.3.2 RR Algorithm . . . . . . . . . . . . . . . . . . . . . 170
    3.3.3 WRR Algorithm . . . . . . . . . . . . . . . . . . . . 170

4 Experimental Evaluation 170
  4.1 Experiment Setup . . . . . . . . . . . . . . . . . . . . . . . 170
    4.1.1 Scenario 1 . . . . . . . . . . . . . . . . . . . . . . . 171
    4.1.2 Scenario 2 . . . . . . . . . . . . . . . . . . . . . . . 172
  4.2 Experimental Results . . . . . . . . . . . . . . . . . . . . 172
    4.2.1 Scenario 1 . . . . . . . . . . . . . . . . . . . . . . . 172
    4.2.2 Scenario 2 . . . . . . . . . . . . . . . . . . . . . . . 173

5 Conclusion and Future Works 175

6 Acknowledgment 177
List of Appended Papers


Comments on my Participation

Paper I I am the main author and was responsible for writing the entire book chapter. My co-authors gave me useful comments and valuable feedbacks during the writing and revising process of the book chapter.

Paper II I came up with the initial idea of the paper and am responsible for the entire writing process. My co-authors helped me develop the idea, especially the structure of the paper, the formulation of the survey taxonomy, and the construction of the comparison tables. My co-authors also gave me a number of constructive comments and feedbacks during the writing and two-round revising process.

Paper III This work was carried out in collaboration with Truong-Xuan Do, who came up with the initial idea. I actively contributed to the design of the entire proposed architecture, its protocol operations, and evaluation methods. I participated in the entire writing process, and was the main author of Section 2 and Section 5.2. I also participated in revising and responding to the reviewers’ comments during two rounds of revisions. Young Han Kim contributed to proof-reading process, and provided feedbacks.

Paper IV I am the main author of the paper and wrote the entire paper. The idea for this paper came from the discussion between the co-authors and myself. My co-authors gave me constructive inputs, comments throughout the whole process of developing the idea as well as conducting the experiments.
Other Publications


1 Introduction

Over the years, we have witnessed an explosive growth in the volume of mobile data traffic due to the ever-increasing number of new mobile devices, the growth of the Internet of Things market, and the appearance of many new disruptive services (e.g., virtual reality and autonomous driving). According to a forecast by Cisco [1], the total volume of mobile data traffic is expected to reach approximately 31 Exabytes per month by 2020, with a compound annual growth rate of about 46 percent. To cope with this rapid growth of user demand, the current mobile system (4G) needs to be upgraded or re-designed. This has triggered the investigation of the fifth generation of mobile networks, i.e., 5G.

Recently, researchers from industry and academia have been proposing many different potential solutions to contribute to the development of 5G. Although various definitions and visions of what 5G would be have been discussed, they have the same target to roll out 5G technology worldwide by 2020 [2, 3]. By that time, 5G is expected to support a diversity of use cases with greater capabilities than today’s 4G network, ranging from ultra-low latency (less than 1ms) to ultra-high data rate (up to 10 Gbps) [4]. In order to achieve its targets, 5G is recognized as a big paradigm shift rather than a minor evolution from today’s 4G network. It is expected to bring a radical change in both the radio access as well as mobile core parts of the mobile system by using advanced technologies. Among other technology candidates, Software Defined Networking (SDN) [5] and Network Function Virtualization (NFV) [6] have emerged as two promising technologies for the development of the 5G network architecture, especially the Mobile Packet Core (MPC) network. While SDN features the decoupling between the control and the data planes from the conventional networking stack, NFV enables the transformation of network functions from dedicated hardware implementation to run as software appliances in a general-purpose server or in a cloud environment. In that way, SDN and NFV offer a great cost reduction, flexibility in management, shortened time to launch new services, etc. to operators.

In this thesis, we are interested in the MPC network which provides the connectivity for mobile users to use services provided by external IP networks such as the Internet or the IP multimedia subsystem (IMS). Our main goal is to address some challenges raised from the adoption of SDN and NFV in the re-design of the MPC architecture towards 5G. To this end, we first conduct a review on 5G mobile networks including typical use cases, key requirements, and enabling technologies in order to understand the driving factors and the role of SDN and NFV in the development of 5G. Since the 5G network architecture is yet to be standardized and due to the flexibility, scalability, and cost reductions offered by SDN and NFV, a large part of the current research on the 5G network architecture is to re-design the 4G EPC architecture using the principles of SDN and NFV. As the second contribution, we provide a comprehensive survey on how the SDN and NFV technologies can change the 4G EPC architecture. We cover not only the work done in academia but also...
industry efforts as well as standardizations initiatives. Moreover, we identify many open questions and gaps for future research. Based on this, we propose solutions to address two of the research questions. Firstly, since most of the current works focus on unicast services, we propose an SDN/NFV-based MPC architecture for providing multicast and broadcast services. Our numerical results show that the proposed architecture can reduce the total signaling cost compared to the traditional architecture. Secondly, we address the question regarding the scalability of the control plane. We take the Mobility Management Entity (MME) – one of the key control plane entities of EPC – as a case study. In our work, the MME is deployed as a cluster of multiple virtual instances (vMMEs) and a front-end load balancer. We focus on investigating different approaches to achieve better load balancing among these vMMEs, which in turn improves scalability. Our experimental results suggest that carefully selected load balancing algorithms can significantly reduce the control plane latency.

The rest of the introductory summary is structured as follows. Section 2 provides the required background knowledge for the work presented in this thesis. Section 3 outlines the main objectives of this thesis and the research questions. The main contributions of the thesis are summarized in Section 4. Section 5 describes the employed research methodology. Section 6 presents a short summary of the appended papers. Finally, Section 7 concludes with some final remarks and a discussion about future work.

2 Background

In this section, we provide the necessary background to understand the context and research work conducted in this thesis. The research context of this thesis falls into the development of the MPC architecture, 5G, and two related enabling technologies: SDN and NFV. As Paper II provides a comprehensive background on the current MPC and the existing SDN/NFV-based MPC solutions, in the following we limit the description to the current MPC architecture and its main entities, and summarize the main concepts of SDN and NFV. A broader view on 5G and related activities are provided in Paper I.

2.1 Mobile Packet Core Architecture

The most recent mobile packet core network is the EPC, which is the core of the Long Term Evolution (LTE) 4G system [7, 8]. Together, the EPC and the LTE forms the Evolved Packet System (EPS), which is a flat, all-IP network with the separation of the control and data planes. The LTE/EPC typically aims at delivering services and applications to users using the unicast mode wherein each user gets its own radio channels for transmitting and receiving data. This unicast mode is often used for applications such as mail, web-browsing, and media downloads, etc. However, there was a huge demand for using multimedia services such as video live streaming over the mobile networks. To meet this demand, the 3GPP defined the Multimedia Multicast and
Broadcast Services (MBMS) in the Release 6 [9], which allows 3G mobile operators to efficiently deliver media content to a large group of users at the same time. In the 3GPP Release 9 [10], the MBMS became the evolved MBMS (eMBMS) to support multicast and broadcast services over the 4G LTE networks. Figure 1 depicts the EPS system which supports unicast services provided via the EPC, and multicast and broadcast services provided via the eMBMS. The EPC architecture is composed of four main network elements: the Mobility Management Entity (MME) which is responsible for controlling all signaling events coming from user devices via the S1-MME interface, and performs bearer setup for the data plane traffic via the S11 interface; the Serving Gateway (SGW) which is an intermediate gateway between the radio access network and the EPC, and is responsible for transferring all packets across the data plane from eNodeBs to the Packet Data Network Gateway (PGW) via the S1-U and the S5 or S8 interfaces; the PGW which provides the connectivity from the user devices to external IP networks, e.g., the Internet, via the SGi interface; and the Home Subscriber Server (HSS) which is a server which stores user-related and subscriber-related information, and exchanges this information with the MME via the S6a interface.

The eMBMS architecture has three new types of nodes working together with the MME. The Multicell/Multicast Coordinator Entity (MCE) is responsible for coordinating multiple cells, allocating resources, and performing admission control for an MBMS service. The MCE can either be co-located with the eNodeB or work as an independent physical node. It communicates with the MME via a newly defined M3 interface. The MBMS Gateway (MBMS-GW) acts as an intermediate gateway to forward user-plane traffic to the eNodeBs by IP multicast mechanisms through the M1 interface. The MBMS-GW interacts with the MME via the Sm interface. The Broadcast Mul-
ticast Service Center (BMSC) is responsible for authentication/authorization of both content providers and users. It is also responsible for charging and configuration of multimedia data flows through the core network via the SG-mb interface in the control plane and the SGI-mb interface in the data plane (see Figure 1). As shown in Figure 1, an MBMS service area is referred as a group of eNodeBs or a region within which the same multimedia content is distributed.

2.2 Software Defined Networking
SDN [5] is a new networking paradigm featuring the decoupling of the control and data planes. Figure 2 provides the overview of the SDN architecture. The control plane functions are abstracted and moved to a logically centralized control entity, called an SDN controller. The data forwarding plane is composed of Data Forwarding Elements (DFEs) such as virtual/physical switches and routers which are programmed by the SDN controller via an open Application Programming Interfaces (API), e.g., OpenFlow. OpenFlow [12] is the most common protocol used between the SDN controller and the data plane. Network functions such as routing, monitoring, load balancing, etc., which are previously coupled with the data forwarding plane, have now become applications on top of the SDN controller. The communication between these applications and the SDN controller is often done via a Representational State Transfer (REST) API.

Nowadays, SDN has gained a lot of attention from both academia and industry in many networking areas, not only wired networks such as campus or data center [13, 14] but it has also been expanding quickly in the field of mobile and wireless networks [15]. Regarding the application of SDN in the MPC, SDN is essentially applied to decouple the control and data planes of the EPC’s gateways, i.e., the SGW and the PGW. In that way, it brings a great flexibility in the control and management of the EPC.
2.3 Network Function Virtualization

NFV [6] is essentially a networking initiative to transform individual network functions that previously were implemented as dedicated hardware to become software appliances running on general-purpose commodity servers. Figure 3 shows the reference architectural framework for NFV. In this figure, Virtual Network Functions (VNFs) represent the implementation of network functions. The NFV Infrastructure (NFVI) provides resources that can be utilized by VNFs. The NFV management and orchestrator (NFV MANO) [16] is responsible for orchestrating and managing VNFs through a set of VNF Managers (VNFM) and NFVI through a Virtualized Infrastructure Manager (VIM). The NFV orchestrator (NFVO) is in charge of network services (NS) life-cycle management, on-boarding of new NS, etc. In addition, the NFV MANO also allows the integration with external Operational and Business Support Systems (OSS/BSS) as shown in Figure 3.

Among a variety of NFV use cases covered in [17], virtualizing the EPC’s entities is one of the most important use cases and has been attracting a lot of attention. In that way, NFV promises significant cost saving either in deployment or in operation for mobile operators.
3 Research Questions

The overall objective of this thesis is to identify the challenges of applying SDN and NFV to re-design the current MPC architecture, and to address some of these challenges. In order to achieve the objective above, we are going to address the following research questions:

1. **RQ1: What are the main challenges while re-designing the 4G EPC architecture towards 5G using SDN and NFV principles?**
   
   As 5G-related research activities are actively being conducted around the world, understanding the concept of 5G, the driving factors behind it, and the enabling technologies is a crucial step for us before conducting the research and contributing to this field. Furthermore, we want to know the roles of SDN and NFV in the development of 5G, in particular, the MPC network. Because 5G is still in its infancy, we realized that the most typical approach adopted by most researchers is to take the current MPC architecture, the EPC, as the starting point and re-design it by applying SDN and NFV principles. A large number of such works has been proposed over the last few years, and we aim at identifying challenges and open issues raised from this adoption.

2. **RQ2: In what way could an SDN/NFV-based mobile packet core network be beneficial for multicast and broadcast services?**
   
   Multicast and broadcast services have been provided by several operators over their mobile networks to efficiently deliver multimedia content to multiple users while consuming minimum radio and network resources. While addressing **RQ1**, we saw that most of the existing works focus on the delivery of unicast services. This motivated us to analyze if such a new SDN/NFV-based design approach could also be beneficial in providing multicast and broadcast services.

3. **RQ3: How could the scalability of the SDN/NFV-based mobile packet core be improved?**
   
   In most SDN/NFV-based MPC network architectures, the MME – one of the EPC’s key control plane entities – still remains functionally unchanged. However, it now becomes a virtual instance running in a cloud data center instead of on dedicated hardware. It has been responsible for handling up to hundreds of thousands of control messages per second in 4G during busy hours. As the number of devices increases, the number of signaling events to be handled by the MME will drastically increase, thus resulting in a scalability problem. This motivates us to investigate solutions to solve this problem for the MME.
4 Contributions

While addressing the research questions outlined in Section 3, this thesis makes the following contributions:

1. **A review of essential aspects of 5G mobile networks**
   The first contribution of this work is provided in Paper I. We conducted a review on typical use cases, key requirements, and enabling technologies which are essential for the development of 5G. This paper also provides an overview of research activities that are being conducted around the world including main standardizations bodies and major research communities. This contributes to addressing research question RQ1.

2. **A comprehensive survey of SDN/NFV-based MPC network architectures**
   The second contribution identifying open issues and key future research challenges is presented in Paper II. In this paper, we conducted a comprehensive survey of existing solutions which adopt the principles of SDN and NFV to re-construct the 4G EPC architecture. More than 60 research papers have been reviewed and categorized by using a proposed multi-dimensional taxonomy framework. Based on that, we identified several open research questions and challenges that need to be further addressed in the future. This contributes to answering research question RQ1.

3. **A proposed SDN/NFV-based MPC network architecture supporting multicast and broadcast services**
   A research gap observed from our second contribution is that most of the existing SDN/NFV-based solutions focus on the delivery of unicast services. In Paper III, we addressed this issue, by proposing a new SDN/NFV-based MPC architecture for supporting multicast and broadcast services. We answer research question RQ2 by showing that the proposed architecture not only gains advantages of SDN and NFV in terms of cost reduction and flexibility, but also helps reduce the signaling load compared to the traditional approach.

4. **An approach to improve the scalability of the key control plane entity of an SDN/NFV-based MPC network**
   As the final contribution, we tackle the scalability of the MME entity which still remains as the key control plane entity of an SDN/NFV-based design of the 5G core. Applying the NFV concept, the MME becomes a virtual instance (vMME) running in a cloud environment. The scalability can be improved by running multiple vMMEs as a cluster. One remaining question is how to efficiently balance the signaling load among these vMMEs within the cluster. In Paper IV, we show that carefully selected load balancing algorithms can significantly improve the performance compared to simple random or round-robin schemes, thus further improving the scalability of the vMME architecture, something which contributes to answering research question RQ3.
5 Research Methods

Research methods are referred to as approaches, procedures, and guidelines used in conducting research [18]. There are a variety of different methods and they can be classified in many different ways. For example, they can be categorized as either being qualitative or quantitative methods [19]. While qualitative methods involve case studies and surveys, and provide a detailed non-numerical description, quantitative methods focus on statistical hypothesis testing associated with measurements. In the computer networking field, research methods can be classified as theoretical methods which are mainly based on mathematics and logic, and experimental methods which concern practical aspects of research [18]. The research work presented in this thesis falls in all of these categories as further discussed below. We also followed the traditional scientific approach [20] in conducting our research. It refers to an iterative cycle of observations, hypotheses formulation, testing and verification.

To tackle RQ1, we employed a systematic literature review [21] as a qualitative research method. Understanding the current knowledge and existing body of work in the topic of interest is a crucial step to identify research gaps and open research questions. In this case, a systematic literature review serves as the best research tool. The main steps of a systematic literature review include formulating the research question, setting inclusion or exclusion criteria, selecting and accessing the literature, assessing the quality of the literature, analyzing, synthesizing and disseminating the findings [22]. By following these steps, we aim at understanding the role of SDN and NFV in the evolution and development of the MPC network towards 5G, and more importantly identifying the challenges and research questions raised from the adoption of SDN and NFV into the MPC. The outcomes of this research are presented in Paper I and Paper II.

To address research question RQ2, we adopted an analytical model to verify our hypothesis. Analytical modeling is one of the four main methods for hypothesis verification; the other ones are real measurements, simulations, and emulations. Since this research topic is novel and lack available testing equipment and environments, analytical modeling seems to be the most appropriate method. The numerical data used in our modeling is taken from standard specifications. The numerical results are presented in Paper III.

The problem addressed in research question RQ3 was formulated based on the outcomes of my literature review. The hypothesis for this question is that the scalability of the control plane can be improved by deploying multiple control plane entities of the same type (e.g., MME) in a virtualized environment, and efficiently balancing the load among them. An experimental method to verify this hypothesis is adopted by using a standards-compliant emulation platform named Open5GCore. This is mainly due to the availability of the software and advantages of emulation over simulation. The experimental results are presented in Paper IV.
6 Summary of Appended Papers

Paper I – 5G Mobile Networks: Requirements, Enabling Technologies, and Research Activities

In this paper, we present an overview of 5G mobile networks and provide essential information in relation to the development of 5G. We start by summarizing the three main categories of use cases proposed by the ITU and 3GPP namely, enhanced mobile broadband, massive machine type communications, and ultra-reliable and low latency communications. Depending on different use case categories, a large number of different requirements has been defined. We briefly describe the key requirements from two different perspectives: a user performance perspective such as high data rate and ultra-low latency, and a system and management perspective such as flexibility and programmability. In order to fulfill these stringent requirements, advanced technologies should not only be adapted to the radio access side but also the core network. In the second part of the paper, we review 5G-enabled technologies and their roles in the 5G system. Examples of these enabling technologies include mmWave, Cloud-RAN, and Mobile Edge Computing for the radio access network; SDN, NFV, and cloud computing for the core network; and network slicing and management and orchestration for the end-to-end 5G system. In addition, we also identify several research challenges and questions that need to be resolved in future research for each described technology. Finally, we provide a summary of the status of ongoing research activities related to 5G around the world including standardization organizations and research communities. This paper serves as a basis for the rest of the thesis work.

Paper II – SDN/NFV-based Mobile Packet Core Network Architectures: A Survey

In this paper, we provide a comprehensive survey of proposed solutions to incorporate SDN and NFV into the current 4G MPC architecture. We first describe the most typical ways to re-design the 4G EPC architecture using the principles of SDN and NFV and discuss their advantages and disadvantages. Over the past few years, there has been a large number of SDN/NFV-based MPC architectures. Therefore, we propose a taxonomy which helps us classify and analyze the existing SDN/NFV-based MPC proposals in four main dimensions including architectural approach, technology adoption, functional implementation, and deployment strategy. This proposed taxonomy identifies the trade-offs that need to be taken between selected technologies, architectural and functional designs, and deployment options. On the basis of the identified trade-offs, the taxonomy suggests the most suitable implementation and deployment options. In addition, we also defined some additional criteria and attributes to make comparison between the existing SDN/NFV-based MPC solutions. Another important contribution of this paper is a list of challenges and research questions that need to be addressed in future research. We categorize them into six main research directions including the develop-
ment of standard interfaces, the improvement of scalability and reliability, the placement optimization and allocation of resources, management and orchestration, network slicing and sharing, and network performance evaluation and benchmarking. Among the identified challenges in this paper, we propose solutions to address two of them: the lack of solutions for supporting multicast and broadcast services addressed in Paper III and the scalability of the control plane addressed in Paper IV, respectively.

Paper III – SDN-based Mobile Packet Core for Multicast and Broadcast Services

In this paper, we focus on the adoption of SDN and NFV in the EPS system for multicast and broadcast services, i.e., the eMBMS. The eMBMS has been deployed by several operators to efficiently deliver multimedia services to multiple users simultaneously, thus helping them to reduce costs. However, similar to the EPC architecture, the eMBMS has still several limitations including hardware-based functional deployment, and a tight coupling between the control and data planes. Through our survey in Paper II, a large number of solutions has been proposed to solve these limitations in the EPC by using SDN and NFV, but we have not seen any solution for the eMBMS. To this end, we propose a new architecture where we re-designed the current eMBMS architecture by using the principles of SDN and NFV. We introduce several modules to enable multicast and broadcast services such as a Multicast Broadcast Handler (MB Handler) and a Multicast Broad Routing (MBR). These modules are implemented as applications on top of a mobile controller. The protocol operations for enabling multicast and broadcast services are also elaborated within our proposed architecture. By numerical evaluation, we show that our proposed architecture can reduce the total signaling cost as compared to the 3GPP eMBMS architecture. We also discuss the capital expenditure (CAPEX) and operational expenditure (OPEX) cost reduced by our proposed architecture. Still, there are some aspects that falls outside of the scope of the paper and which we recognize need to be addressed by future work. Firstly, we propose to let the mobile controller and other control plane functions run in a cloud environment to solve the scalability issue. However, there is a lack of detailed solutions on how the scalability can be completely addressed. Secondly, our work is limited to the protocol operations without any practical implementation to see how it works. This is due to the lack of available testing equipment and environments. It is also required us to carry out a more extensive evaluation to prove the benefits of the proposed architecture.

Paper IV – On Load Balancing for a Virtual and Distributed MME in the 5G Core

In this paper, we present a solution to improve the scalability of the MME, which still remains a key control plane entity in most SDN/NFV-based MPC architectures. We do this by improving the load balancing in a virtual and distributed MME (vMME) architecture. In particular, we exploit the effect on
Towards SDN/NFV-based Mobile Packet Core: Benefits, Challenges, ...

the control plane latency and resource usage of load balancing algorithms used in this kind of MME design. In most of the existing designs, simple algorithms such as Random (RD) and Round-Robin (RR) are employed by the front-end Load Balancer (LB) to balance the incoming traffic, which we found not efficient enough. In addition, there has not been any performance evaluation and comparison of these algorithms. To this end, we first implement a Weighted Round Robin (WRR) algorithm which takes into account the heterogeneity of resources such as link latency between the LB and the vMME and the capacity of vMMEs. Then, we compare the performance of this algorithm with an RD and an RR algorithm in terms of the attachment delay and CPU usage. Experiments are conducted using a standards-compliant emulation platform, and with traffic that closely resembles that of a realistic traffic pattern. Experimental results show that our proposed WRR algorithm significantly reduce the attachment delay and has better resource usage. Still, there are some known limitations with our proposed WRR. Firstly, the weight assigned to each vMME in the current WRR implementation is set to a fixed value. Secondly, as we know that the processing time of the attachment request is always longer than that of the handover and detachment requests we can utilize this information to make a better load balancing decision. For example, the attachment request will be forwarded to the most powerful MME in the pool while the detachment request is sent to others. However, we have not considered this in the current implementation. Therefore, we are currently working on more advanced load balancing solutions which overcome these two limitations. Moreover, several scaling strategies such as proactive/reactive and hybrid scaling will be investigated in conjunction with the load balancing function in order to create a scalable and efficient vMME solution.
7 Conclusions and Future Work

Although the 4G EPC architecture, the current MPC architecture, has served us well for a decade or so, there are some inherent limitations with the 4G EPC architecture. Two limitations are the coupling of the control and data planes and the dependence on dedicated hardware. These limitations cause challenges for mobile operators to manage and upgrade their networks, and thus to meet the unprecedented requirements of the upcoming 5G. Therefore, a radical change in the architecture of the current MPC network, i.e., the EPC, is needed. Two emerging technologies, namely SDN and NFV, are widely considered as critical enablers for making this change possible.

In this thesis, we have presented our contributions to the knowledge in the evolution and development of MPC network architectures towards 5G. At first, we review the essential aspects of 5G mobile networks such as important use cases and key requirements, and related activities currently being conducted around the world. A detailed description of technology candidates considered to facilitate the development of 5G spreading from radio access networks to the MPC network is also provided. Secondly, we narrow down our scope to the MPC network and contribute to identify weaknesses and research gaps in the currently proposed SDN/NFV-based MPC architectures, and to suggest ways to address some of these weaknesses. Therefore, an extensive survey on the efforts made by standards, industry and academia in using SDN and NFV to re-design the 4G MPC architecture is conducted. Based on the outcomes of our survey, we attempt to address two research questions regarding the benefits of an SDN/NFV-based MPC solution to deliver multicast and broadcast services, and the scalability of the MME in an SDN/NFV-based MPC architecture.

In the future, we will continue our work on improving the scalability of the MME. As we shown in Paper IV, having a better load balancing algorithm, i.e., the WRR, can not only improve the scalability, but also significantly reduces the control plane latency compared to the simple algorithms, e.g., the RD and the RR. However, our current WRR scheme relies on a manual assignment of the weight to each MME in the pool. Therefore, we are currently developing more dynamic and adaptive load balancing algorithms which take into account the necessary information fed back from the vMME instances as well as the difference between the types of signaling events (e.g., attachment, detachment, and handover). Moreover, we also intend to employ some mechanisms to prevent the MME from being congested and to trigger scaling of the MME when needed in conjunction with the development of new load balancing algorithms. Finally, we would also like to contribute to another research challenge: the placement and allocation of resources for the MPC network in 5G as outlined in Paper II. For this, we aim at proposing placement or allocation solutions to not only guarantee stringent requirements of 5G (e.g., end-to-end latency) but also to provide high availability and scalability in the MPC network.
References


Towards SDN/NFV-based Mobile Packet Core
Benefits, Challenges, and Potential Solutions

In mobile networks, the mobile core plays a crucial role in providing connectivity between mobile user devices and external packet data networks such as the Internet. After more than three decades, the mobile core has been gradually evolved through four generations and is called the Evolved Packet Core (EPC) in the current generation (4G). In recent years, the explosion of mobile data traffic and devices and the advent of new services have led to the investigation of the next generation of mobile networks, i.e., 5G. Among other technology candidates, Software Defined Networking (SDN) and Network Function Virtualization (NFV) have been widely considered to be key enablers for the network architecture of 5G, especially the mobile packet core (MPC) network.

This thesis aims at identifying benefits and challenges of introducing SDN and NFV to re-architect the current MPC architecture towards 5G and addressing some of the challenges. To this end, we conduct a comprehensive survey of the existing SDN/NFV-based MPC architectures. Through this survey work, several research questions for future work have been identified and we contribute to address two of the research questions. Firstly, we propose an SDN/NFV-based MPC architecture for providing multicast and broadcast services. Secondly, we tackle the scalability problem of the Mobility Management Entity (MME) - one of the EPC key control plane entities. In particular, we investigate different approaches to achieve better load balancing among virtual MMEs in a virtual and distributed MME design, which in turn improves scalability.