An Adaptive Mechanism for LTE P-GW Virtualization using SDN and NFV

Luciano Jerez Chaves\textsuperscript{1,2}, Islene Calciolari Garcia\textsuperscript{2}, Edmundo R. Mauro Madeira\textsuperscript{2}

\textsuperscript{1}Federal University of Juiz de Fora (UFJF), Brazil
\textsuperscript{2}University of Campinas (Unicamp), Brazil
Outline

❖ Introduction
❖ Long-Term Evolution (LTE) networks
❖ Software-Defined Mobile Network (SDMN) architecture
❖ Packet GateWay (P-GW) user plane virtualization
❖ Performance evaluation
❖ Conclusions and future work
Introduction

❖ LTE networks have been de facto employed for high-speed wireless communication
❖ Mobile data traffic has grown 18-fold over the past 5 years
❖ Half a billion new mobile devices added in 2016
❖ Video traffic accounts for 60% of total data traffic

LTE mobile gateways

❖ Few proprietary elements designed to handle millions of users with high availability
❖ Complex and expensive equipments

Virtualize and “softwarize” existing mobile architecture

Software Defined Networking (SDN)
Network Function Virtualization (NFV)
Software Defined Networking

- Decouples the control plane from the data plane of network elements
- Network intelligence is centralized in software
- Simplified distributed forwarding hardware
Network Function Virtualization

- Decouples the physical equipment from the network functions that run on them
- Network functions are implemented in software
- Commercial Off-The-Shelf (COTS) physical servers
Contributions

❖ This work contributes with a virtualization model for the LTE Packet Data Network (PDN) GateWay (P-GW)
  ❖ SDN is used to split control and user planes
  ❖ NFV is used to model each segment independently
  ❖ User-plane virtualization modeled on top of OpenFlow switches
    ❖ Hardware or software-based infrastructure platforms
    ❖ Dynamic adaptation to workload changes
Long-Term Evolution networks

- 4G standard for high-speed wireless communication
- Maintained by the 3rd Generation Partnership Project
- **Evolved Packet System (EPS)**
  - Evolved Universal Terrestrial Radio Access Network (E-UTRAN)
  - Evolved Packet Core (EPC)
EPS architecture

Packet domain only
Standard interfaces for control/user planes
GPRS Tunneling Protocol (GTP)
EPS bearers

**EPS bearers for common QoS treatment**
- Guaranteed Bit Rate (GBR)
- Non-Guaranteed Bit Rate (Non-GBR)
- QoS Class Identifier (QCI)
- Traffic Flow Template (TFT)
Adopted SDMN architecture

- Single EPC controller
- OpenFlow backhaul network
- S/P-GW user plane virtualization on top of OpenFlow switches
Adopted SDMN architecture

- **Single EPC controller**
  - Modeled as a Virtual Network Function (VNF) on the cloud
  - Configures backhaul network and S/P-GW user plane switches
  - Complies with standard LTE control interfaces
  - Coordinates GTP Tunnel Endpoint ID (TEID) allocation
Adopted SDMN architecture

- **OpenFlow backhaul network**
  - Enhanced OpenFlow switches for GTP tunneled traffic
  - Only GTP TEID is used for matching and routing
Adopted SDMN architecture

- **S/P-GW user plane virtualization on top of OpenFlow switches**
- High amount of traffic load on the user plane
- OpenFlow switches used for fast packet processing
- Hardware physical switches or software-based switches
P-GW user plane virtualization

- Downlink IP packet filtering using Traffic Flow Templates (TFTs)
  - Match the 5-tuple to identify GTP TEID and S-GW address
  - One or more OpenFlow entries for each active bearer
    - Number of flow entries is prone to grow quickly
    - A single OpenFlow switch will not scale
P-GWu internal design

- Pool of TFT OpenFlow switches
  - Hardware or software switches
  - Tradeoff between flow table size \times packet processing capacity

<table>
<thead>
<tr>
<th>OpenFlow parameter</th>
<th>Hardware</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow table size</td>
<td>16K</td>
<td>750K</td>
</tr>
<tr>
<td>Processing capacity</td>
<td>240 Gbps</td>
<td>10 Gbps</td>
</tr>
</tbody>
</table>
P-GWu internal design

- **UpLink and DownLink (UL/DL) switches for TFT load balancing**
- OpenFlow logical ports used for GTP de/encapsulation
- All traffic of a single UE is processed by the same TFT switch
- UE IP address masked matching for load balancing
- For $2^n$ active TFT switches: match the $n$ LSB bits of IP address
Adaptive mechanism

- Dynamically adjust the number of TFT switches
  - Monitoring flow table usage and traffic processing load
  - Increase/decrease the TFT pool size: *split* / *join* thresholds
  - Using a factor of 2 when changing the number of switches
  - Moving TFT rules and updating UL/DL switches
Blocking strategy

- EPC controller may block bearer requests when no more TFT switches can be activated
  - Flow table is full: block
  - Traffic load exceeds switch processing capacity:
    - Block none requests, allowing switches to drop random packets
    - Block GBR requests, ensuring QoS requirements
    - Block all requests, regardless of the bearer type
Performance evaluation

- **Network Simulator 3 + OFSwitch13 module**
  - Backhaul of 6 OpenFlow switches in ring topology
  - LTE sites with 3 base stations laid out on a hexagonal grid
  - 5 different types of traffic flows with distinct QCIs
Blocking strategy evaluation

Bearer requests and traffic in steady state
- Block none strategy with high drop ratio
- Block GBR strategy expunges GBR traffic
- Block all strategy brings better results
Adaptive mechanism evaluation

Split threshold: 90%
Join threshold: 30% (conservative)
Traffic changes reflected on the number of active TFT switches
Steady-state P-GWu evaluation

Different number of active TFT switches for hardware and software platforms

High flow table usage on hardware

High traffic load on software
Steady-state P-GWu evaluation

Adaptive number of TFT switches can improve network throughput and mitigate bearer request block ratio.
Remarks for real-scale scenario

- **DL/UL switches are the bottleneck**
  - High performance hardware platform: ~ 400K UEs
    limited by processing capacity

- **TFT switches have distinct limitations**
  - Hardware platform: ~ 4K UEs
    limited by table size
  - Software platform: ~ 1.6K UEs
    limited by processing capacity

<table>
<thead>
<tr>
<th>OpenFlow parameter</th>
<th>Hardware</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow table size</td>
<td>16K</td>
<td>750K</td>
</tr>
<tr>
<td>Processing capacity</td>
<td>240 Gbps</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>TFT pool size</td>
<td>128</td>
<td>256</td>
</tr>
</tbody>
</table>
Conclusions and future work

- This paper shows...
  - A virtualization model for LTE gateways
  - P-GWu virtualization using OpenFlow switches
  - Adaptive mechanism for improved resource usage
  - Performance evaluation by means of simulations

- As future work...
  - Evaluate different TFT load balancing strategies
luciano@lrc.ic.unicamp.br