Optimum Placement of S/PDN-Gateways in the LTE Evolved Packet Core

Gerd Windisch
TU Chemnitz
Chair of Communication Networks
Chemnitz, Germany
Gerd.Windisch@etit.tu-chemnitz.de

I. INTRODUCTION

A significant increase of mobile data traffic had been observed in the last years. This trend is expected to continue due to new bandwidth consuming Internet applications [1]. Keeping the traditional EPC design with few centralized S-/PDN gateway sites will then require gateways with very high data throughput. In case of a gateway failure a large amount of traffic would be affected. Moreover, centralized gateways impose limitations on traffic engineering and load distribution. Thus a decentralized gateway placement of an increased number of gateways (with medium throughput) seems to be a better option which also leads to reduced transport cost and reduced delay.

This contribution focuses on the optimum placement of S-/PDN-gateways in an LTE EPC, assuming that the topology of the underlying IP transport network and the number and location of the Internet Exchange Points (IXPs) are predefined. The problem is formulated as a binary linear program and the LP solver CPLEX is applied. Results are presented for a realistic network scenario with 50 nodes.

II. PROBLEM STATEMENT

The LTE EPC consists of two nodes (which we assume to be collocated): the Serving Gateway (S-GW) and the Packet Data Network Gateway (PDN-GW) [2]. In order to determine the optimal position of S/PDN-GWs within the EPC (wrt. to the transport costs) it is required to consider the whole path from the access regions up to the IXPs the mobile Internet traffic can take - see Fig. 1a. Here an access region denotes a cluster of eNodeB sites with reasonable traffic aggregation (comparable to RNC sites in 3G).

III. MODELING ASSUMPTIONS

The functional elements of Fig.1a are mapped to the underlying IP-based transport network - see Fig.1b. Each site (PoP) of the core transport network represents the ingress/egress point of one access region. In our model, for simplicity, it is assumed, that the transport network topology is given and that all access regions account for the same amount of traffic. Furthermore it is assumed that the number and locations of the IXPs are given (and thus are not subject to optimization) and that also the number of gateways is given. Thus, only the gateway locations are subject to optimization.

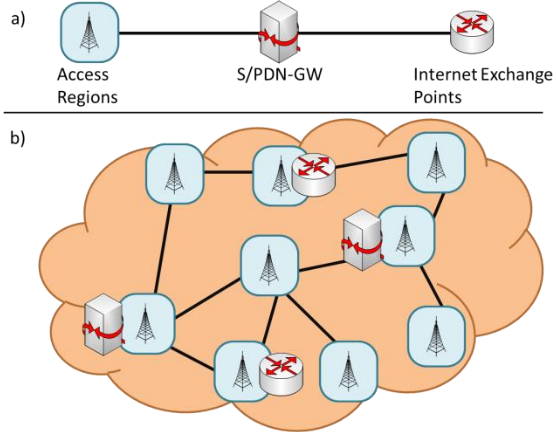

Figure 1: Network Model

Each site of the transport network could potentially host a S/PDN-GW. Regarding the placement of the S/PDN-GWs the following assumptions are taken:

- S- and PDN-GWs are always collocated
- traffic from an S/PDN-GW is routed to the closest IXP
- load balancing is applied among the S/PDN-GWs
- one access region is connected to one S/PDN-GW only

IV. FORMULATION AS AN OPTIMIZATION PROBLEM

The gateway placement problem belongs to the class of facility location problems and can be described as a binary linear program. The optimization objective is to minimize the overall path length of all traffic flows between access regions and IXPs (1) so as to minimize the transport costs.

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} (\omega_{ij} + \gamma_j) v_{ij} \quad (1)$$

Following constraints apply:

$$-\beta(1+\alpha) < \sum_{j=1}^{N} (\delta_j^d + \delta_j^u) v_{ij} - \beta(1+\alpha) u_i \leq 0 \quad \forall i = 1, \dots, |N| \quad (2)$$

Constraint (2) ensures that the S/PDN-GW capacity is not exceeded.

$$\sum_{j=1}^{N} v_{ij} = 1 \;\; \forall i = 1, \dots, |N| \;\; (3)$$

Constraint (3) ensures that an access region is only connected to one S/PDN-GW.

$$\sum_{i=1}^{N} u_i = m \;\; (4)$$

Constraint (4) ensures that m S/PDN-GW are placed.

$$\sum_{i=1}^{N} \sum_{j=1}^{N} ((\rho^1_{ijk} + \eta^1_{jk})\delta^d_i + (\rho^2_{ijk} + \eta^2_{jk})\delta^u_i) v_{ij} = \varepsilon_k \;\; \forall k = 1, \dots, |L| \; (5)$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} ((\rho^2_{ijk} + \eta^2_{jk})\delta^d_i + (\rho^1_{ijk} + \eta^1_{jk})\delta^u_i) v_{ij} = \varepsilon_k \;\; \forall k = 1, \dots, |L| \; (6)$$

The constraints (5) and (6) ensure that the capacity of a link is not exceeded.

Sets:

- N: set containing all transport network sites (PoPs)
- N_{IXP}: subset of N, containing all IXP locations
- L: set containing all links

Variables:

- u_i: = 1 if at site i a GW is placed
- v_{ij}: = 1 if access region i (site i) is served by GW j

Parameters:

- m: number of GWs to be placed
- ω_{ij}: link weight of link between site i and site j
- δ^u_i: traffic demand at site i in uplink direction
- δ^d_i: traffic demand at site i in downlink direction
- α: constant (given)
- β: GW capacity
- γ_i: path length between site i and the closest IXP
- ε_k: capacity of link k
- ρ^1_{ijk}: = 1 if direction 1 of link k is part of the downlink path between site i and j
- ρ^2_{ijk}: =1 if direction 2 of link k is part of the downlink path between site i and j
- η^1_{jk}: = 1 if direction 1 of link k is part of the downlink the path between GW site j and the closest IXP site
- η^2_{jk}: = 1 if direction 2 of link k is part of the downlink path between GW site j and the closest IXP site

The optimization model is implemented in MatLab and the CPLEX solver is applied.

V. RESULTS

As a first example the “germany50” network topology [3] is used. This topology represents a realistic network covering Germany, consisting of 50 sites and 88 links.

Fig. 2 shows the average path length for an increasing number m of S/PDN-GWs and 3 IXP sites. The red line shows the overall average path length between all access regions and their closest IXPs.

It can be seen that for more than 3 S/PDN-GWs no further reduction of the path length is achieved (even if more S/PDN-GWs would be provided). The reason is, that during optimization S/PDN-GWs are tried to be placed on the shortest path between an access region and its closest IXP. Since 3 IXPs are assumed, the first three S/PDN-GWs are placed at the IXP sites therefore being on the shortest path.

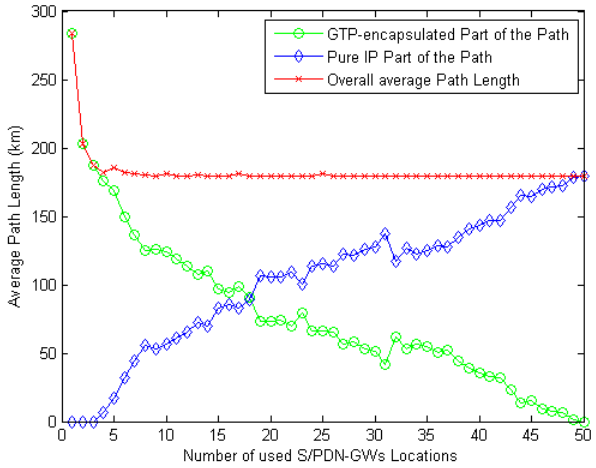

Figure 2: Average Path Length vs. Number of S/PDN-GWs

The green line in Fig. 2 denotes the path segment before the GW (where GTP-tunneling is applied) and the blue one denotes the path after the GW (pure IP). It can be seen that the first three S/PDN-GWs are collocated to IXPs and therefore no pure IP path segment is present.

VI. SUMMARY

This contribution investigates the influence of the number and location of S/PDN-GW on the average path length (for given locations and number of IXPs). One outcome is, that the path length cannot be reduced further, if the number of S/PDN-GWs exceeds the number of IXPs. However, the number of S/PDN-GWs influences the share of GTP-encapsulated and pure IP traffic. In case the encapsulation overhead of GTP would be included into the objective function the optimum solution would be to provide more S-/PDN-GWs so as to increase the pure IP traffic share in the network.

VII. ACKNOWLEDGEMENT

This work is carried out within the CELTIC MEVICO project funded by the Federal Ministry of Education and Research (BMBF) of Germany (grant 01BU1013).

VIII. REFERENCES

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2008-2013", 2009, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[2] 3GPP: “TS 23.401 Technical Specification, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”, V10.4.0, June 2011. ”

[3] Pioro, M., Wessaly, R., et al.: ”Survivable fixed telecommunication Network Design (sndLib)”, 2012 sndLib.zib.de