Rendon Schneir, J, Bradford, J, Konstantinou, K, Sumanasena, A, Zimmermann, G, Gramaglia, M. (2022) Cost assessment of elasticity for a 5G broadband network. *Transactions on Emerging Telecommunications Technology*, 33(4):e4425

COST ASSESSMENT OF ELASTICITY FOR A 5G BROADBAND NETWORK

Juan Rendon Schneir[1]*, Julie Bradford[2], Konstantinos Konstantinou[2],
Abhaya Sumanasena[2], Gerd Zimmermann[1], and Marco Gramaglia[3]

[1] Deutsche Telekom, Landgrabenweg 151, 53227 Bonn, Germany
[2] Real Wireless, PO Box 2218, Pulborough West Sussex RH20 4XB, United Kingdom
[3] University Carlos III of Madrid, Av. Universidad 30, 28911, Leganes, Spain

* Corresponding author: juan.rendon-schneir@telekom.de

Acknowledgements

This work has been performed within the framework of the 5G PPP program partially funded by the European Commission within the Horizon 2020 Framework Program. Specifically, this work has been performed within the 5G PPP Phase II project 5G-MoNArch (Grant Agreement No. 761445) with the concepts and tools developed here also being utilised on the 5G PPP Phase III project 5G-TOURS. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein.

ABSTRACT

In the highly competitive telecommunications market, operators need to consider techniques that might lead to a more cost-efficient use of the network resources. The recently proposed concept of resource elasticity focuses on the inherent cost reductions achieved by tailoring the resource utilisation of network elements, such as radio and cloud elements, to the real network demand. However, no research has been conducted so far regarding the cost benefits that such schemes could bring to 5G networks, in terms of total cost of ownership (TCO). This article makes a novel contribution by analysing the cost benefits of employing elasticity in a 5G network providing users with a broadband service in a hotspot area. The area exemplarily analysed was the Port of Hamburg in Germany, which is a realistic scenario and which receives a considerable number of additional mobile users when a cruise ship arrives at the port. The results show that, for the different scenarios analysed, the total radio access network (RAN) cost reductions of using elasticity compared to a non-elastic network range from 2.5% to 4.4%. Considering the edge cloud site-related cost component alone, a 16.4% saving from elasticity was found. Further research is needed to understand the benefits of elasticity for other types of 5G use cases in different types of areas.

Keywords: elasticity, 5G, cost, savings

1. INTRODUCTION

The deployment of 5G networks will initially address specific use cases delivering tailored services to target users in defined coverage areas. While it is clear that over time close to nationwide coverage will provide users with an improved broadband service, the provisioning of some 5G use cases will require the deployment of specialised and localised 5G infrastructure. The transition to standalone 5G networks will include major architectural changes compared to 4G [1]. These changes translate to major cost implications in the deployment of 5G networks. When planning 5G networks, the telecommunications industry and mobile network operators (MNOs) will need to consider different techniques that could lead to a cost-efficient use of the network resources while at the same time defining a network that meets customer requirements [2].

The future 5G mobile network design will be affected by heavy use of virtualisation and cloudification technologies not only for core network (CN) functionalities, but also for those of the radio access network (RAN). O-RAN is an example of a distributed and cloudified RAN architecture that is being investigated by the industry [3]. Based on the principles of software-defined radio (SDR), software-defined networking (SDN), and network function virtualization (NFV), the corresponding control plane (CP) and user plane (UP) processing functions will be flexibly orchestrated. This flexibility will be achieved by employing a highly *softwarised* infrastructure environment consisting of distributed sites with processing capabilities interconnected by broadband fixed fibre links. RAN processing functions for a high number of radio cells are typically concentrated in so-called edge cloud sites - following a virtual RAN (vRAN) concept -, whereas CN functions are hosted in central clouds. Related architectural approaches are described in [4-6].

By *softwarising* the network through NFV and SDN, operators can embrace a more flexible way of operation, based on software deployment cycles rather than recurring to the hardware maintenance phases typical in the physical networking functions. Moreover, beyond this already clear advantage, *softwarised* networks allow for a more flexible usage of the resources that are able to match the real demand with the available resources, quickly adapting to the demand without providing any disruption to the user. Examples of telco cloud resources are the central processing unit (CPU) and random access memory (RAM), whereas examples of radio resources are frequency bands and spectrum. We refer to this capability as resource elasticity.

In such environments the concept of resource elasticity, which adapts computational and radio resources in both temporal and spatial domains to the changing demands of the network, can dramatically increase resource efficiency. These improvements in resource efficiency can be achieved by reducing over-provisioning while allowing for network self-dimensioning, smart resource distribution, and orchestration [7].

Several studies have investigated different aspects of cloud radio access networks [8-10]. While the advantages and employment of resource elasticity in 5G wireless networks have been studied in the literature [7, 11, 12], no studies have been published so far regarding the cost implications in terms of the impact on the total cost of ownership (TCO) of elasticity in large-scale 5G networks. A few cost analyses regarding 5G have been published over the last few years, but they have not covered all cost aspects of 5G [13-15]. In this article we make a novel contribution by addressing the following research question: *What are the cost benefits of a 5G broadband network that employs elasticity?*

To this end we have calculated the TCO by employing a cost model that enabled us to quantify the capital expenditures (CAPEX) and operational expenditures (OPEX). We have analysed the network resources that would be needed to provide an enhanced mobile broadband (eMBB) service in an example area. The selected area was the Port of Hamburg in Germany, which receives a considerable number of visitors when cruise ships arrive at the port. These visitors are temporarily concentrated at the cruise ship terminals in the port creating temporary demand hotspots that put pressure on the existing mobile network. In our network dimensioning we consider the use of elasticity to reduce over-provisioning for these temporary hotspots. The research was conducted as part of the European Union (EU) 5G-PPP Phase 2 project 5G-MoNArch, in which different aspects of elasticity in a 5G environment were studied [16]. The cost modelling tool helped us to estimate the cost savings achieved via network elasticity when providing services in temporary demand hotspots.

The remainder of this article is organised as follows. In section 2 we explain the concept of elasticity and its potential importance for the planning of telecommunications resources in a 5G environment. Section 3 describes the 5G network architecture employed in the analysis. The cost analysis approach is explained in section 4. Following this, we show in section 5 the results of the cost analysis and potential savings in the baseline case. The results of our sensitivity analysis are then presented. Furthermore, a summary of key findings across all results is presented and areas of further research are indicated. Finally, we conclude the article in section 6.

2. ELASTICITY IN 5G NETWORKS

Future 5G networks will require a flexible and dynamically configurable network infrastructure. This more flexible network infrastructure will enable the support of a large variety of 5G services with their wide range of diverging requirements from eMBB to ultra-reliable and low-latency communications (URLLC) and massive machine-type communication (mMTC) services [17]. In addition to this service variety, the use of network slicing leads to an additional dimension to be considered in the network set-up both from a technical and from a business perspective [4, 5]. Network slicing will permit the provisioning of isolated logical networks, which may be dedicated to different tenants e.g. from vertical industries, onto a common infrastructure [18].

To cope with such demands, virtualisation and cloudification have become central features for the realisation of 5G networks in an economically suitable fashion. 5G network architectures, as described, allow the setup of slice instances across a network infrastructure typically consisting of a higher number of edge cloud sites than central cloud sites [5, 6, 19, 20]. These cloud sites, which are interconnected by mesh-, ring-, and/or tree-based broadband transport networks, provide the needed processing infrastructure usually based on servers with general purpose processors (GPPs) capabilities. Edge cloud sites are connected to antenna sites which host only a minor part of the whole processing chain, primarily just the radio front-end. Dependent on the service requirements (latency, reliability, throughput, etc.) virtual network functions (VNFs) of the control and user planes will be orchestrated during the network slice initiation onto the different servers of the cloud sites.

By applying elasticity to VNF lifecycle management, the network can flexibly adapt its resources to the changing demands in both temporal and spatial domains, minimising both over-provisioning and performance degradation [7], regardless of the underlying virtualisation technology (either using virtual machine or container technologies). That is, resource elasticity makes the VNF implementation aware of the underlying computational and radio resources, so they can be optimised jointly as done traditionally with other kind of resources (e.g., resource blocks in a radio scheduler). This helps to utilise resources better while fulfilling the stringent key performance indicators (KPIs) that 5G services will require. While elasticity is a well-known principle from cloud computing and also from radio resource management, combining elasticity in both of these areas in the NFV environment is a very novel concept as, with 5G deployments, the advantages that elasticity brings have to be considered especially for the mixture of computing and radio resources in the RAN processing [7, 11, 12]. Elasticity can be applied in several ways in 5G networks. In the approach derived within the 5G-MoNArch project, elasticity was applied for a multi-slice/multi-tenant network architecture differentiating between computational elasticity, orchestration-driven elasticity, and slice-aware elasticity [7, 16, 21].

As discussed in [7], elasticity has three dimensions in the context of networking: computational elasticity, orchestration driven elasticity, and slice-aware elasticity. They are summarised in Fig. 1.

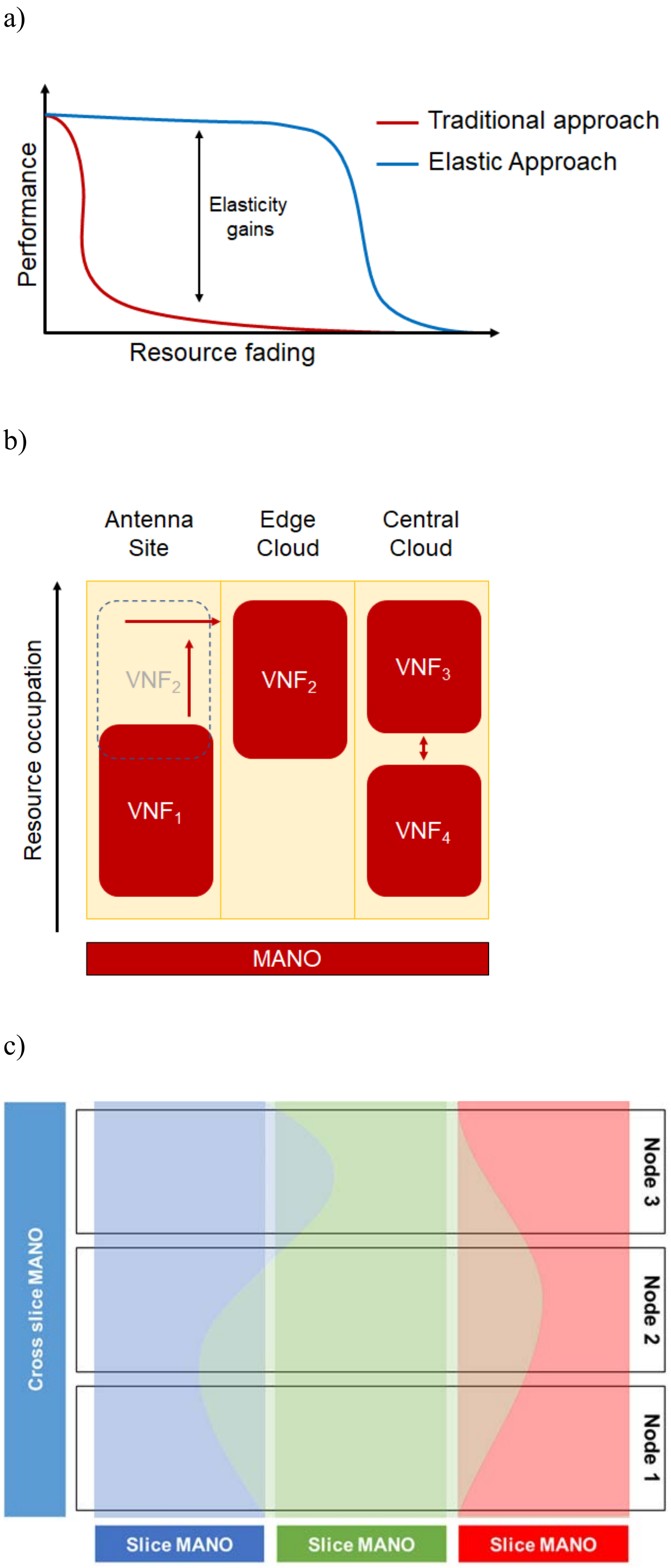

Figure 1. Three dimensions for elasticity: a) Computational elasticity; b) Orchestration-driven elasticity; c) Slice-aware elasticity.

Computational elasticity acts at the VNF design level by introducing the ability to scale them based on the available computational resources. For example, in case of computational resource shortage, VNFs would autonomously adjust their operation to reduce the consumption of computational resources while minimising the impact on network performance. This is depicted in Fig. 1a that is showing the behaviour of a traditional VNF when CPU is fading, with a “graceful degradation” instead of an abrupt decrease. The other two dimensions of elasticity, orchestration-driven elasticity and slice-aware elasticity, operate at the orchestration level. Orchestration-driven elasticity means increasing the flexibility of the orchestrator in terms of VNF placement decisions. This means that VNFs may be placed in a dynamic manner in the infrastructure layer according to resource occupation and availability, as depicted in Fig. 1b. This placement may also be under the constraint of service requirements such as latency or the availability of resources, as also depicted in Fig. 1b, which depicts the re-orchestration of a VNF from the antenna site to the cloud due to the lack of resources. In slice-aware elasticity, an end-to-end cross-slice optimisation is carried out. That is, multiple network slices deployed on a common infrastructure can be jointly orchestrated and controlled in an efficient way while guaranteeing slice isolation over a set of heterogeneous resources. Fig. 1c depicts a situation in which, thanks to a cross slice orchestration framework, the load experienced by different slices is balanced among the different nodes according to the peaks and troughs. While elasticity has mostly been evaluated in small scale scenarios (see [22] for a recent proposal of orchestration and computational elasticity solution) in this paper we propose a larger scale evaluation.

Fig. 2 shows another example of the usage of orchestration-driven elasticity, which is the type of elasticity assumed in our study. In the case of a temporarily heavy load in the edge cloud site on the right-hand side of the 5G radio network, additional computational resources from other edge clouds serving more quiet areas may be used. As this requires a change in delay due to the shift of VNFs to cloud sites further away from the antenna sites, the orchestrator must consider the latency restrictions for services transferred.

Resource elasticity approaches will be further assisted by artificial intelligence/machine learning (AI/ML) methods reducing the need for human interaction in the management and network orchestration (MANO) process [7, 23]. The type of elasticity employed in our analysis was orchestration-driven elasticity.

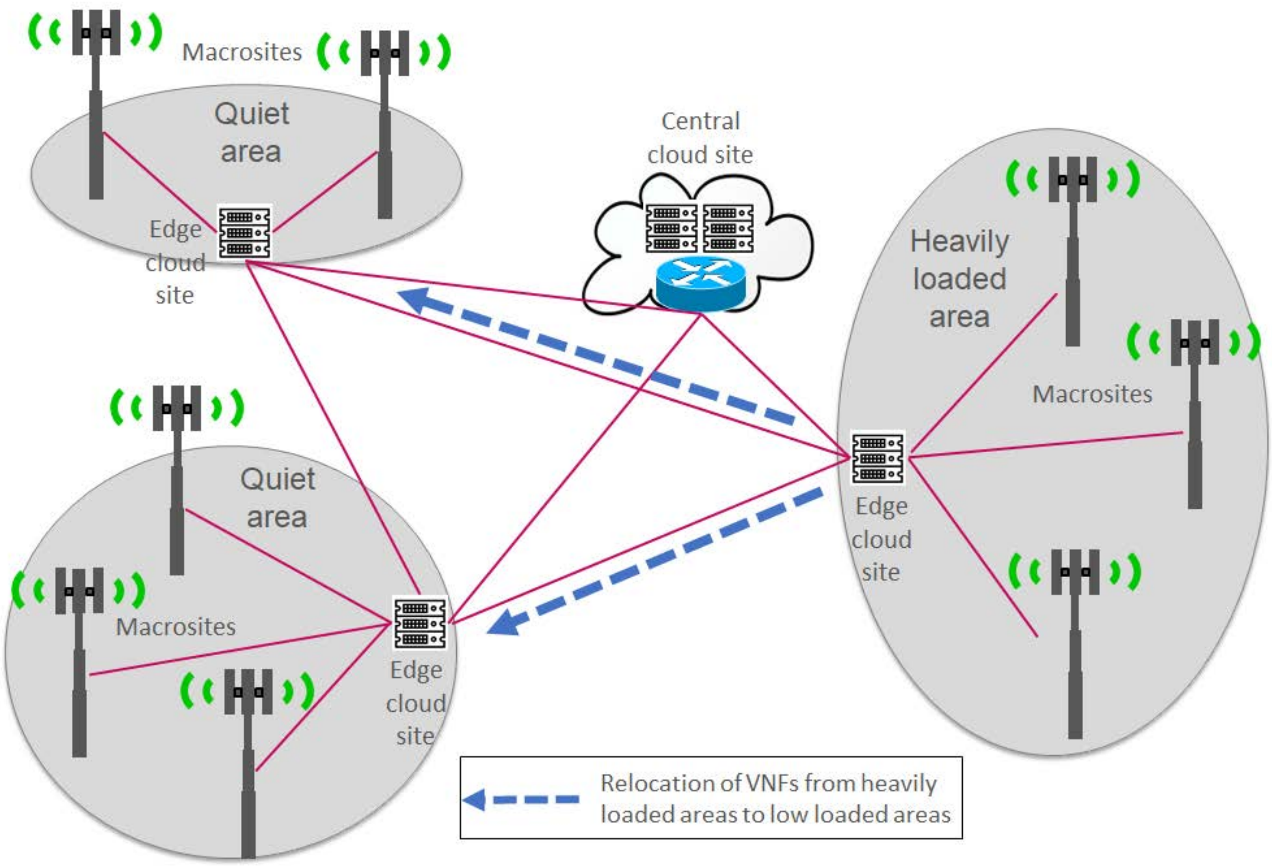

Figure 2. Concept of orchestration-driven resource elasticity between edge cloud sites. VNFs from edge cloud sites located in heavily loaded areas are relocated to edge cloud sites located in quiet areas.

3. NETWORK ARCHITECTURE

3.1 Study area

We selected Hamburg, in Germany, as the setting for our analysis as it is representative of a large European city with a busy city centre located close to its port area. Hamburg contains three cruise ship terminals which are good examples of temporary demand hotspots. Fig 3. shows the area considered which includes the city centre of Hamburg, to the north of the river Elbe, and the more industrial port area of the city where the Hamburg Port Authority (HPA) operates, to the south of the river Elbe. The city centre has a residential population of approximately 150,000 people compared with only 3,000 people for the industrial port area [21]. Fig. 3 also shows the locations of the assumed existing cellular sites in the area [21]. There is noticeably a higher density of sites in the north of the study area compared with the south reflecting the stark difference in population density between the two areas. In addition to cellular sites, Fig 3. shows the assumed location of four edge cloud sites. These would be required to run the network functions for the cellular sites under a virtualised network (as described in section 3.3). The edge cloud site locations have been selected based on the density of existing fixed telecoms exchanges and data centres in the area [21]. Reusing existing infrastructure in this way helps to reduce the cost of deploying new edge cloud sites.

The industrial port area includes Hamburg's Steinwerder cruise ship terminal. As this terminal is in one of the deeper parts of the port, it regularly accommodates the largest cruise ships visiting the Port of Hamburg. These ships can have up to 4,300 passengers and 1,500 staff onboard putting pressure on the existing mobile network in the area [21]. However, these ships typically arrive early on a Saturday morning when the city eMBB traffic is not yet at its peak.

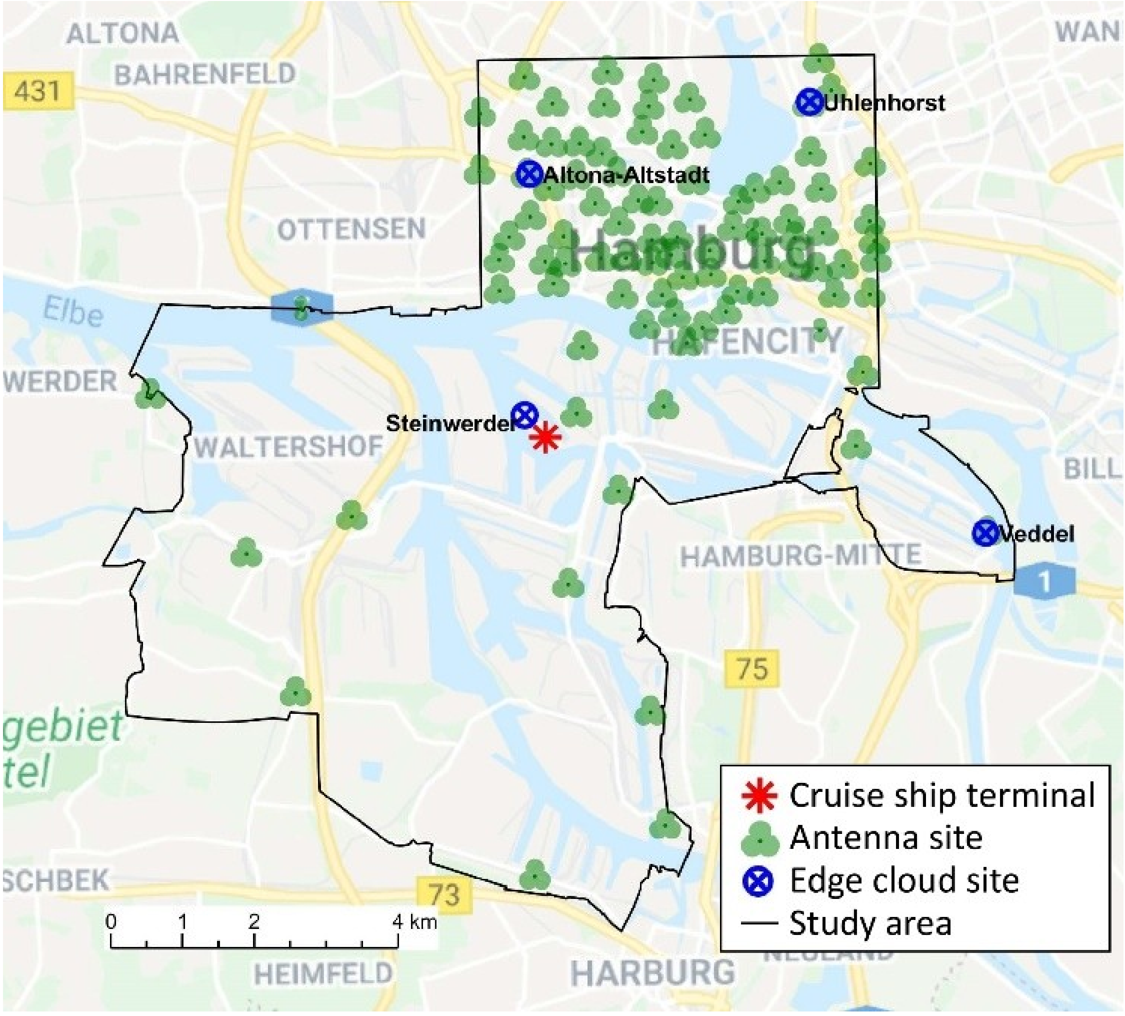

Figure 3. Study area considered of Hamburg city centre, the river Elbe and Hamburg's port area.

3.2 5G simulated service

Telecommunications operators are cautious about the size of investment in 5G networks that they make. The first service that 5G networks are already providing in initial 5G rollouts worldwide is an improved broadband service. Therefore, in our analysis we follow a conservative approach and we focus on eMBB services in temporary hotspot locations that put pressure on the existing wider area network.

Large cruise ships arriving at the Steinwerder cruise ship terminal, shown in Fig. 2, create temporary hotspots of eMBB demand until all passengers have dispersed from the terminal area. Provision for the hotspot eMBB service is made against the backdrop of delivering eMBB to consumer portable devices across the wider city area also. We dimension our eMBB network to allow for a minimum cell edge throughput of 10 Mbps to outdoor users. This would be sufficient to stream high definition video, which is challenging to achieve in cell edge conditions on existing LTE networks. We do not consider extreme 5G use cases such as immersive applications requiring high bandwidths and low latencies or URLLC services for controlling machinery.

Regarding the business model, it was assumed that the telecommunications operator deploys the required network infrastructure. The telecommunications operator would then get revenue either from Cruise Gate Hamburg, who belongs to the HPA and who manages the cruise ship terminal, or from visiting shipping liners which provide their passengers with enhanced connectivity around the terminal area. The end-users would not pay extra money to the HPA for this service. Revenues are not considered further in our analysis presented here as our focus is on TCO and potential cost savings from elasticity.

The forecast of the mobile traffic demand over the time period 2020-2030 is based on the following assumptions:

- Cruise ship passengers and staff consume 25% of their daily eMBB traffic at the cruise ship terminal [21]. This factor is higher than typical daily to busy hour traffic conversion assumptions for eMBB. The assumed higher concentration of data demand at the time when passengers arrive into port is due to two effects. Firstly, connectivity on the ship whilst outside of the port will have been limited, leading to a surge in demand when connectivity is re-established in port. Secondly, passengers are likely to download maps and employ other applications to plan their day on arrival at the port.

- We consider low, medium, and high traffic demand scenarios derived and extrapolated from the Cisco visual networking index (VNI) forecast for Germany [24]. For the low demand scenario an initial compound average growth rate (CAGR) of 20% was assumed which slowly reduces over time. The high growth scenario follows the Cisco VNI forecasts initially, for the years where Cisco VNI forecasts are available, and then applies an initial 30% CAGR which reduces over time. The medium scenario uses the average CAGR of the low and high scenarios [21]. The daily eMBB traffic per user in 2020 for our medium demand case was 77 MB, i.e., 2.3 GB per month for each user. By 2030 this grows to 480 MB per day or 14.7 GB per month.

3.3 Network architecture

Fig. 4 shows the network architecture considered. A C-RAN architecture was assumed. Four geographically separated edge cloud sites are connected to the MNO's core network and interconnected with each other to make use of elasticity. One edge cloud site is located at the Steinwerder's cruise ship terminal, while the other three edge cloud sites are located in other areas across the city (see Fig. 3). Each edge cloud site has several macrocell sites connected via a fronthaul connection to serve the eMBB traffic across the study area. Edge cloud sites are also equipped with GPP cores, installed in server cabinets, to provide the radio access protocol stack for attached macrocell sites. All small cells (SCs) deployed in and around the terminal building are connected to the Steinwerder edge cloud site. The 5G user equipment (UE) can be connected to macrocell and small cell sites.

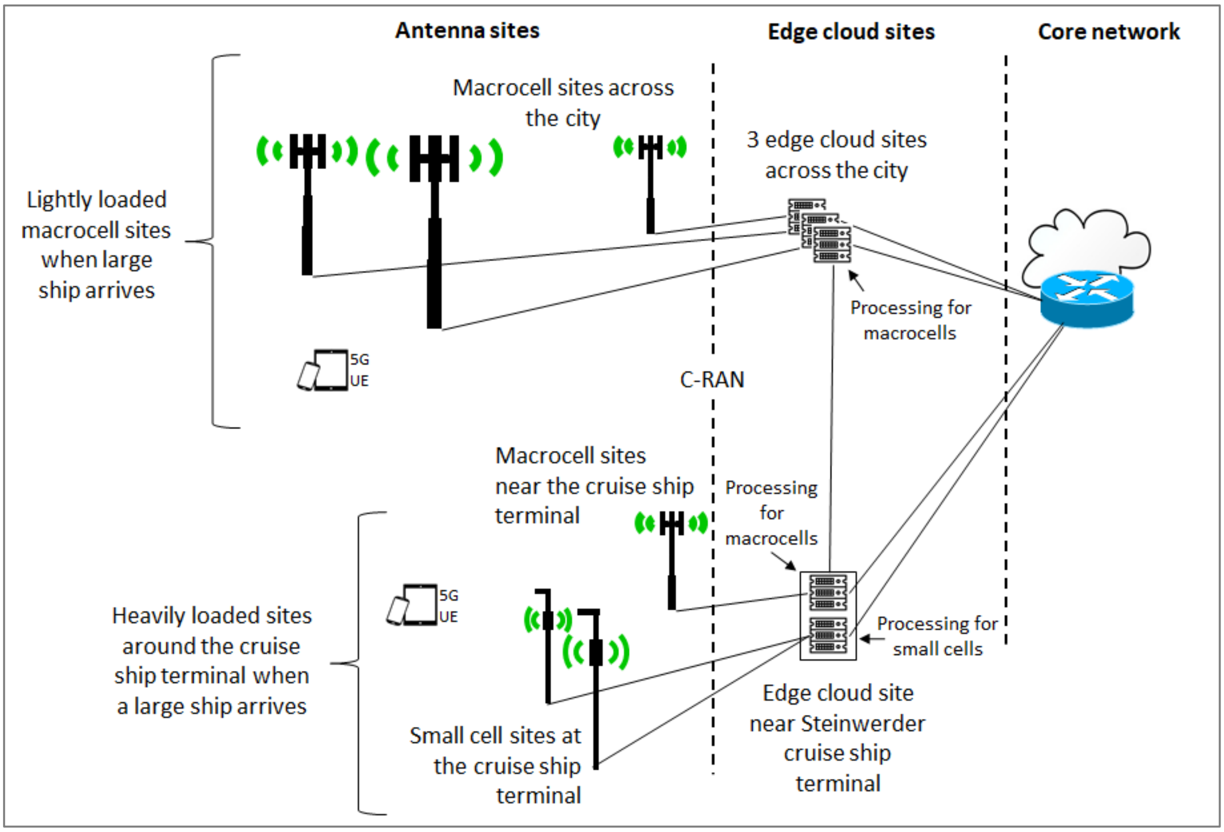

Figure 4. Network Architecture.

The wider area eMBB network is based on the existing antenna and edge cloud sites in the area. This wide area network is expanded over the time period 2020-2030 to accommodate the growing demand for eMBB services anticipated across the study area over this time. Network expansion includes upgrades of existing macrocell sites - including additional frequency bands, bandwidth or increases in multiple input multiple output (MIMO) order - or construction of new antenna sites. As new macrocell sites are added or upgraded, capacity at the existing edge cloud sites is expanded to accommodate the requirements of all antenna sites being served. Table 1 shows several parameters employed for the network dimensioning and cost calculation.

A brownfield rollout was employed for macrocell, edge cloud and small cell sites. In the case of macrocell sites, a network of existing antenna site locations and configurations is assumed to already exist at the start of 2020. The existing antenna site locations and configurations used have been selected to be representative of a typical existing mobile operator in the area [21].

In the case of edge cloud sites, we assume that fixed telecom exchanges already exist at the start of 2020 and that the MNO has access to these to install edge cloud equipment. As such, site acquisition for edge cloud data centres is not considered in our analysis but CAPEX for additional equipment as demand grows and on-going OPEX for these sites are.

In the case of small cells, we assume a low volume of these in the existing wide area network at the start of the study period, as would be representative of a typical operator in the area. We assume no existing infrastructure at the cruise ship terminal. The incremental cost for serving the demand hotspot at the cruise ship terminal comes from deploying a SC network at the cruise ship terminal. Note that in terms of the spectrum bands shown on Table 1, for SCs each MNO is assumed to start with 20 MHz of spectrum at a low frequency band (i.e., 1.8 GHz, 2.1 GHz, and 2.6 GHz spectrum bands) for their small cell layer in 2020 which increases over time. We assume that a further 20 MHz per MNO at a medium frequency band (i.e., around 3.5 GHz) becomes available for their SC layer by 2023 [21].

Our analysis focuses on elasticity in the sense of flexible use of processing resources in the network, as opposed to radio resources. For this reason, the impact of elasticity is limited to the server equipment installed on the edge cloud sites and will not impact the dimensioning or costs of the macrocell and small cell sites.

4. COST MODELLING

To calculate cost efficiencies, the TCO was calculated. In our analysis, TCO is the sum of CAPEX and OPEX values over the time period 2020-2030. No discount rate was applied or net present value (NPV) employed for the cost analysis. We considered the CAPEX and OPEX of macrocells, small cells and edge cloud sites. The cost of the backhaul is included for each of these components.

The network dimensioning behind the cost analysis utilises a tool called CAPisce for the wide area network. CAPisce is a tool for dimensioning cellular networks that has been developed by Real Wireless and was used initially for the United Kingdom (UK) regulator Ofcom [25] before being enhanced to support 5G virtualised networks under 5G-NORMA [26] and 5G-MoNArch [21]. This tool looks at the mobile demand generated across the study area in each year from 2020 to 2030. The demand in each year is compared with the capacity of existing sites in the area. In the case of a deficit in network capacity, the tool selects the most cost-effective option of extending capacity by either upgrading existing sites or adding new sites.

For each year, we capture the site and equipment volumes and translate these to CAPEX and OPEX for running such a network. The unit costs shown in Table 1 reflect average Western European values that were derived under the 5G-MoNArch project [21], and do not reflect the specific values of any single telecommunications operator. The frequency bands and assumed available bandwidths match average MNO spectrum holdings for Germany and assumptions on the future release of spectrum, again reviewed under the 5G-MoNArch project. We note that there is a mix of Frequency Division Duplex (FDD) and Time Division Duplex (TDD) spectrum available depending on the frequency band being considered. For the TDD spectrum, we assume that any bandwidth is used as a supplemental downlink to add extra downlink capacity to an existing FDD connection rather than being divided between uplink and downlink.

Our cost analysis is exclusive of CN, orchestration, administrative, and marketing costs. The CAPEX and OPEX of the macrocell sites and small cells refer to an example of a configuration with low band, 2x4, and 20 MHz.

Parameter	Value
Spectrum bands: - Sub 1GHz - Low – paired - Low – unpaired - Medium	 700, 800, 900 MHz 1800, 2100, 2600 MHz 2600 MHz 3400-3700 MHz
Spectrum bandwidths, with respect to above bands	10 MHz (FDD) 20 MHz (FDD) 20 MHz (TDD) 20 MHz (TDD)
Number of sectors: - Macrocell sites - Small cell sites	 3 2
Propagation model	ITU Hata model
Shadow fading	Log normal
MIMO orders	Up to 32 port active antennas
Number of sites (2020): Macrocell Small Cell Edge Cloud	 90 5 4
Macrocell site CAPEX (Civil works and site acquisition, antennas/feeder, and RF front end)	€ 80,000
Macrocell site OPEX (Site rental, power costs, licensing and maintenance, and site transmission)	€ 20,000 per year
Small Cell CAPEX (Civil works and site acquisition, antennas/feeder, RF front end, labour, and site transmission)	€ 27,000
Small Cell OPEX (Site rental, power costs, licensing and maintenance, and site transmission)	€ 2,000 per year
Edge cloud site CAPEX (Processing servers and cabinets, labour, and site transmission)	€ 51,000 (Example of adding 4 new servers and renewing 4 existing servers, every 3 years)
Edge cloud site OPEX (Site rental, power costs, site visits on-site maintenance, licensing and maintenance, and site transmission)	€ 52,000 per year (Example of typical edge cloud site required in the study area by 2025)

Table 1. Simulation parameters for network planning and cost analysis.

5. RESULTS

Three different cases were studied. The first case corresponds to the baseline case, whereas cases 2 and 3 were made as part of a sensitivity analysis. For all the cases studied we calculate the TCO both with and without elasticity. Each case contains a number of scenarios which are described in the following:

• **Case 1:** Baseline cost saving of elasticity. This case examines the computational resources required in the macrocell network and small cell network at the following two times of day:

- Scenario 1a: Midweek 5pm to 6pm, busy time for the city, cruise terminal is quiet.
- Scenario 1b: Saturday 6am to 7am, cruise ship arrives at the terminal; the terminal is busy, the city is quiet.

Comparing the above two scenarios we can determine if the processing required for the small cell network at the cruise terminal can be accommodated in the spare processing of the wider area macrocell network at the time when the city is quiet. This indicates the potential computational resource and related cost savings of elasticity. The baseline case corresponds to a situation where the traffic demand is medium and the market share of the corresponding MNO is 33%.

• **Case 2**: Sensitivity analysis - the effect of mobile traffic demand on the benefits of elasticity. The effect of increasing or decreasing the assumed eMBB traffic per user is studied in this case. The process of modelling the computational resources required at two times of day, as done for case 1, was re-applied for the following three scenarios to determine the potential savings from elasticity:

- Scenario 2a: Low traffic demand.
- Scenario 2b: Medium traffic demand. Note this is the baseline case for comparison and the same as case 1.
- Scenario 2c: High traffic demand.

For all case 2 scenarios the market share of the corresponding MNO is kept at the baseline level of 33%. Only the demand per user is varied. As outlined in section 3.2, different CAGR trends are assumed for the low, medium and high mobile traffic forecasts employed in case 2. The resulting eMBB traffic forecasts for Germany are shown in Fig. 5.

• **Case 3:** Sensitivity analysis - the effect of market share of hotspot traffic on elasticity savings. The effect of the number of MNOs at the cruise ship terminal, and hence market share of hotspot traffic, is analysed in this case. The process of modelling the computational resources required at two times of day, as done for case 1, was re-applied for the following three scenarios to determine the potential savings from elasticity:

- Scenario 3a: 33% market share assuming 3 MNOs. Note this is the baseline case for comparison and the same as case 1.
- Scenario 3b: 50% market share assuming 2 MNOs.
- Scenario 3c: 100% market share assuming only 1 MNO addressing the demand hotspot.

For all case 3 scenarios the eMBB demand is kept at the baseline medium traffic demand level.

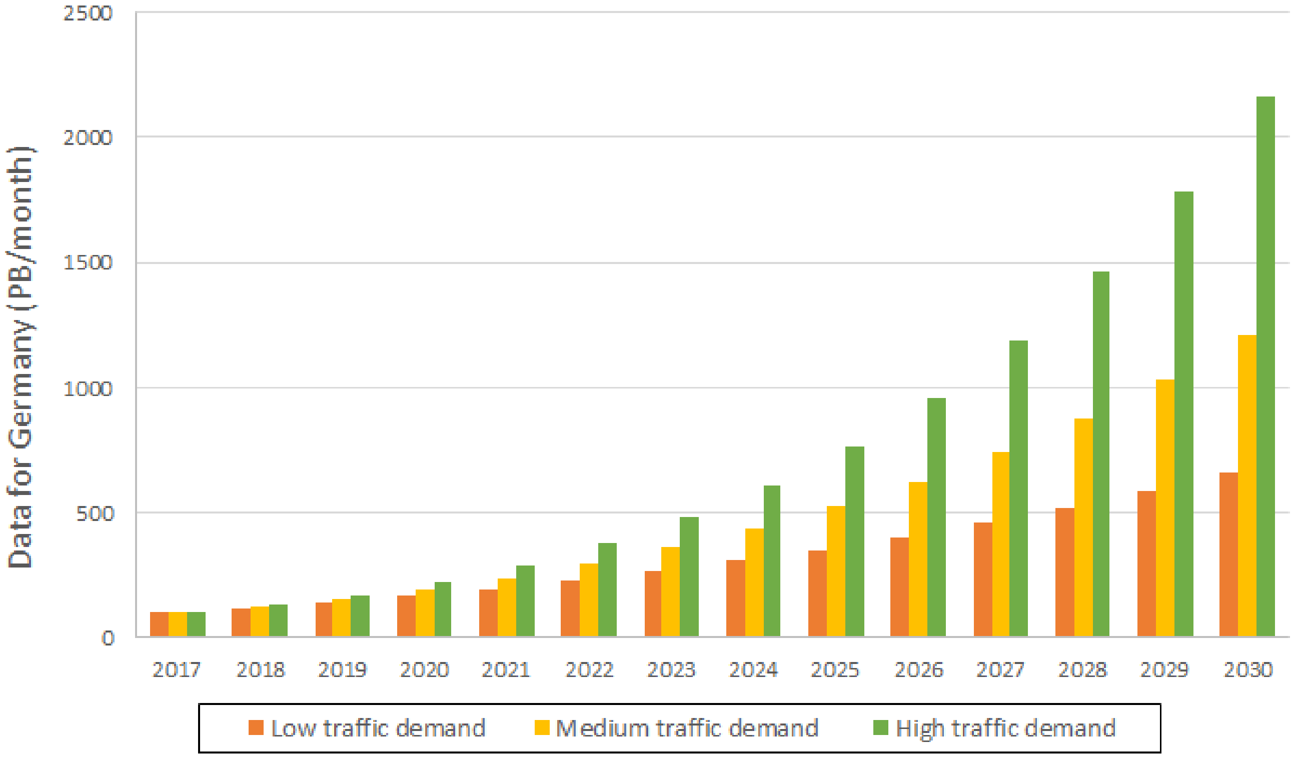

Figure 5. eMBB demand forecast for Germany for low, medium and high traffic scenarios.

5.1 Case 1: Baseline

Fig. 6 shows the cost drivers of the whole RAN infrastructure of the wide area network plus hotspot network over the time period 2020-2030 for the baseline case when no elasticity is employed. Fig. 6 shows that, for this time period, the edge cloud site CAPEX is 7.5% and the edge cloud site OPEX 7.7%. Therefore, the sum of CAPEX and OPEX for the edge cloud sites correspond to only 15.2% of the whole RAN infrastructure costs. As computational elasticity only applies to the processing at edge cloud sites, 15.2% represents the maximum possible cost benefit of elasticity in this case. In practice, the benefit will be much less than this as even with elasticity some edge cloud sites and processing will still be needed.

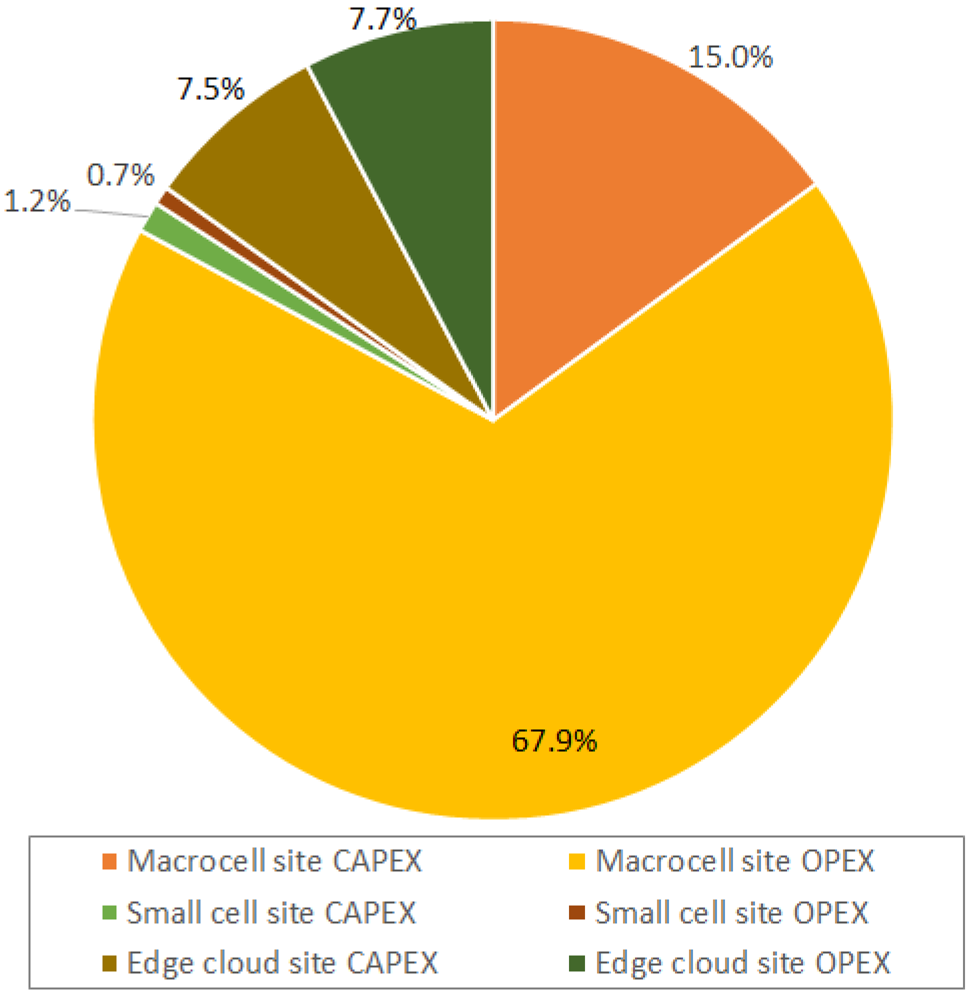

Figure 6. Baseline case (case 1): CAPEX and OPEX components of the RAN infrastructure (2020-2030), elasticity not employed.

Fig. 7 shows the number of GPP cores installed in total across all four edge cloud sites to serve the macrocell network in 2030 based on the wider area eMBB demand across the study area at this time. The figure shows how many of these GPP cores are utilised and not utilised at the two different times of day modelled. The two time intervals shown in Fig. 7 correspond to the peak demand times for the city centre (i.e., midweek 5pm to 6 pm) and the cruise ship terminal (i.e., Saturday 6am to 7am), respectively.

During the busy hour of the macrocell layer, the total number of unutilised GPP cores is only 34. However, when a large cruise ship arrives at the terminal during the off-peak time of the macrocell layer, i.e., Saturday morning, the total number of unutilised cores available from the macrocell layer increases to 1175. The number of GPP cores required to process the traffic generated by the cruise ship passengers via the SC network during this time is 113. Therefore, during the cruise ship terminal's peak time, there are more than enough unutilised cores available from the macrocell servers in the edge cloud sites to process the traffic generated from the SC network.

If the unutilised cores can be dynamically assigned to and used for the cruise ship terminal SC network via elasticity, then any dimensioning of additional processing (i.e., 113 cores) and subsequent edge cloud site costs for the SC network can be saved. In other words, without elasticity the edge cloud sites would be dimensioned for the sum of the peak requirements of both the macrocell network (2674 cores) and SC network (113 cores). With elasticity the overprovisioning of 113 cores for the SC network can be avoided.

Avoiding this overprovisioning of cores due to elasticity results in a saving of edge cloud costs of 16.4%. It is then possible to derive the total RAN cost savings by combining this saving of 16.4% with the result from Fig. 6 showing that edge cloud site cost components make up 15.2% of overall RAN costs. The total cost saving due to elasticity in the whole RAN infrastructure over the 2020-2030 period for the baseline case is 2.5%.

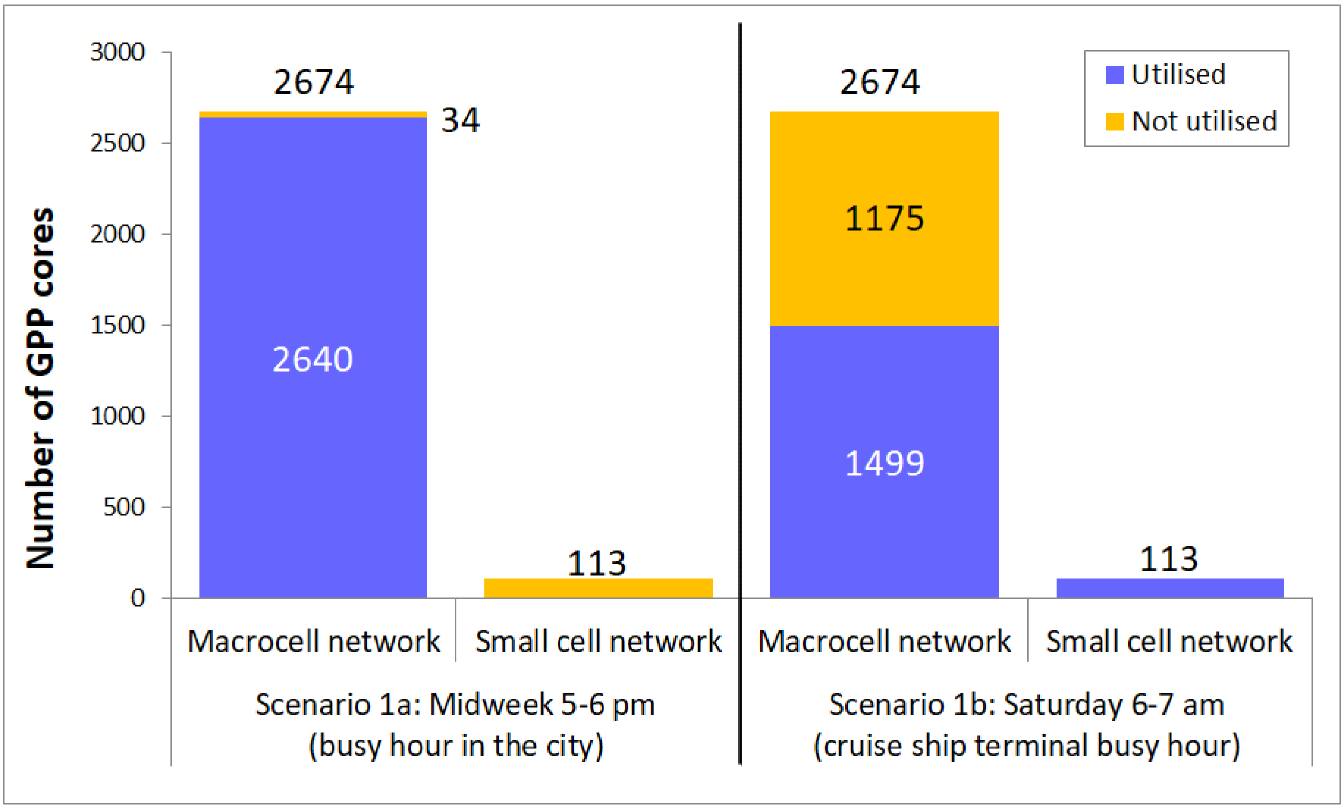

Figure 7. Baseline case (case 1): Amount of GPP cores used and not utilised in 2030. Modelling conducted for the peak demand time for Hamburg city (scenario 1a) and for the cruise ship terminal (scenario 1b).

5.2 Case 2: Effect of traffic

Fig. 8 shows the cost savings from elasticity under different demand scenarios. It is shown that the cost savings from elasticity vary from 2.5%, for the baseline case with medium traffic, up to 2.7%, for the low and high traffic demand scenarios. The cost savings of the RAN infrastructure for the overall area due to elasticity are, therefore, largely insensitive to the demand. This is because in the context of network costs for the wide area network, the incremental cost of the SC network is quite small and the elasticity impacts only the edge cloud site compound of the SC network costs. Focusing purely on the incremental cost of serving the demand hotspot at the cruise ship terminal, we see a diminishing returns effect with demand.

The cost savings on the additional cost of serving the demand hotspot at the cruise ship terminal decreases from 48% to 43% and further to 38% with low, medium and high demand, respectively. This is because the number of edge cloud sites is fixed regardless of demand level, although the volume of servers will increase with demand, but the number of antenna sites will increase with demand. Therefore, in the low demand scenario the edge cloud site costs, impacted by elasticity, make up a larger proportion of the TCO of the cruise ship terminal network compared with the high demand scenario.

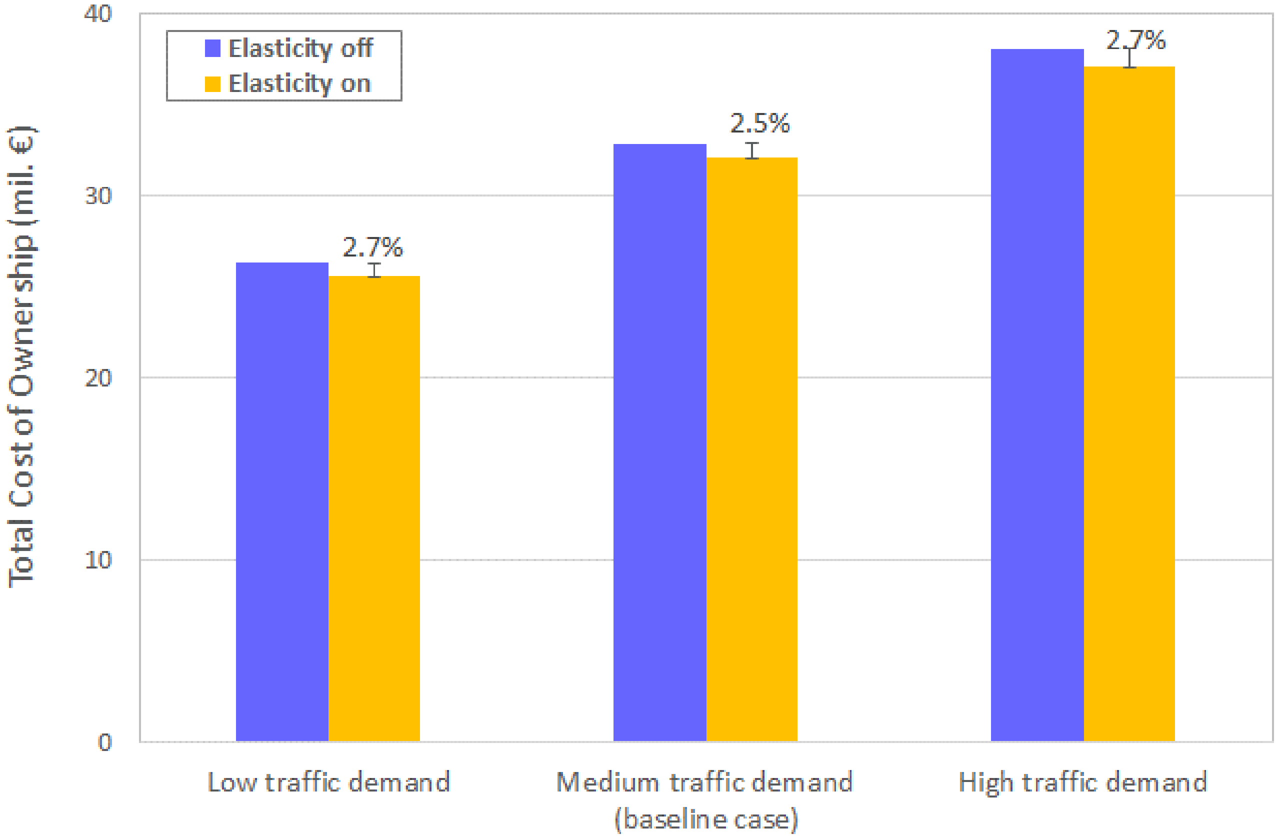

Fig. 8. Sensitivity analysis (case 2): Impact of traffic demand on cost savings achieved by elasticity.

5.3 Case 3: Effect of market share

So far, we have assumed that 3 MNOs serve the traffic generated by the cruise ship terminal and that the market share at the cruise ship terminal is equally distributed amongst these, i.e., each MNO is serving 33% of the total traffic. Fig. 9 shows the impact of this market share assumption on the deployment cost.

In this case, we consider the cost impact if the deployment is limited to two or a single MNO (i.e., 50% and 100% of the market share, respectively). In the absence of elasticity, both antenna site and edge cloud site costs increase for an increase in market share. Fig. 9 shows that the cost saving achieved via elasticity increases with increased market share or traffic on the network. The cost savings due to elasticity in the cases to serve 50% and 100% market share are 2.9% and 4.4%, respectively. This means that there is an increase of 0.4% and 1.9% percentage points, respectively, compared to the cost saving of serving 33% market share. This is because when there is more traffic for the cruise ship terminal hotspot network to carry, it will require more small cells and edge cloud site processing. The higher volume of equipment in the cruise ship terminal hotspot network means that the cost of this network becomes more significant compared to the wide area network. However, the processing required in the cruise terminal hotspot network is still relatively small compared with the wider study area macrocell network. In this way, even in the 100% market share scenario, the edge cloud site processing of the cruise terminal hotspot network can still be accommodated in the spare processing of the wider area network at the time of day when the cruise terminal is busy. Therefore, in the high demand scenario the cost savings relative to the overall network become more significant.

Note, we found that when elasticity is enabled in the Steinwerder edge cloud site alone, additional GPP cores were still required to serve the traffic from the cruise ship terminal when the market share is increased from 33% to 100%. However, when elasticity is enabled in all 4 edge cloud sites, the additional spare GPP cores available from the other edge cloud sites are enough to absorb the processing requirements of the SC network.

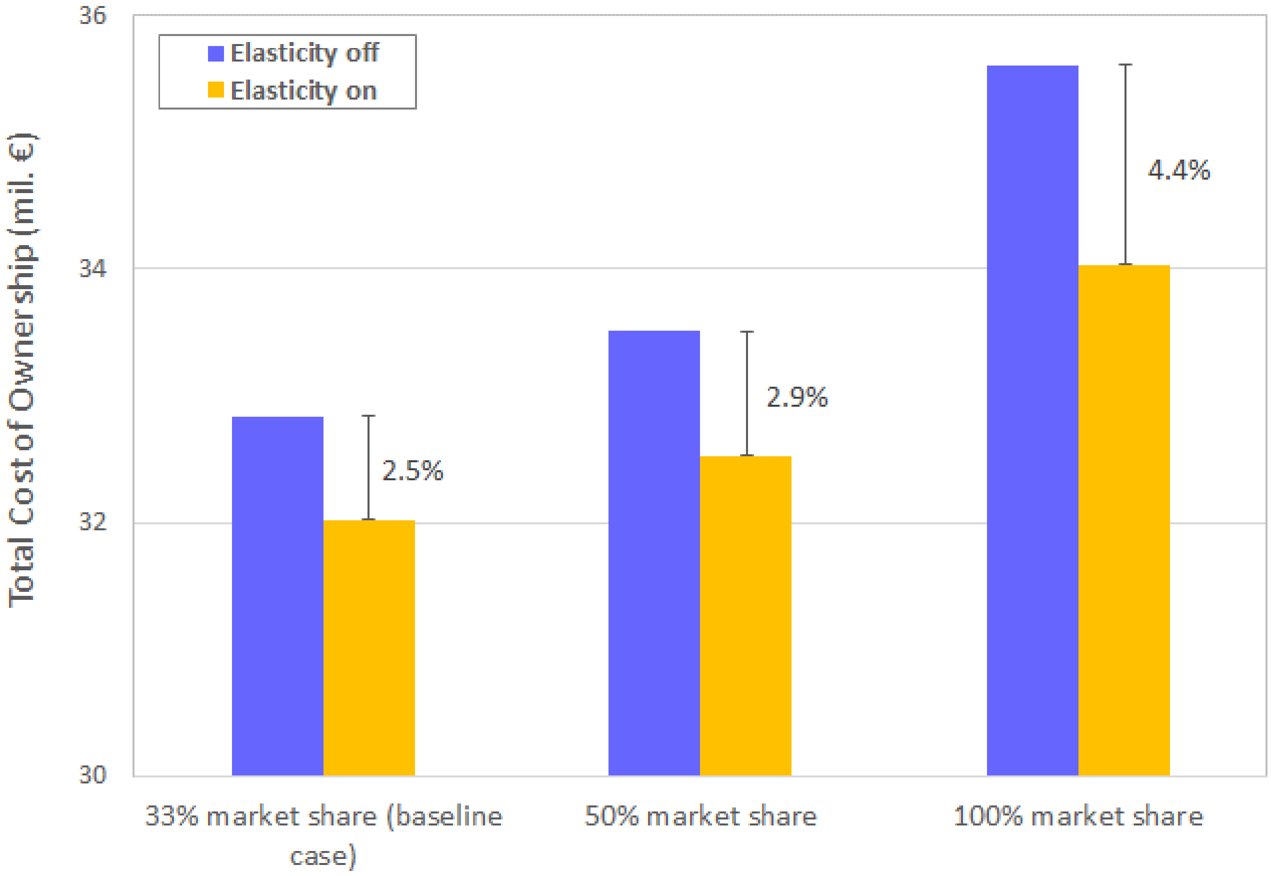

Fig. 9. Sensitivity analysis (case 3): Impact of market share on cost savings achieved by elasticity.

5.4 Overall analysis of the results

The benefits of elasticity in reducing network costs are limited to the processing and related edge cloud site costs. In networks that focus on macrocells, the total cost of ownership is dominated by the antenna site costs rather than the edge cloud site costs due to the high site rentals and civil works associated with large macrocells. In SC focused networks that target localised demand hotspots, the antenna sites are more straightforward to implement than macrocells (i.e., can be placed on existing street furniture) and are less costly. Therefore, in virtualised SC networks, the edge cloud site costs typically make up a larger proportion of the total cost of ownership compared with macrocell focused networks. Hence, SC focused networks stand to benefit more from elasticity.

There are a number of aspects that could be further studied when considering different scenarios and assumptions. Firstly, the study was performed for a specific scenario, which is the port of Hamburg. Other areas might require other network architectures with different network dimensioning, which will lead to different cost structures and, therefore, to different implications of the benefits of elasticity.

Secondly, the analysis was done for the specific 5G use case of eMBB. Other 5G use cases might require different network dimensioning. In the case of more extreme 5G use cases at the cruise ship terminal, this would increase the capacity and hence processing required to support the small cell network. Given the large volume of spare processing capacity on the wider macrocell network during the cruise ship terminal's busiest time, it is likely that the cost benefits of elasticity would increase further in this scenario.

Thirdly, in the TCO analysis some cost items, such as the cost of a 5G network orchestrator or the spectrum costs, were not taken into account. The cost of a 5G network orchestrator to provide elasticity was not included because this type of orchestrator was not commercially available when writing this article.

6. CONCLUSIONS

This article has shed light on the cost benefits of employing elasticity for a 5G network that provides a broadband service in a hotspot area. The results show that for the baseline scenario the RAN cost reduction achieved by the employment of elasticity is 2.5%. The sensitivity analysis performed for different scenarios shows that the TCO reduction ranges from 2.5 to 4.4%. Considering the edge cloud site-related cost component alone, a 16.4% saving from elasticity was found in the baseline scenario. In conclusion, elasticity is a technique to be considered when applying cost-efficient 5G network dimensioning. Further research is needed to understand the benefits of elasticity for other types of 5G use cases, such as augmented reality for tourism, hotspots for industrial areas or stadiums. Moreover, the effect of elasticity in suburban or rural areas, which can have different traffic profiles, should be studied.

References

1. Bega D, Gramaglia M, Bernardos Cano C, et al. Toward the network of the future: From enabling technologies to 5G concepts. *Trans. Emerging Tel Tech*. 2017; e3205. https://doi.org/10.1002/ett.3205
2. Oughton E, Frias Z. The cost, coverage and rollout implications of 5G infrastructure in Britain. *Telecommunications Policy*. 2018;42(8):636-652.
3. O-RAN. Open Radio Access Network Alliance. https://www.o-ran.org/. Retrieved April 14, 2021.
4. Rost P, Banchs A, Berberana I, et al. Mobile network architecture evolution toward 5G. *IEEE Communications Magazine.* 2016;54(5):84-91.
5. Banchs A, Gutierrez-Estevez D, Fuentes M, et al. A 5G Mobile Network Architecture to Support Vertical Industries. *IEEE Communications Magazine.* 2019;57(12):38-44.
6. Bruschi R, Bolla R, Davoli F, et al. Mobile Edge Vertical Computing over 5G Network Sliced Infrastructures: An Insight into Integration Approaches. *IEEE Communications Magazine.* 2019;57(7):78-84.
7. Gutierrez-Estevez D, Gramaglia M, De Domenico A, et al. Artificial Intelligence for Elastic Management and Orchestration of 5G Networks. *IEEE Wireless Communications*. 2019;26(5):134-141.
8. Paglierani P, Neokosmidis I, Rokkas T, et al. Techno-economic analysis of 5G immersive media services in cloud-enabled small cell networks: The neutral host business model. *Trans. Emerging Tel Tech*. 2019; e3746. https://doi.org/10.1002/ett.3746

9. Bassoli R, Granelli F, Arzo ST, Di Renzo M. Toward 5G cloud radio access network: An energy and latency perspective. *Trans Emerging Tel Tech*. 2019;e3669. https://doi.org/10.1002/ett.3669

10. Farhadi Zavleh A, Bakhshi H. Resource allocation in sparse code multiple access-based systems for cloud-radio access network in 5G networks. *Trans Emerging Tel Tech*. 2020;e4153. https://doi.org/10.1002/ett.4153

11. Bellavista P, Corradi A, Edmonds A, et al. Elastic Provisioning of Stateful Telco Services in Mobile Cloud Networking. *IEEE Transactions on Services Computing*. 2018. DOI: 10.1109/TSC.2018.2826003.

12. Zhang J, Ji Y, Zhang J, et al. Baseband Unit Cloud Interconnection Enabled by Flexible Grid Optical Networks with Software Defined Elasticity. *IEEE Communications Magazine*. 2015;53(9):90-98.

13. Wisely D, Wang N, Tafazolli R. Capacity and costs for 5G networks in dense urban areas. *IET Communications*. 2018;12(19):2502_2510.

14. Oughton E, Katsaros K, Entezami F, et al. An Open-Source Techno-Economic Assessment Framework for 5G Deployment. *IEEE Access.* 2019;7:155930-155940.

15. Rendon Schneir J, Ajibulu A, Konstantinou K, et al. A Business Case for 5G mobile broadband in a dense urban area. *Telecommunications Policy.* 2019;43(7):101813.

16. 5G-MoNArch. Deliverable D2.3, Final overall architecture. 2019. https://5g-monarch.eu/deliverables/.

17. ITU. Setting the Scene for 5G: Opportunities & Challenges. 2018. https://www.itu.int/pub/D-PREF-BB.5G_01-2018

18. Foukas X, Patounas G, Elmokashfi A, et al. Network Slicing in 5G: Survey and Challenges. *IEEE Communications Magazine*. 2017;55(5):94-100.

19. Marinova S, Rakovic V, Denkovski D, et al. End-to-End Network Slicing for Flash Crowds. *IEEE Communications Magazine.* 2020;58(4):31-37.

20. Taleb T, Mada B, Corici M, et al. PERMIT: Network Slicing for Personalized 5G Mobile Telecommunication. *IEEE Communications Magazine*. 2017;55(5):88-93.

21. 5G-MoNArch. Deliverable D6.3, Final report on architectural verification and validation. 2019. https://5g-monarch.eu/deliverables/.

22. Ayala-Romero J, Garcia-Saavedra A, Gramaglia M, et al. A Deep Learning Approach Tailoring Computing and Radio Resources in Virtualized RANs. In: The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19), Association for Computing Machinery, New York, NY, USA, October 2019, Article 30, 1–16. DOI:https://doi.org/10.1145/3300061.3345431

23. Zhu G, Liu D, Du Y, et al. Toward an Intelligent Edge: Wireless Communication Meets Machine Learning. *IEEE Communications Magazine*. 2020;58(1):19-25.

24. Cisco. Cisco Visual Networking Index (VNI): Forecast and Trends, 2016-2021. 2016. https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Germany_2021_Forecast_Highlights.pdf

25. Real Wireless. Techniques for increasing the capacity of wireless broadband networks: UK, 2012-2030. 2012. http://static.ofcom.org.uk/static/uhf/real-wireless-report.pdf

26. 5G-NORMA. Deliverable D2.3, Evaluation architecture design and socio-economic analysis -final report. December 2017. http://www.it.uc3m.es/wnl/5gnorma/pdf/5g_norma_d2-3.pdf