Combined Virtual Mobile Core Network Function Placement and Topology Optimization with Latency bounds

TECHNISCHE UNIVERSITÄT CHEMNITZ

Andreas Baumgartner
Varun Reddy
Thomas Bauschert

Chair of Communication Networks
Technische Universität Chemnitz

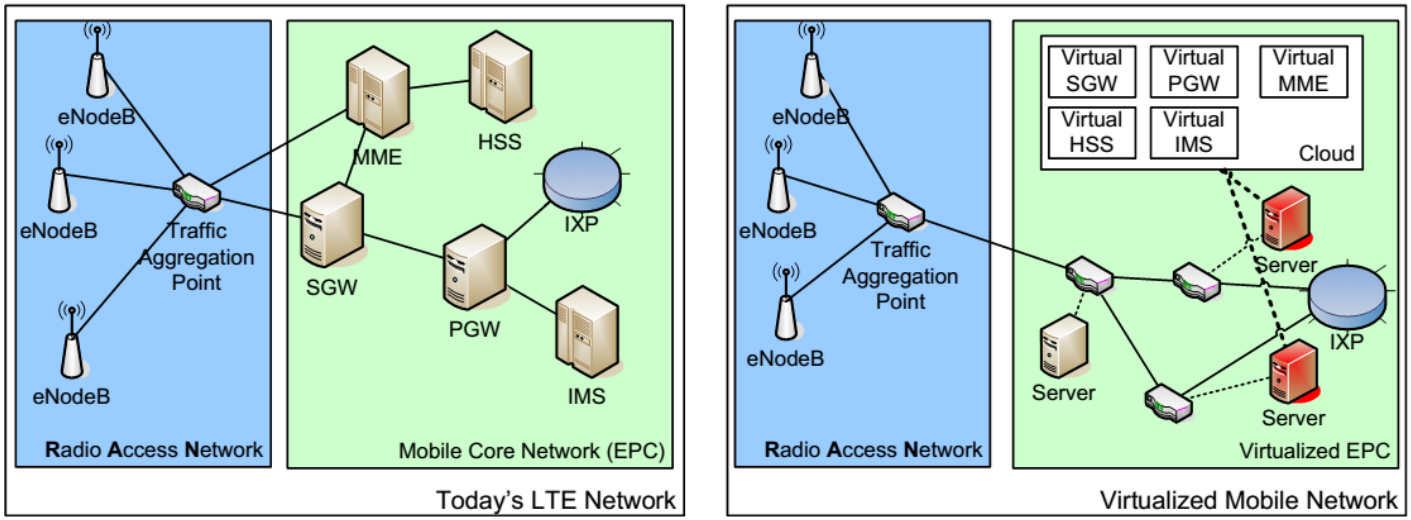
October 2, 2015

Overview

Overview

Introduction

Placement decisions for virtualized network functions in a fully virtualized mobile core network scenario:



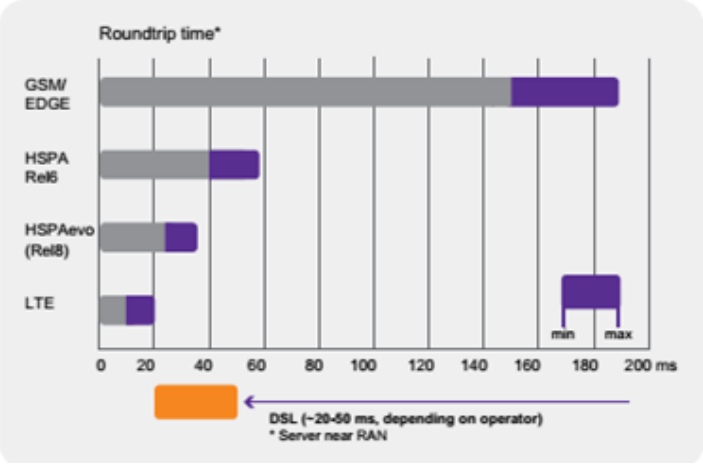
Introduction

Mathematical optimization model:

- Find optimum topology for virtual mobile core network topology
- Find optimum embedding for virtual mobile core network
- Assure roundtrip time (RRT) delay bounds for service chains

VINO Project (BMBF funded)

Virtual Network Optimization



Overview

Problem Formulation

Problem Statement

Find an optimum embedding with respect to embedding costs of the virtual mobile core network on a physical substrate network

Goal

Minimization of the overall link embedding costs and the costs for VNF placement on physical substrate nodes

Problem Formulation

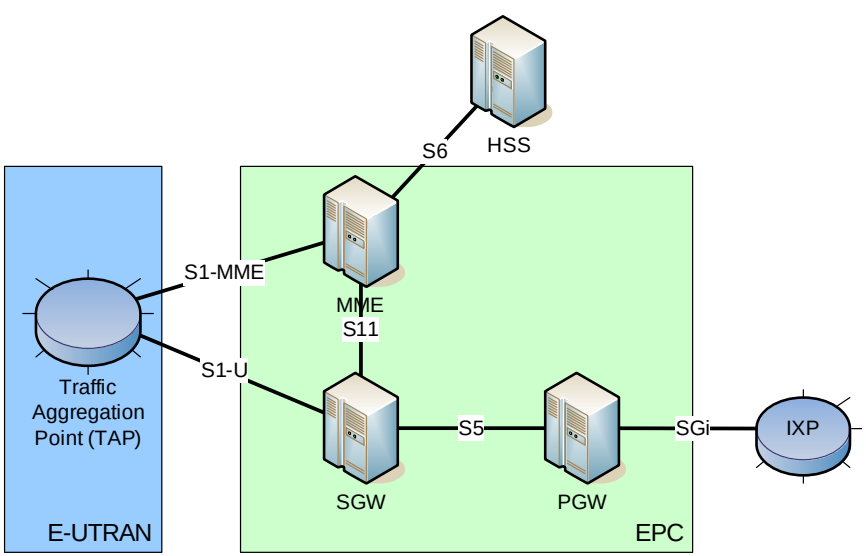
Approach

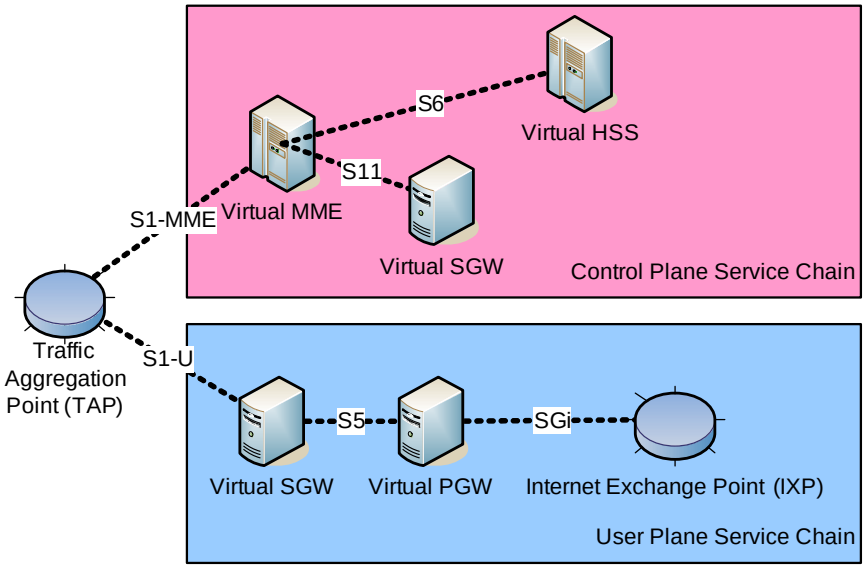
Virtual core network topology not given, but subject to optimization

- Do not focus on the embedding of a fixed topology but on the embedding of multiple service chains where each service chain comprises a single RAN traffic aggregation point (TAP) as traffic end point
- Mixed Integer Linear Programming Formulation

Problem Formulation

Considered LTE network functions and service chains



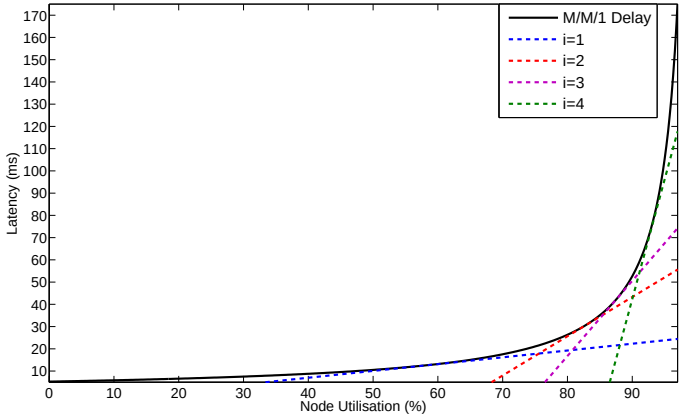


Problem Formulation

Consider Delay caused by:

- Propagation: Depending on link length
  $\rightarrow$ linear
- Packet Forwarding/Queueing: Depending on node's traffic utilization
  $\rightarrow$ Non-Linear, typically convex curve
- Processing: Depending on CPU utilization
  $\rightarrow$ Non-Linear, typically convex curve

M/M/1 model assumed for convex delay curve



Problem Formulation

Given input data

- Physical (substrate) network: Directed graph (N_s, L_s)
- Capacity $c_{(n_{s1}, n_{s2})}$ of every physical link $(n_{s1}, n_{s2}) \in L_s$
- Set of traffic aggregation points $T \subseteq N_s$
- Processing/Storage/Switching capacity of every node $n_s \in N_s$, $c^m_{n_s}$ $(m \in \{\text{pro,stor,bdw}\})$
- Link costs per occupied bandwidth unit: $\text{cost}(n_{s_i}, n_{s_j})$ for $(n_{s_i}, n_{s_j}) \in L_s$
- Basic node costs $\text{cost}(n_s)$ for $n_s \in N_s$
- Cost per occupied pro, sto, bdw unit at node n_s:

$$\text{cost}(x, n_s), \forall x \in \{\text{pro,stor,bdw}\}$$

Problem Formulation

Assumptions I

- Consider one virtual mobile core network for the embedding
- Single path routing
- Every node $t \in T$ is a traffic aggregation point for RAN traffic with demand

$$d_t^{\mathrm{VNF},m} \quad (\mathrm{VNF} \in \{\mathrm{SGW, PGW, IXP, MME, HSS}\}, m \in \{\mathrm{pro,stor,bdw}\}, t \in T)$$

- Only certain nodes are capable of hosting certain virtual network functions VNF $\in$ {SGW, PGW, IXP, MME, HSS}

$$\mathrm{suit}_{n_s}^{\mathrm{VNF}} = \begin{cases} 1 & \text{if VNF can be embedded on } n_s, \\ 0 & \text{else.} \end{cases}$$

Problem Formulation

Assumptions II

- Convex delay curves
  - $\alpha_i, \beta_i, \gamma_i, \delta_i$ Linearization parameters for the convex delay curve
  - $z_{n_s}^{\text{pro, max}}, z_{n_s}^{\text{bdw, max}}$ Big constants

Problem Formulation

Variables I

- $x_{n_s}^{t,\mathrm{VNF}} \quad \forall n_s \in N_s, \mathrm{VNF} \in \{\mathrm{SGW, PGW, IXP, MME, HSS}\}, t \in T$
  Binary variable: True, if a virtual network function of type VNF is embedded on node n_s for the traffic aggregation point $t \in T$
- $f_{(n_{s1},n_{s2})}^{t,(\mathrm{VNF}_1,\mathrm{VNF}_2)} \quad \forall t \in T, (n_{s1}, n_{s2}) \in L_s, (\mathrm{VNF}_1, \mathrm{VNF}_2) \in \{(\mathrm{TAB,SGW}), (\mathrm{SGW,PGW}), \dots\}$
  Binary variable: True, if the virtual link between $(\mathrm{VNF}_1, \mathrm{VNF}_2)$ is embedded on the physical link (n_{s1}, n_{s2}) for the demand of traffic aggregation point $t \in T$
- $z(n_{s1}, n_{s2})$ Delay on the physical link from n_{s1} to n_{s2}

Problem Formulation

Variables II

- $z_{n_s}^{t,\mathrm{VNF,pro}}$ Delay caused by the processing of VNF for TAP $t \in T$ on the physical node n_s
- $z_{n_s}^{t,\mathrm{bdw}}$ Delay caused by packet queueing for TAP t on the physical node n_s
- $u_{n_s}^{\mathrm{pro}}, u_{n_s}^{\mathrm{bdw}}$ Auxiliary variables for the utilization of the processing/forwarding resources at the physical node n_s

Problem Formulation

TECHNISCHE UNIVERSITÄT CHEMNITZ

Objective Function

Minimize:

$$a \cdot \sum_{n_s \in N_s} x_{n_s} \cdot \text{cost}(n_s) + b \cdot \sum_{n_s \in N_s} \sum_{t \in T} \sum_{\text{VNF} \in N_v} \sum_{x \in \{\text{pro,stor,bdw}\}} x_{n_s}^{t,\text{VNF}} \cdot d_t^{\text{VNF},x} \cdot \text{cost}(x, n_s)$$

$$+ c \cdot \sum_{(n_{s1}, n_{s2}) \in L_s} \text{cost}(n_{s1}, n_{s2}) \cdot \sum_{t \in T} \sum_{(\text{VNF1,VNF2}) \in L_v} f_{(n_{s1}, n_{s2})}^{t,(\text{VNF1,VNF2})} \cdot d_t^{(\text{VNF1,VNF2})}$$

Problem Formulation

Constraints I

$$\sum_{n_s \in N_s} x_{n_s}^{t,\mathrm{VNF}} \cdot \mathrm{suit}_{n_s}^{\mathrm{VNF}} = 1 \qquad \forall t \in T, \mathrm{VNF} \in N_v$$

$$x_{n_s}^{t,\mathrm{VNF}} \leq \mathrm{suit}_{n_s}^{\mathrm{VNF}} \qquad \forall t \in T, n_s \in N_s, \mathrm{VNF} \in N_v$$

$$\sum_{(w,n_s) \in L_s} \sum_{t \in T} \sum_{(\mathrm{VNF1},\mathrm{VNF2}) \in L_v} f_{(w,n_s)}^{t,(\mathrm{VNF1},\mathrm{VNF2})} \cdot d_t^{(\mathrm{VNF1},\mathrm{VNF2})} \leq c_{n_s}^{\mathrm{bdw}}$$

$$\sum_{t \in T} \sum_{\mathrm{VNF} \in N_v} x_{n_s}^{t,\mathrm{VNF}} \cdot d_t^{\mathrm{VNF},\mathrm{pro}} \leq c_{n_s}^{\mathrm{pro}}$$

$$\sum_{t \in T} \sum_{\mathrm{VNF} \in N_v} x_{n_s}^{t,\mathrm{VNF}} \cdot d_t^{\mathrm{VNF},\mathrm{stor}} \leq c_{n_s}^{\mathrm{stor}} \qquad \forall n_s \in N_s$$

Problem Formulation

Constraints II

$$\sum_{t \in T} \sum_{(\text{VNF1},\text{VNF2}) \in L_v} f^{t,(\text{VNF1},\text{VNF2})}_{(n_{s1},n_{s2})} \cdot d_t^{(\text{VNF1},\text{VNF2})} \leq c(n_{s1}, n_{s2}) \qquad \forall (n_{s1}, n_{s2}) \in L_s$$

$$\sum_{(n_{s1},w) \in L_s} f^{t,(\text{VNF1},\text{VNF2})}_{(w,n_{s1})} - f^{t,(\text{VNF1},\text{VNF2})}_{(n_{s1},w)} = x^{t,\text{VNF1}}_{n_{s1}} - x^{t,\text{VNF2}}_{n_{s1}}$$

$$\forall t \in T, n_{s1} \in N_s, (\text{VNF1},\text{VNF2}) \in L_v$$

Problem Formulation

Constraints III

$$\alpha_i + \beta_i \cdot u_{n_s}^{\text{pro}} \leq z_{n_s}^{t,\text{VNF,pro}} + (1 - x_{n_s}^{t,\text{VNF}}) \cdot z_{n_s}^{\text{pro,max}}$$
$$z_{n_s}^{t,\text{VNF,pro}} \leq z_{n_s}^{\text{pro,max}} \cdot x_{n_s}^{t,\text{VNF}} \qquad \forall n_s \in N_s, \text{VNF}, t \in T, i \in I$$

$$\gamma_i + \delta_i \cdot u_{n_s}^{\text{bdw}} \leq z_{n_s}^{t,\text{bdw}} + (1 - \hat{f}_{n_s}^{t}) \cdot z_{n_s}^{\text{bdw,max}}$$
$$z_{n_s}^{t,\text{bdw}} \leq z_{n_s}^{\text{bdw,max}} \cdot \hat{f}_{n_s}^{t} \qquad \forall n_s \in N_s, t \in T, i \in I$$

Problem Formulation

Constraints IV

$$\sum_{n_s \in N_s} \sum_{\text{VNF} \in \tilde{N}_v} z_{n_s}^{t,\text{VNF,pro}} + \sum_{(n_{s1},n_{s2}) \in L_s} \sum_{l_v \in \tilde{L}_v} f_{(n_{s1},n_{s2})}^{t,l_v} \cdot z(n_{s1}, n_{s2}) + \sum_{n_s \in N_s} z_{n_s}^{t,\text{bdw}} \leq d_{\text{user}}$$

$$\forall t \in T, \forall \tilde{N}_v, \tilde{L}_v$$

$$\sum_{n_s \in N_s} \sum_{\text{VNF} \in \hat{N}_v} z_{n_s}^{t,\text{VNF,pro}} + \sum_{(n_{s1},n_{s2}) \in L_s} \sum_{l_v \in \hat{L}_v} f_{(n_{s1},n_{s2})}^{t,l_v} \cdot z(n_{s1}, n_{s2}) + \sum_{n_s \in N_s} z_{n_s}^{t,\text{bdw}} \leq d_{\text{control}}$$

$$\forall t \in T, \forall \hat{N}_v, \hat{L}_v$$

Problem Formulation

Constraints V

$$x_{n_s}^{t,\mathrm{VNF}}, x_{n_s}, f_{(n_{s1},n_{s2})}^{t,(\mathrm{VNF}_1,\mathrm{VNF}_2)}, \hat{f}_{n_s}^{t}, x_{n_s} \in \{0,1\}$$

$$z_{n_s}^{t,\mathrm{VNF,pro}}, z_{n_s}^{t,\mathrm{bdw}} \in \mathbb{R}^+ \qquad \forall t \in T, \mathrm{VNF} \in N_v, n_s \in N_s, (\mathrm{VNF}_1, \mathrm{VNF}_2) \in L_v, (n_{s1}, n_{s2}) \in L_s$$

$$N_v = \{\mathrm{TAP,SGW,MME,HSS,PGW,IXP}\}$$

$$L_v = \{(\mathrm{TAP,SGW}), (\mathrm{SGW,TAP}), (\mathrm{TAP,MME}), (\mathrm{MME,TAP}), (\mathrm{SGW,PGW}), (\mathrm{PGW,SGW}), (\mathrm{SGW,MME}), (\mathrm{MME,SGW}), (\mathrm{PGW,IXP}), (\mathrm{IXP,PGW}), (\mathrm{MME,HSS}), (\mathrm{HSS,MME})\}$$

$$\tilde{N}_v = \{\mathrm{TAP,SGW,PGW,IXP}\}$$

$$\tilde{L}_v = \{(\mathrm{TAP,SGW}), (\mathrm{SGW,TAP}), (\mathrm{SGW,PGW}), (\mathrm{PGW,SGW}), (\mathrm{PGW,IXP}), (\mathrm{IXP,PGW})\}$$

$$\hat{N}_v = \{\mathrm{TAP,MME}\}, \{\mathrm{MME,HSS}\}, \{\mathrm{MME,SGW}\}$$

$$\hat{L}_v = \{(\mathrm{TAP,MME}), (\mathrm{MME,TAP})\}, \{(\mathrm{MME,HSS}), (\mathrm{HSS,MME})\}, \{(\mathrm{MME,SGW}), (\mathrm{SGW,MME})\}$$

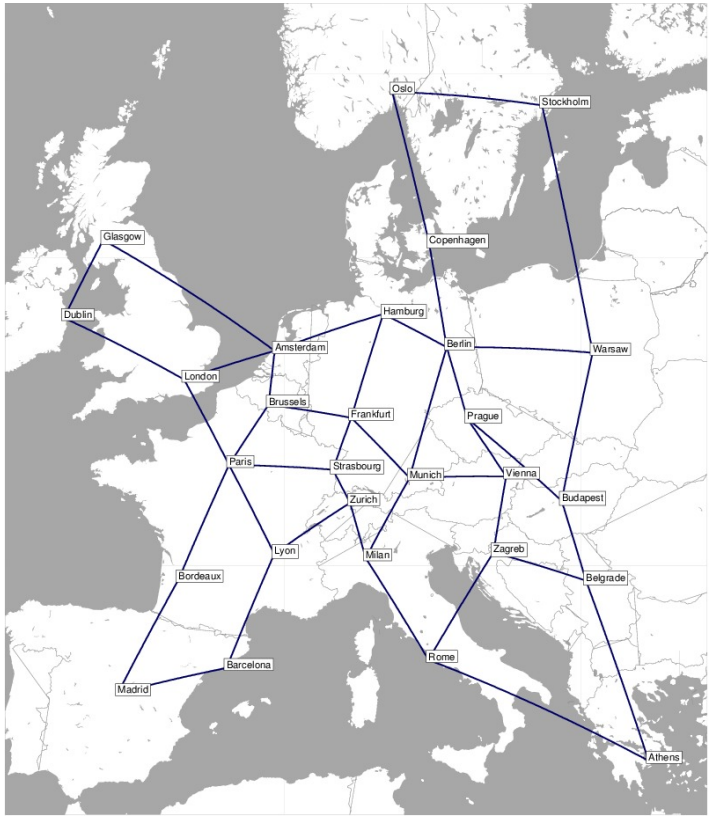
Overview

Performance Evaluation

Performance Evaluation using network topology examples from SNDlib

Considered Scenarios

- I: Given IXP nodes and given nodes for VNF placement
- II: Given IXP nodes, every node can host any type of VNF
- III: Every node can be IXP node and host any type of VNF



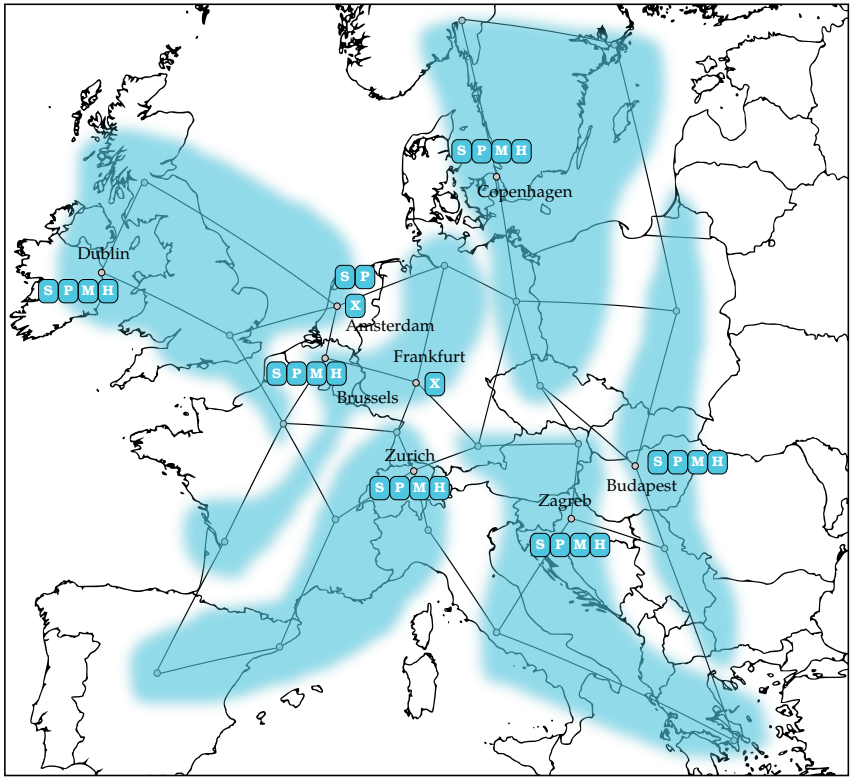
Performance Evaluation

Example: Nobel-EU network (28 nodes, 41 links)

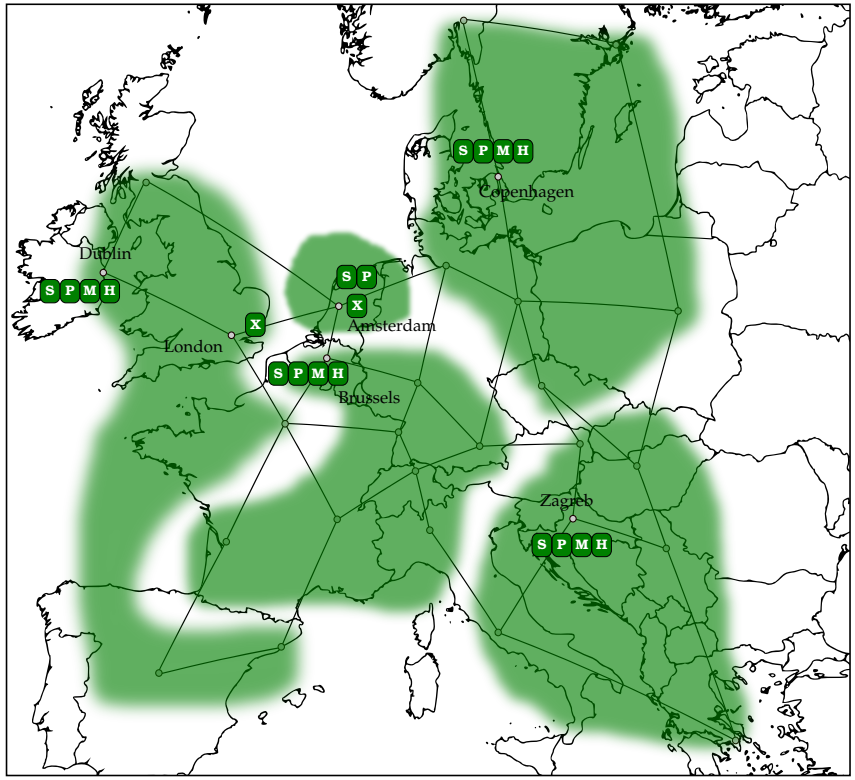
- Node costs with respect of the GDP of the specific country
- Each node is assumed to be a traffic aggregation point (TAP)
- Traffic demand scales with number of Internet users in the specific country
- Different utilization scenarios realized by a global demand scaling factor k
- Signal propagation speed $0.67 \cdot c_{\text{light}}$ (fiber connections)
- Linearized M/M/1 delay curve for processing and switching (Central buffer model)
- Use of Python/CPLEX 12.5 on Intel i7-3930K CPU, 64Gbyte RAM

Performance Evaluation

VNF placement and TAP allocation example



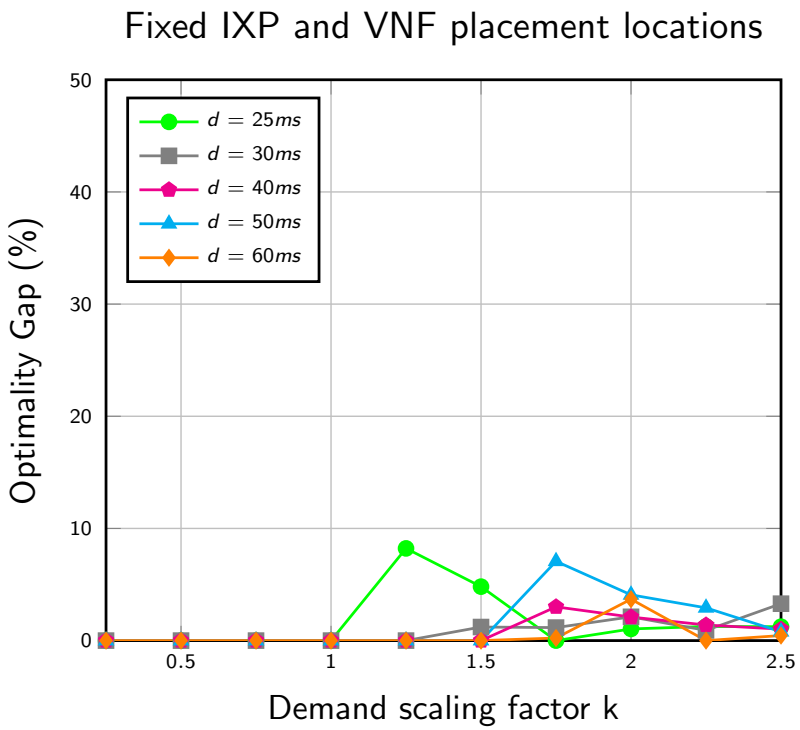
$k = 3, d_{\text{control}} = d_{\text{user}} = 25$ ms

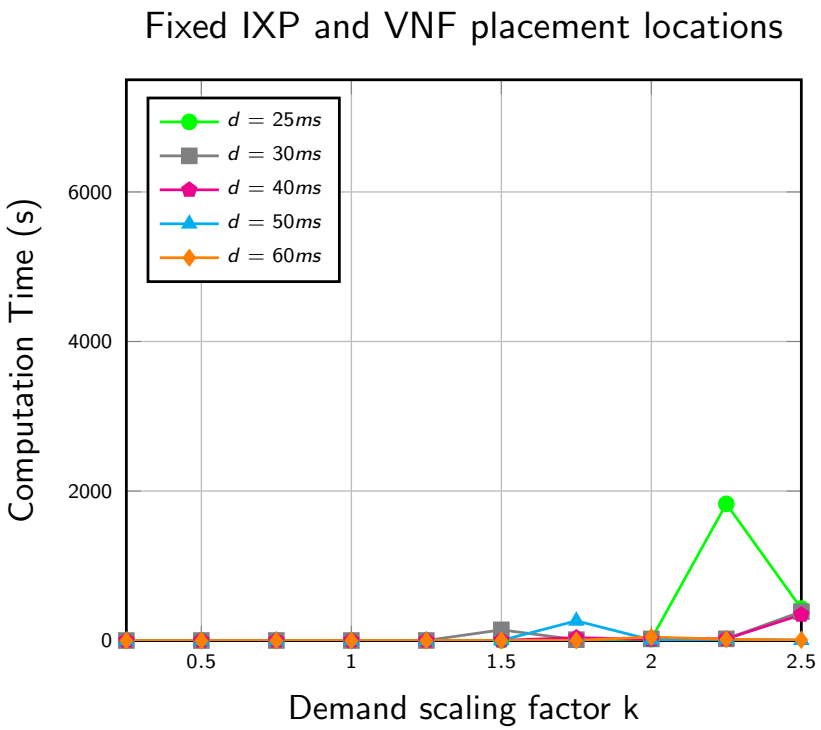


$k = 3, d_{\text{control}} = d_{\text{user}} = 50$ ms

Performance Evaluation

Scenario I

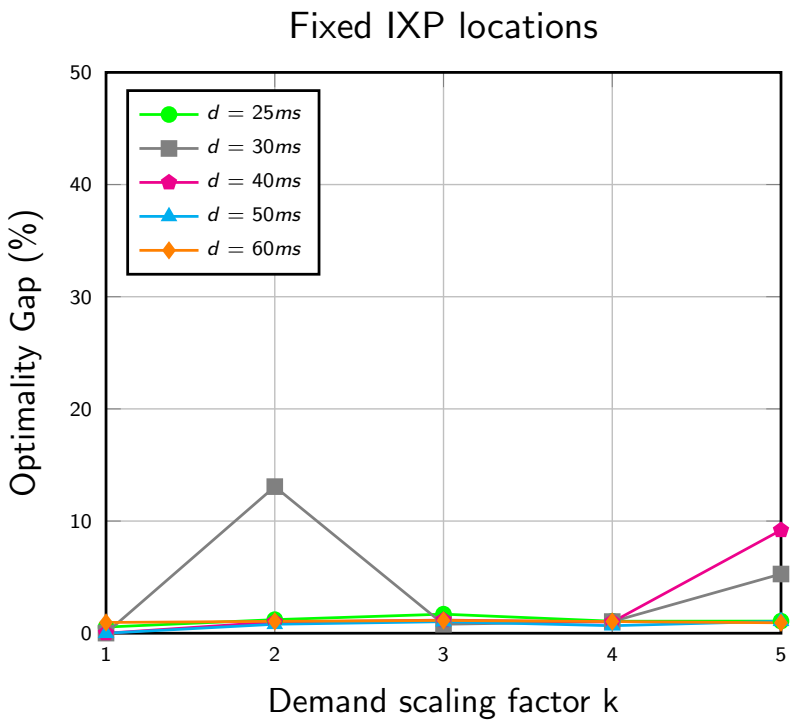


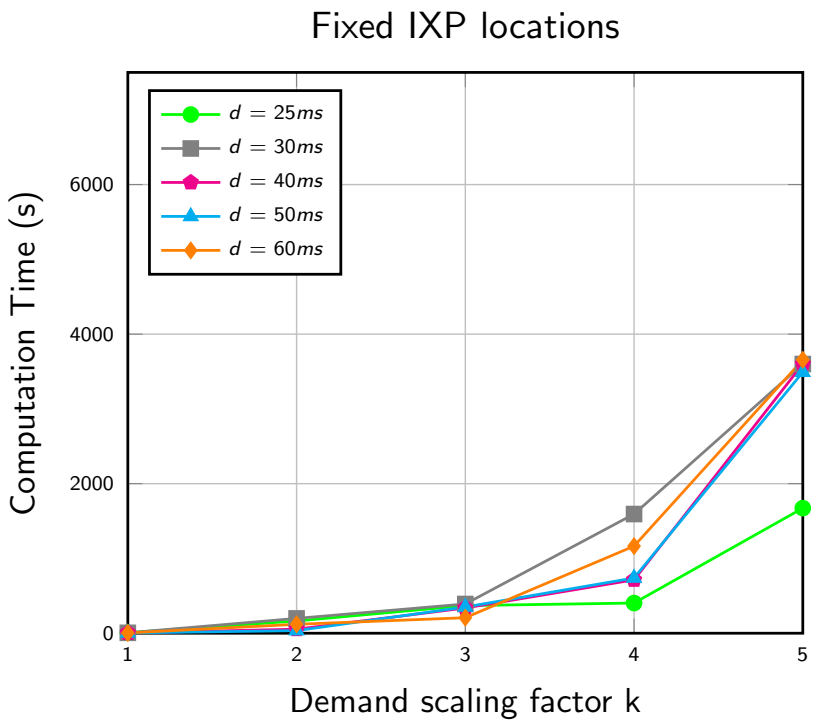


Optimality gaps and solution times for different demand scaling factors and delay bounds, Scenario I.

Performance Evaluation

Scenario II

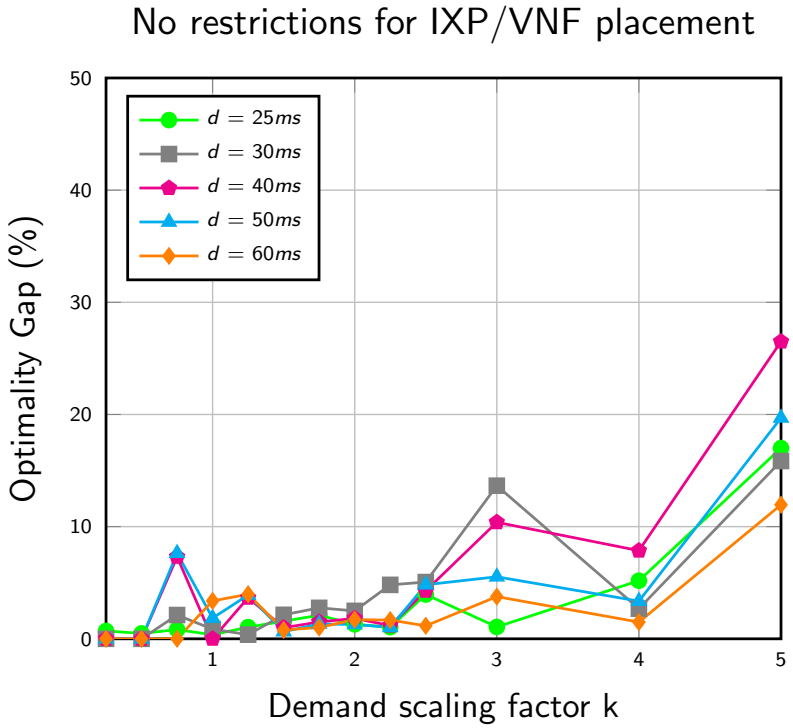


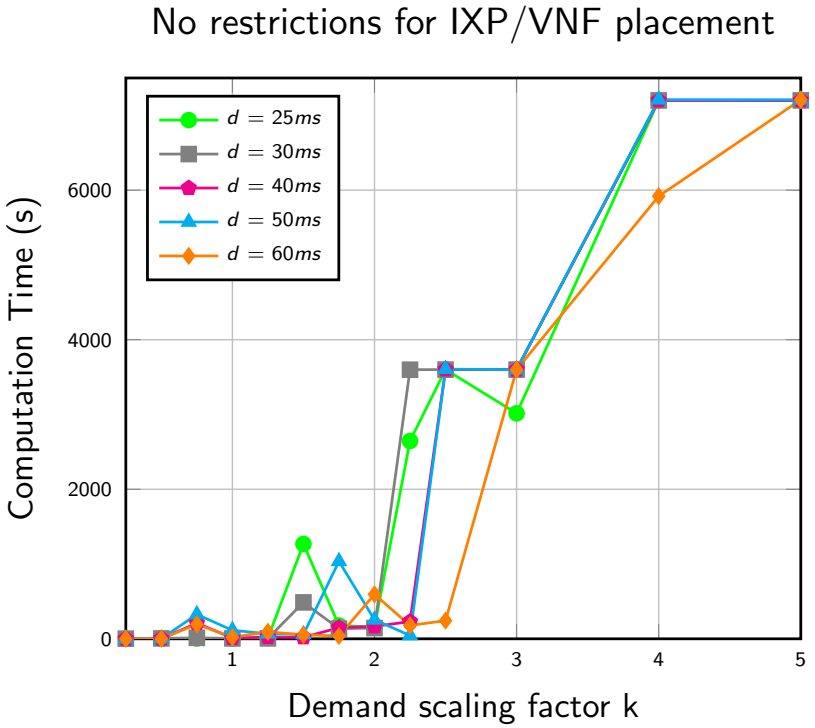


Optimality gaps and solution times for different demand scaling factors and delay bounds, Scenario II.

Performance Evaluation

Scenario III





Optimality gaps and solution times for different demand scaling factors and delay bounds, Scenario III.

Overview

Summary

MILP formulation for virtual mobile core network embedding

- Finding optimum virtual network topology together with it's embedding on a given physical substrate network
- Delay bounds are considered for each specific service chain and subchain
- Solvable in reasonable time with standard CPLEX settings for moderate network sizes

Summary

Ongoing work

- Consider robustness of VNF placement with respect to traffic variation (robust optimization)
- Consider robustness of VNF placement with respect to substrate network failures (reliability)
- Develop mathematical model for service chain embedding within data centers
- Consider control and user plane separation (SDN controller placement)