D1.3 "Industrial application of MAINS solutions and their impact on future broadband demand"

Status and Version:		
Date of issue:	06.02.2013	
Distribution:	Public	
Author(s):	Name	Partner
Fernández-Palacios	Juan	TID
Gonzalez-Rodriguez	Beatriz	TID
Basham	Mark	INT
Kiss-Kalló	Csaba	INT
Zervas	Georgios	UoEssex

Table of Contents

1. Introduction

Purpose and Scope

One of the main MAINS objectives is the development of new metro network architectures to be used by network operators in order to new services based on network resident IT resources and optimize the network resources consumption. Therefore, final projects results (architectures, SW developments, etc) are expected to be implemented in real ISPs' networks. According to it, this report aims to define what would be the steps to be taken in order to evolve from the final MAINS results towards a Carrier Class system:

- Identification of application scenarios for MAINS concepts.
- Identification of the functionalities that are missing and should be included in commercial systems.
- Performance studies to evaluate the MAINS solution deployment in a real scenario.

Reference Material

Reference Documents

[1] Cisco White Paper, "Entering the Zettabyte Era", June 2011.

[2] T. Lohmar, M. Elisove, "Evaluation of the File Repair Operations for Multicast/Broadcast Download Deliveries" April 2005.

[3] Cs. Kiss Kalló, M. Basham, J. Dunne, J. P. Fernandez-Palacios, "Cost Reduction of 80% in Next-Generation Virtual PC Service Economics Using an OPST Sub-Wavelength Metro Network", In Proceedings of NOC 2011.

[4] J.E. Gabeiras, V. López, J. Aracil, J.P. Fernández Palacios, C. García Argos, Ó. González de Dios, F.J. Jiménez Chico and J.A. Hernández: Is Multi-layer Networking Feasible?, in Elsevier Journal on Optical Switching and Networking, April 2009, Vol. 6, Issue 2, Pages 129-140.

[5] Intune Networks, "Optical Packet Switch and Transport. A Technical Introduction", 2009. http://www.intunenetworks.com/home/shapeup/core_innovation/opst_technical_introduction/

[6] N. Metropolis, S. Ulam, "The Monte Carlo Method", September 1949.

[7] Kiss Kalló, Raymond Carley, Conor O'Malley, Victor Lopez, J. P. Fernández-Palacios, "Economic Modelling of Uncertain Next-Generation Network Evolution", submitted to Optical Network Design and Modeling, ONDM 2012, Colchester, UK.

[8] Y. Lee et. al., "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", IETF Draft, Feb. 2011.

[9] Cisco White Paper, "Cisco Visual Networking Index: Forecast and Methodology, 2010–2015" June 2011.

[10] P. Pavon-Marino et. Al., "On the Myths of Optical Burst Switching", IEEE Transactions on Communications, vol. 59, no. 9, pp. 2574-2584, September 2011.

[11] Kiss Kalló, Victor Lopez, John Dunne, Mark Basham, J.P. Fernández-Palacios, “Benefits of Optical Packet Switching for Router By-pass in Metro Networks“,Future Networks and Mobile Summit, Berlin, Germany, July 2012.

[12] A. Rosas et al., Evolutionary projects in Content Delivery [online], Future Internet Assembly, 2010.

[13] M. Zink et. al, “Characteristics of YouTube network traffic at a campus network - Measurements, models, and implications”, in Computer Networks 53(4), 501-514, 2009.

[14] M. Cha et. al, “Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems”, in ACM/IEEE Transactions on Networking, 5(17), 2009.

[15] J. P. Fernández-Palacios, L. Perez, J. Rodriguez, J. Dunne, M. Basham, IP Offloading over Multi-granular Photonic Switching Technologies In proc. of ECOC 2010.

[16] C.Kiss-Kalo, J.Dunne, et. al.,” Economic Modelling of Uncertain Next-Generation Network Evolution”, ONDM 2012, accepted.

[17] N. Metropolis, S. Ulam, “The Monte Carlo Method”, September 1949

[18] Network-on-and-off-Chip Architecture on Demand for Flexible Optical Intra-Datacenter Networks. B.R. Rofoee (1), G. Zervas (1), Y. Yan (1), N. Amaya(1),Y. Qin(1), D. Simeonidou(1).

[19] A. Vahdat et al., OFC’11, OTuH2 (2011).

[20] C. F. Lam et al., Coms. Mag., July (2010).

[21] Data Center Networking Enterasys (2011)

[22] G. S. Zervas et al, J OPEX, (19) 26, (2011)

[23] N. Amaya et al., Proc. ICTON’11, (2011)

[24] N. Amaya et al., Proc. ECOC’11, (2011

[25] Kiss Kalló, J, Shields, Raymond Carley, Víctor López, J.P. Fernández-Palacios, “Cost-Effective Sub-wavelength Solution for Data Centre Location in Scaled Next-Generation Networks“, submitted to Optical Network Design and Modeling, ONDM 2012, Colchester, UK.

Acronyms

AN	Access Node
CN	Concentration Node
CO	Central Office

CAPEX	Capital Expenditures
CPE	Customer Premises Equipment
CS	Concentration Switch
DWDM	Dense Wavelength Division Multiplexing
E2E	End to End
EN	Edge Node
E-NNI	External Network-to-Network Interface
FTTH	Fiber To The Home
GbE	Gigabit Ethernet
GMPLS	Generalized Multi-Protocol Label Switching
L2VPN	Level 2 Virtual Private Network
NCS	Network Centric Services
OAM	Organization Administration and Management
OBS	Optical Burst Switching
OBST	Optical Burst Switching Technology
OLT	Optical Line Termination
OPEX	Operational Expenditures
OPST	Optical Packet Switching Technology
PCE	Path Computation Element
QoE	Quality of Experience
QoR	Quality of Resilience
QoS	Quality of Service
SC	Sub-wavelength Concentrator
SOAP	Simple Object Access Protocol
TE	Traffic Engineering
TMF	Telemanagement Forum
TNA	Transport Network Address
VM	Virtual Machine
VoD	Video on Demand
WADL	Web Application Description Language
WSON	Wavelength Switched Optical Networks
XML	eXtensible Markup Language
XSLT	eXtensible Style sheet Language Transformations

Executive Summary

According to the results of both technoeconomic and performance analysis carried out during the project, MAINS industrial application scenarios can be summarized as follows:

- **High capacity multiservice metro networks**. Technoeconomic analysis over multiservice metro networks in Madrid metro area [11] demonstrates a 42% capital cost reduction by 2015 in favour of the proposed MAINS architecture, when compared against currently used, typical all-IP architectures.
- **Cost effective adaptation to uncertain traffic evolution**. The uncertainties and the implied level of risk associated with next-generation network architectures were modelled in [7] using Monte Carlo simulation, aimed at understanding network economics evolution. The core result of this analysis is that a subwavelength optical packet forwarding technology can de-risk network investments by 500% when compared to a next-generation IPoDWDM solution.
- **CDN (Content Delivery networks) optimization.** As a practical way to measure MAINS architectural benefits for content delivery networks a comparison was made between common H-VPLS and OPST Metro deployments for streaming (Unicast) video services. Cost and power figures are compared with the H-VPLS alternative, highlighting the savings as well as additional potentials (in terms of scalability, restoration times, etc.) brought by the use of MAINS technologies.
- **Elastic VPNs.** MAINS architectures would also enable new Elastic VPNs services which are not available yet. As demonstrated in the industrial performance analysis (section4). Subwavelength switching technologies can provide guaranteed bandwith on demand from Mbps to tens of Gbps. Some potential use cases for elastic VPNS are the interconnection of datacenters with BoD for data intensive transmissions or the modification of guaranteed bandwidth between datacenters and access nodes according to dynamic traffic patterns.
- **Flexible Data Centre deployments**. The impact of data centre location in the network on the network cost is analyzed in MAINS. Data centre location, routing strategies, service mix, traffic growth and subscriber service take-up is modelled to obtain a broad view about the cost sensitivities in these networks. MAINS analysis [25] demonstrates that resilience to data centre location and changing traffic patterns enable the sub-wavelength packet optical solution to achieve cost savings when compared to the IPoDWDM approach.
- **Intra-Datacenter networking**. MAINS is proposing a novel network on-and-off chip approach for highly efficient and transparent intra-datacenter communications based on sub-wavelength switching [18]. The implemented FPGA-based network on-chip line card enables hitless adaptation between Ethernet and TSON, which is supported by a flexible network off-chip of AoD, demonstrating end-to-end high performance results
- **Cost effective Virtual PC services**. MAINS proposed architecture for nomadic virtual PC services would be a 20% of the cost of the current network architecture approach, revealing a new and counter-intuitive network design [3].

The suitability of Intune´s OPST platform as the Telco infrastructure for the delivery of above mentioned advanced services has been demonstrated in a complex test bed installed Telefonica labs in Madrid resembling some realistic deployment scenarios.

2. Application scenarios for MAINS concepts

2.1. Scalable multiservice metro networks

Subwavelength switching technologies proposed in MAINS (i.e TSON and OPST) have been designed to support services with different QoS constraints (e.g best effort, real time, streaming…) over the same optical infrastructure. Such multiservice capabilities have been demonstrated in different experimental performance analysis carried out in University of Essex and Telefonica Labs (see section 3).

Such capabilities enable MAINS architectures to be deployed in multiservice metro networks combining voice, video and Internet applications. In fact, as demonstrated in the MAINS field trial in Cyprus, MAINS architectures can be operated and controlled like current Carrier Ethernet or IP/MPLS solutions so that the evolution from current metro solutions towards MAINS architectures could be done in a smooth and transparent way for network operators.

2.1.1. Cost effectiveness for increasing traffic demands

The rationale behind the evolution from current Carrier Ethernet and IP/MPLS networks towards MAINS architectures relies on the increased scalability and cost effectiveness of subwavelength switching solutions for increasing traffic demands generated by intensive bandwidth consuming applications such Internet live streaming applications (HD, 3D, Ultra HD…), HD video conference, cloud services (e.g Virtual PC, IT outsourcing,).

According to recent traffic forecasts [1], by 2015 the global data consumption will cross the Zettabyte (1021 bytes) threshold. The biggest chunk of this large amount of data is some form of video content. The most popular video service over the Internet is YouTube. It is estimated that 48 hours of YouTube video is uploaded per minute and that the number of views per day is 3 billion.

The scalability and cost effectiveness of MAINS proposed architectures were demonstrated in the technoeconomic analysis reported in [1] and published in [11]. The main contribution of this analysis was that it demonstrated a 42% capital cost reduction in favour of the proposed MAINS architecture, when compared against currently used, typical all-IP architectures. This is achieved by performing packet transport, aggregation, switching and grooming in the optical layer within an IP off-loading architecture. Such results show that multi-granular optical technologies are a strong candidate for solving the bottleneck problem caused by video streaming, cloud services and mobility in metropolitan area networks.

2.1.2. Economic modeling of uncertain network evolution

Another advantage of MAINS architecture relies in its capability for flexible network resources allocation according to dynamic traffic patterns. Such property would enable network operators to significantly de-risk network investments in multiservice aggregation networks.

2.1.2.1. Traffic uncertainty in next generation networks

There is general consensus in the telecommunications industry that traffic patterns in service provider networks are dynamically changing and are hard to predict. Such uncertain behaviour is due to a series of factors, unpredictable service evolution, changing user habits and user mobility being among the most important ones.

Firstly, new Internet applications are being released every day and they introduce unpredictable bandwidth demand and traffic profiles to the network. Second, the number of devices connected to IP networks is expected to be twice as high as the global population in 2015 [1]. This will likely lead to new user habits as they learn to use networked devices in new areas of their everyday lives. Third, the proliferation of high quality of experience (QoE) mobile data services running on comfortably usable smartphones is making its impact on the traffic profiles. Service providers cannot be certain anymore about the geographical distribution of demand. Consumer traffic patterns can change instantaneously based on social or business gatherings' location.

Fourth, on the server side, interest for a specific content can be sparked by a single news article or a post on a social networking webpage, creating a phenomenon known as the Slashdot effect [2]. Fifth, as cloud services gain more traction, highdemand services like the Virtual Personal Computer [3] and gaming [1] will significantly contribute to the bandwidth needed for providing excellent QoE to cloud consumers.

The list can continue, but the widely accepted view in the industry is that

- traffic is going to increase in the future, but it is impossible to know by how much; and

- traffic will change very dynamically, but no-one can tell on what pattern.

This uncertainty represents an important risk in the network operators' business and addressing that risk implies a significant cost. One big chunk of this cost is stemming from some form of overprovisioning and/or inefficient use of resources (e.g., provisioning for peak customer bandwidth, stranded bandwidth on static optical paths, building out capacity for changing geographical demand distribution, scaling the network, etc). Another important cost source is operations. Traffic forecasting, frequent network re-optimizations, optical equipment installation, separate control and management for the IP and optical layers are all contributing to high staff and services costs.

Over the past decade network operators and equipment vendors have optimized their solutions and many optimizations targeted the removal of layers and protocols between IP/Layer3 and Optical/Layer1. Thus, IP over optical paths (e.g. IPo(D)WDM, IPoROADM, IPoL1) was proposed by many as a future-proof solution for unpredictably growing, dynamic next-generation networks [4]. In this case study the economic potential of a sub-wavelength optical switching technology called OPST

(Optical Packet Switch and Transport) [5] is analysed in the context of next-generation services and is compared against an equivalent IPoDWDM solution. OPST is an innovative optical switching technology based on ultra-fast tuneable lasers designed with dynamic traffic patterns, efficient resource utilization and growing networks in mind. The data path is managed through a single, unified OPST control and management plane for Layer 1 and Layer 2, which simplifies network operation.

Thus, by design, OPST is capable of adapting and scaling in real time to the fast-changing demand that operators need to face in their networks. In the remainder of this paper, Section II describes the approach and assumptions used for the presented technoeconomic network models. Section III provides an analysis of the Monte Carlo modelling output, while Section IV briefly discusses the importance of the results from a broader industry perspective. Section V concludes the case study.

2.1.2.2. Network modeling for uncertainty analysis

Two next-generation network architectures are modelled and compared to analyse their characteristics at scale and the impact on the capital expenditure (CapEx) of uncertainty regarding service mix, service take-up, traffic growth, traffic pattern and data centre (DC) location. Monte Carlo simulation [6] is used to address the huge configuration space implied by the combination of all these input parameters. In the following subsections all of the modelling assumptions are explained in detail.

A. Monte Carlo Modelling

Monte Carlo simulation is used when there is a large number of possible inputs into a model, which makes it unfeasible to calculate every possible outcome. Instead, a large number of random samples are taken, the idea being that the samples will be representative of all the possible outcomes [6].

For each service modelled over the network architecture considered in this case study, the traffic per subscriber, traffic and service mix (i.e., the destination of traffic at service granularity) and the subscriber take-up are allowed to take on random values within a relevant range. Therefore for each trial performed, the traffic for each service could take on a random value. It is impossible to forecast how the traffic for each service will increase in the future, so the Monte Carlo approach allows a valuable insight into the vast space of possibilities.

B. Generic Model Inputs

Each of the modelled architectures focus on dense metropolitan areas that contain a user base of one million distributed around 60 different sites in the metro region.

The model calculates all the equipment and their costs needed between the access head-end node - assumed to be one or more Optical Line Terminals (OLTs) - and the Internet exchange point (see Figure 2.1.1 1). A set of service and traffic baselines from Telefónica data and recent Cisco forecasts [9] are used as guides around which large variations are then used as inputs to the model.

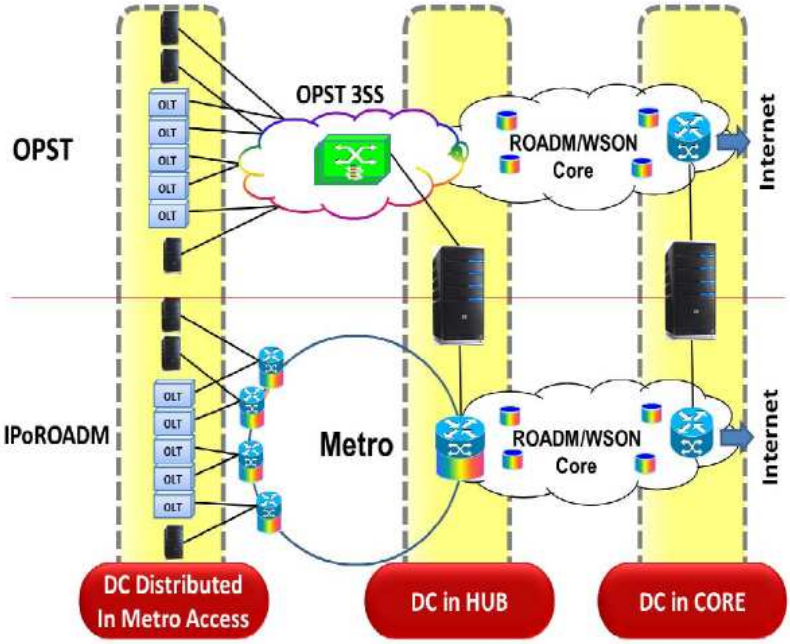


Figure 2.1.2.2-1: OPST (top) and IPoWDM (bottom) architectures and data centre locations. Currently most of the content is sourced from data centres located in the Internet or the core. Interest for moving the data centres in the hub or even in the metro access is increasing, due to the performance benefits

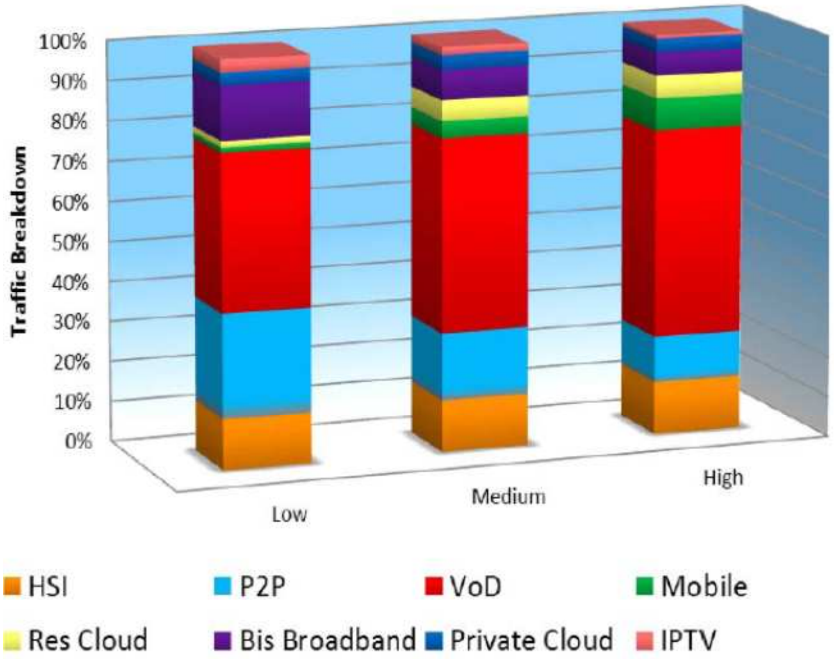


Figure 2.1.2.2-2: Average traffic breakdown per service for the three traffic bands based on Cisco forecasts and a Telefonica traffic data forecast.

The total traffic for the metro region is calculated as a function of the traffic per subscriber and the number of subscribers for 8 different services. Each service is allocated bandwidth per subscriber. Although the allocated bandwidth is an average value, the OLT uplinks are assumed to be filled at 25% only to allow for peaks. The models calculate the service and transport layer equipment required in the network to deliver the services when the data centre is placed in the core, at the metro hub and/or out in the metro (see Figure 2.1.1 1).

When the resources are derived, the models then calculate, using typical unit pricing, the capital cost of the network. Approximately 25 different pricing values were used in the model, but the key comparative pricing is the average cost of a 10GigE OPST port as 1.2 times the price of an average 10GigE IP Router port cost. This relative pricing variable was validated against current industry pricing. Savings are achieved because of the inefficient use of Router ports per solution, e.g. up to 50% of Router ports in some scenarios are interconnecting Routers together rather than delivering services.

One million trials are performed (i.e. one million different combinations of traffic per service, service take-up, traffic pattern, traffic turn-around location and DC locations were taken) so that a significant range of all possible traffic makeups was examined.

Three different traffic bands are used for the models, representing short, medium and long-term evolutions of the communication environment:

- Low Traffic ~0.2-2Tbps
- Medium Traffic ~2-5Tbps
- High Traffic ~5-25Tbps

Average traffic and service take-up assumptions are visualised in Figure 2.1.2.1 2: and Figure 2.1.2.1 3 and further discussed later in this section.

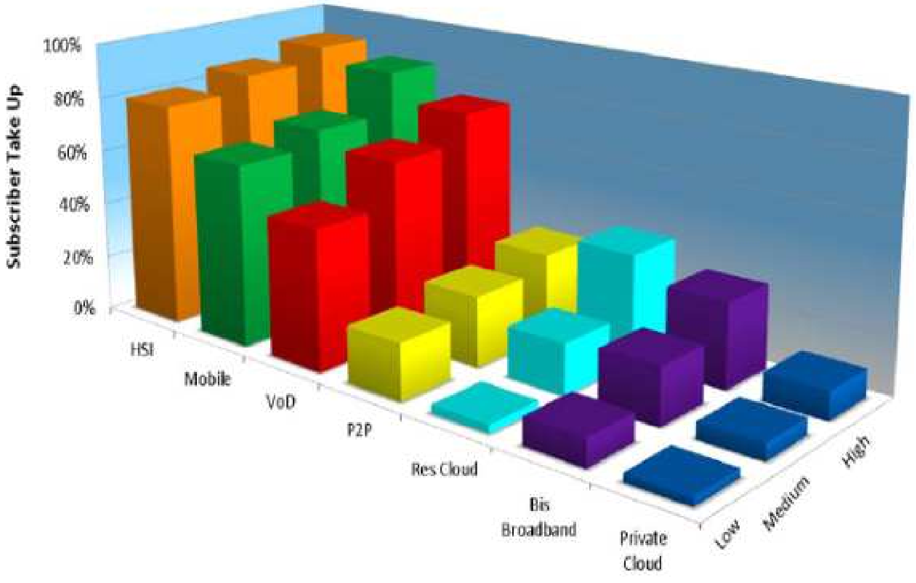


Figure 2.1.2.2-3: Subscriber take-up per service for the three traffic bands. It shows average service popularity based on Cisco forecasts and a Telefonica traffic data forecast.

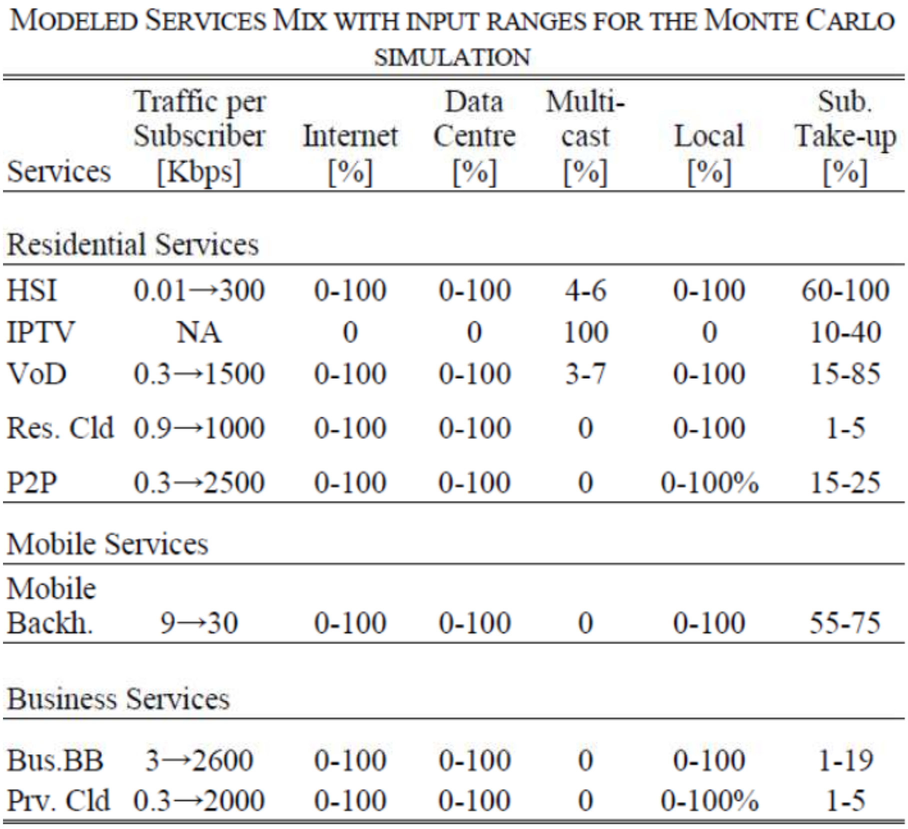
Once the simulations are complete, the results are analysed to understand the economic implications of the network architectures going into the uncertain future. The two network architectures modelled and compared are presented in the next subsections. One of the architectures is based on IPoDWDM technology and is implemented with ROADMs at the transport layers and IP routers at the service layer. Two variants of this IPoROADM architecture are analysed, one in which the local traffic is turned around at the metro head-end, while in the second one the local traffic was turned around at the core routers. On the other hand, the second, OPSTbased, architecture allows for a full optical mesh in the metro regional network, while in the core it uses the same IP and ROADM equipment as the comparative IPoDWDM models [7]. OPST merges the switching and transport functions into a single layer and thus provides a sub-wavelength optical packet forwarding platform that aggregates, grooms, switches and transports packets in a uniquely efficient manner [5]. IPoDWDM is a reasonable solution for carriers to reduce investment, while absorbing traffic increment [4].

C. IPoDWDM Architecture

In the IPoROADM architecture (Figure 2.1.2.1 1, bottom) traffic is picked up by IP access routers and is aggregated by hub routers located at the metro head-end. A dynamic Wavelength Switched Optical Network (WSON) [8] built with ROADMs is used in the backbone to transport the traffic to core routers, which in turn are connected to Internet exchange points. Routers with 2Tbps nonblocking switching capacity are used in the hub and core.

When the capacity of a router is exceeded, additional routers are added to the network through stacking. Router stacks can be placed in the hub or in the core. All of the relevant combinations of router stack and DC placement are analysed.

TABLE I MODELED SERVICES MIX WITH INPUT RANGES FOR THE MONTE CARLO SIMULATION



MODELED SERVICES MIX WITH INPUT RANGES FOR THE MONTE CARLO SIMULATION

Services	Traffic per Subscriber [Kbps]	Internet [%]	Data Centre [%]	Multi-cast [%]	Local [%]	Sub. Take-up [%]
Residential Services						
HSI	0.01→300	0-100	0-100	4-6	0-100	60-100
IPTV	NA	0	0	100	0	10-40
VoD	0.3→1500	0-100	0-100	3-7	0-100	15-85
Res. Cld	0.9→1000	0-100	0-100	0	0-100	1-5
P2P	0.3→2500	0-100	0-100	0	0-100%	15-25
Mobile Services						
Mobile Backh.	9→30	0-100	0-100	0	0-100	55-75
Business Services						
Bus.BB	3→2600	0-100	0-100	0	0-100	1-19
Prv. Cld	0.3→2000	0-100	0-100	0	0-100%	1-5

D. OPST Architecture

In the comparative OPST architecture (Figure 2.1.2.1 1, top) a three-stage optical packet forwarding engine [5], called 3SS, carries out the aggregation, grooming, switching and transport functions. The 3SS substitutes all of the routers and ROADMs used in the IPoROADM solution for the metro space. From a service provisioning perspective, the 3SS appears as a single L2 switch, while it comprises multiple optical platforms interconnected on fibre rings from an operations and maintenance perspective.

E. Data Centre Location

The data centre location can have a significant impact on the cost and performance of a network. In this study different configurations are analysed, where the data centre is placed in the core, in the metro hub or embedded/distributed in the metro access (Figure 2.1.2.1 1).

Currently most of the data centre traffic is coming from the Internet. In order to improve performance and to include the network owners – who play a key role in delivering DC content to the end-users – into the business model of content delivery, DCs are expected to migrate closer to the user. The easiest way to do this is to build huge, centralized DCs in the network core and provide access to them through the core routers. In both cases – with the DC in the Internet and in the core – the network equipment requirements are approximately the same. The difference is that the core router ports that handle the traffic face the internet exchange routers (Internet) or the DCs located in the core. Therefore, these two cases are addressed together in the models presented in this case study.

The DCs can be also placed in the hub to achieve higher performance. The effect of this is that the hub routers carry out more of the grooming and switching functionality. Therefore, this solution results in a more expensive hub and a cheaper core, as more workload is

distributed from the core toward the hub. Finally, the DC can be distributed in many small units in the metro access. The benefit of this – beyond the desired improvement on the user experience – is that central office space can be more efficiently utilized, as servers and storage can be placed wherever there is space in a point of presence, rather than in a few dedicated locations only (i.e., in centralized DCs). The downside of such an architecture is that interconnecting the distributed DC equipment in a cost-effective way, while latency, jitter and ultimately user experience is uncompromised, becomes a challenging task. This is so, because, on one hand, if distributed DC modules are interconnected through a hub router then there is no performance advantage achieved, compared to the case where the DC is placed in the hub. This would actually make the performance of the distributed DC solution worse, rather than improving it as intuitively expected. On the other hand, if the DCs are interconnected through direct optical connections then the cost increases due to the higher port requirements. All of the above DC location considerations are addressed in the models.

F. Service Mix

Eight services types of residential, mobile and business type are modelled. These services are High Speed Internet (HSI, e.g. email, browsing), Internet television (IPTV, e.g. live television), Video on Demand (VoD, e.g, Content Delivery Network (CDN) or Youtube video streaming), Residential Cloud (e.g. Virtual PC, machine-to-machine and other next generation services), Peer-to- Peer (P2P, e.g. file sharing), Mobile Backhaul (e.g. LTE, UMTS, GPRS, GSM), Business Broadband (employee browsing) and Private Cloud (e.g. corporate cloud, includes private lines) (see Table I). For each service an average bandwidth per subscriber was assumed. The traffic is categorized by destination, meaning that traffic can arrive to the metro access sites from the Internet, from DCs located in different places of the network, local metropolitan area traffic and multicast traffic. Multicast

traffic is assumed to be initiated from the DCs and form a separate category for the particular way it is handled in the network. Average traffic breakdown per service and per traffic band is shown in Figure 2.1.2.1 2. Each service is also associated a percentage of subscribers. This percentage is a representation of the popularity of each individual service (see Figure 2.1.2.1 3). The random values are generated according to a uniform distribution within each individual range. Service and traffic assumptions are based on internal Telefónica data and aligned with recent traffic forecasts [9]. However, the core objective of this study is to understand the impact of inaccuracies of such forecasts. This is the reason why in Table I a very large range of possible values for each service and parameter is used and this is how the Monte Carlo simulation approach is applied to this study.

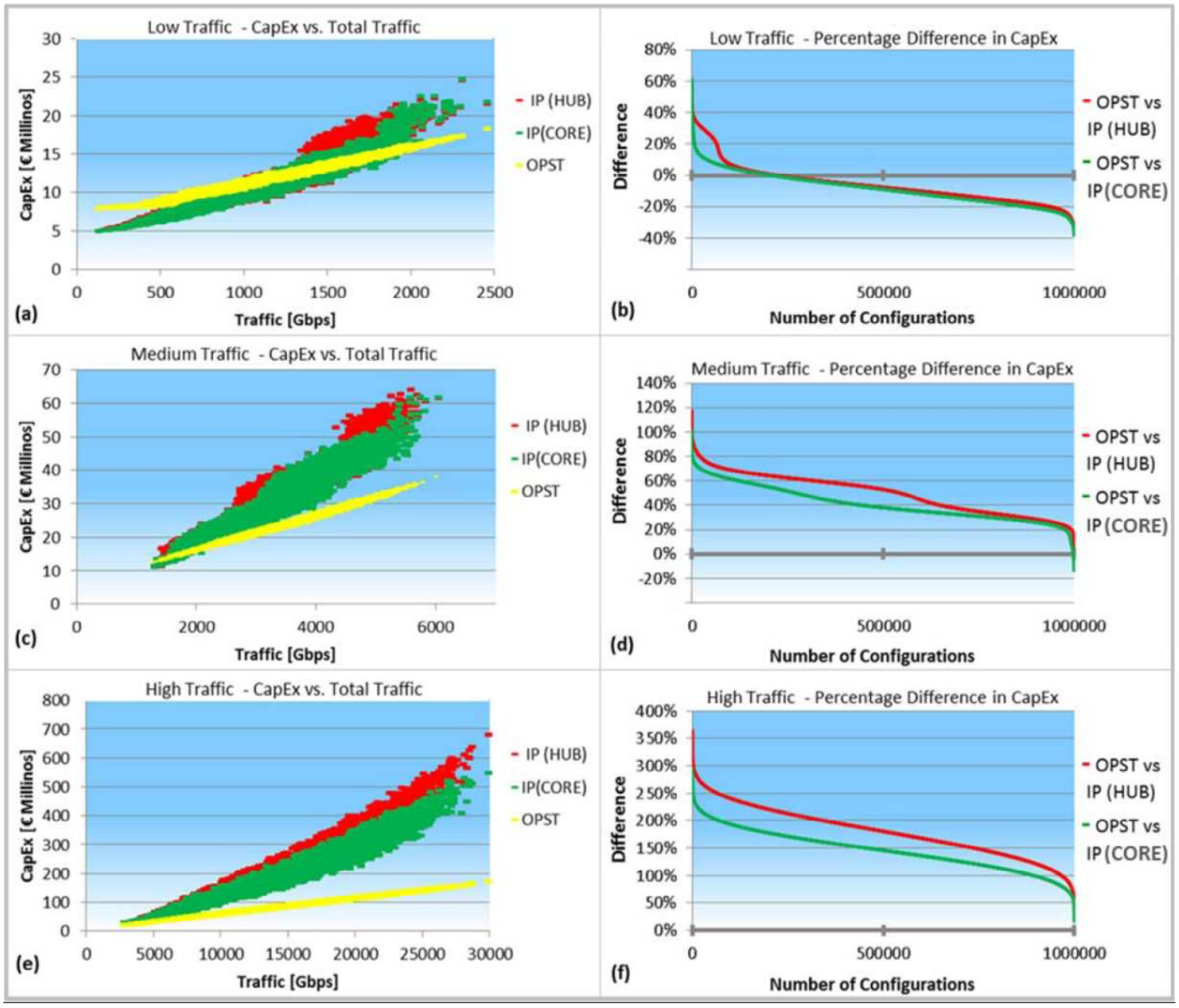


Figure 2.1.2.2-4: CapEx Uncertainty. IPoDWDM: exponential capacity growth at exponential CapEx growth, high uncertainty; OPST: exponential capacity growth at linear CapEx growth, very low uncertainty

The CapEx results for these Monte Carlo simulations are shown for the Low, Medium and High traffic scenarios in the top, middle and bottom graph pairs, respectively, of Figure 2.1.2.1 4. Left side graphs (i.e., Figure 2.1.2.1 4.a,c,e) show the actual configuration CapEx while on the right hand side (i.e., Figure 2.1.2.1 4.b,d,f) the CapEx percentage difference is shown relative to the OPST CapEx. The Low Traffic scenario represents the current and short-term situation, as it assumes a few hundred Gbps traffic for the metro region. This scenario can be comfortably addressed with solutions based on existing single-chassis routers placed in the metro hub or the network core. Smaller IP access routers can aggregate the traffic from multiple clients connected to the same client site and pass traffic on to the metro hub in more consolidated 10G connections. The average utilization on these metro connections typically does not exceed 25%, Therefore, traffic aggregation is performed to make more efficient use of the network resources. As the traffic grows, the CapEx evolution of the IPoDWDM and OPST solutions evens out so that at 1.3Tbps the CapEx gap disappears.

As the traffic grows into the Medium Traffic band the CapEx benefits enabled by OPST become evident. For in excess of 99.8% of the configurations in this band OPST provides significant CapEx savings. This Medium band can be seen as a mid-term traffic scenario that next generation networks are currently being designed for by carriers.

Finally, the High Traffic scenario gives a longer term view over the trends of network CapEx evolution. Figure 2.1.2.1 4.e shows a sustained linear growth for the OPST solution CapEx, while the IPoDWDM CapExes take an exponentially-looking turn upwards. The CapEx gap between the compared solutions is also quite considerable. The average percentage gap between the OPST and IPoDWDM solution CapEx is around 150%, as shown in Figure 2.1.2.1 4.f. Also, the OPST solution CapEx in this latter case is lower for 100% of the configurations.

The main observation of this Monte Carlo analysis, however, is the level of CapEx uncertainty pertaining to the analysed solutions. The effects of uncertainty in traffic, service and data centre evolution as well as in the chosen IP routing strategy can be seen in the dispersion of the solution CapExes, highlighted in Figure 2.1.2.1 5. This CapEx dispersion represents the risks stemming from forecast errors and unpredictable turns in the evolution of network strategy and services. The flexibility and realtime re-configurability of the OPST solution defuses network strategy decisions. Operators do not need to take a bet with OPST and live with it. Instead, they can save the cost of expensive forecasting and decision making; they can start off with a network architecture that meets the short-term requirements and adaptively change it as demand requires. This is possible because the OPST technology virtualises the optical layer. In other words, the optical resources of an OPST network can be reconfigured as easily as the electrical connections of an IP network. This deep and sophisticated optical configurability brings additional dimensions to network flexibility and enables a true virtualisation of the network resources. As shown in Figure 2.1.2.1 5, this OPST flexibility translates into about 6 times narrower CapEx dispersion than that of the IPoDWDM solutions, which in turn means 500% less CapEx uncertainty in favour of the OPST solution.

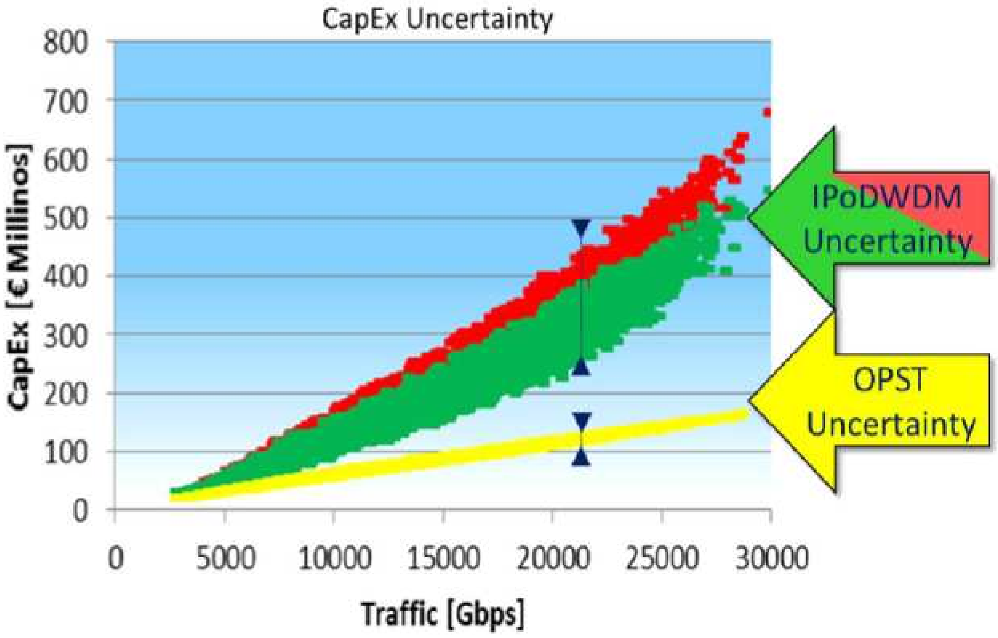


Figure 2.1.2.2-5: CapEx Uncertainty. IPoDWDM: exponential capacity growth at exponential CapEx growth, high uncertainty; OPST: exponential capacity growth at linear CapEx growth, very low uncertainty.

Carriers around the globe have been concerned for the past few years about the high cost of scaling their networks. It has been shown that the way routers and their switch fabrics can be connected together to support the rapidly growing traffic demand is not financially sustainable on the long term. The cost of network equipment is not decreasing fast-enough to compensate for the fading revenue per bit. To aggravate the situation, the uncertainty around the cost to support next generation services with IPoDWDM technology is also very high. As shown in this case study, the cost of such networks follows the exponential growth of the forecast traffic, which dampens the competitiveness of network operators.

Optimizations on the many decades old base IP and optical technologies are not enough for compensating for the rapid surge in the popularity of data networking and the generated traffic.

Sub-wavelength technologies have been extensively studied over the past decade to address the inefficiencies in the financial models of traditional technology. A multitude of optical burst switching technologies have been proposed in an effort to improve network resource utilization and reduce cost as surveyed in [10]. However, as it turns out, not all of these technologies are actually capable of delivering commercial benefits and hence they never make it to commercial implementation [10]. This further proves the very high challenge that this cost optimisation problem is posing.

The OPST sub-wavelength packet forwarding technology is currently one of the very few such technologies that is proving its commercial benefits on the market. OPST is built around a globally unique technology innovation that enables fast tuneable lasers to be used for real-time (nanosecond level) packet switching. This innovation makes it possible to bring the flexibility, efficiency and sophistication of the IP layer down into the optical layer, making it possible to truly virtualize the network infrastructure. This enables linear network cost evolution to address exponential traffic growth; hugely reduced risk in front of unpredictable demand evolution; and a light, sustainable architecture. In other words, the IP Router solution results in efficiency in the network which grows substantially when the size of an electronic switching fabric cannot grow to handle all of the traffic in a region in one fabric. As soon as multiple fabrics need to be interconnected, the resulting complexity of port interconnections required accelerates the costs compared to the linear growth of OPST switch capacity as it is added per optical port to the network. Note that any introduction of protection paths and physical chassis redundancy would have a similar effect, whereas OPST has a dual-ring implementation for its optical switching fabric and therefore in-built protection for the same cost.

2.2. CDN optimization

The growth in content and in particular video services significantly increases the data traffic loads impacting service operator networks. Content Distribution Networks (CDN) are used for storing and distributing content more effectively [12]. This case study considers the deployment of a dedicated network for the distribution of content services. Needless to say, the closer the content is to the user the better, because the video traverses less network elements, with less risk of packet loss and therefore reduced quality of experience for the video audience. The idea is either to use a dynamic DNS (Domain Name Server) that points to the CDN server assigned to the geographical region of the user or to perform a HTTP redirection from a central site. There is an ongoing debate in the industry around small distributed versus large centralized caches. The former can be located closer to the user, thus decreasing the chances for bottleneck formation. The latter concentrates a larger number of users, thus the popular, cached video is closer to the overall user population. Zink et al. estimated that at the campus network level the probability of finding a specific video in a local cache (aka hit rate) is 30% [13]. On the other hand, Cha et al. performed the same estimation considering the whole population of Korea [15]. Three different caching strategies were proposed and the most efficient one produced a 98% hit rate, with a storage size of less than 300,000 videos. This is because the cache population is large and the chance to find a previously downloaded video increases accordingly. The results of [14, 15] motivate the optimal utilization of efficient network architectures in the metro network. The CDN reduces the traffic from other networks, but as the cache must cover a high number of users the traffic load in the metro network is not reduced.

2.2.1. Comparative analysis

In this study, both video on demand (VoD) and live IPTV distribution will be integrated over the metro infrastructure used for residential Internet access. Both VoD and IPTV will be offered by means of Unicast connections to some distributed servers, which are fed from CDN entry points on higher network hierarchical levels.

The overall cost and power consumption that can be deemed to Video service deployment will comprise both Video server platforms and metro nodes and transmission resources. An analysis will be made on joint optimization of network and IT resources needed to provide a high quality video offer at the lowest cost and consumption.

A densely populated metro region will be assumed for the network model that could be representative of the suburban area of some major European city.

Two different implementations will be compared:

- Based on hierarchical VPLS (meshed network).
- Based on OPST rings.

2.2.1.1. Service deployment characteristics and constraints

In the proposed model, the video server streamers (for VoD and live IPTV services) are placed on a subset of metro nodes. There is a tradeoff between the number and performance of deployed streamers and the transmission resources needed to provide connectivity with the servers. Different scenarios will be considered for the first scenario (current H-VPLS architecture), so that the best IT/network combination is obtained and compared with the second (OPST) architecture.

In any specific scenario, IT servers will be dimensioned accounting for peak activity, plus some additional resources (over dimensioning ratio) to cope with server or even complete location failure.

The main characteristics of the planned service are:

- A total of **750K residential users** to be served:
    - 100K users enjoy IPTV and high quality VoD services.
    - There are up to 40K concurrent IPTV/VoD users at peak hour.
    - Average speed of video stream flows (high and low definition mix): 5 Mbps.
- **4 different locations providing connection to Internet exchanges**. The Internet average peak rate per user is 150 Kbps. There is a load balancing policy between the BRASes located at the four access nodes.

  A baseline scenario (before Video service introduction) is considered. In this scenario (Internet service only) 10GbE links are considered for every network connection, even though traffic load could be carried over a limited number of 1GbE links. The reasoning behind that is the low price of 10Gbps interfaces, the expected growth of Internet access and the need to save fiber resources within the network.
- 3 different **Video server models** considered:
    - High Capacity Server (HCS): 20Gbps (4000 users).
    - Medium Capacity Server (MCS): 10Gbps (2000 users).

   - Low Capacity Server (LCS): 5Gbps (1000 users).

2.2.1.2. Network architecture: H-VPLS

The first metro model used in the study is based on a hierarchical VPLS architecture, shown in next figureFigure 2.2.1.2-1: H-VPLS Metro model

.

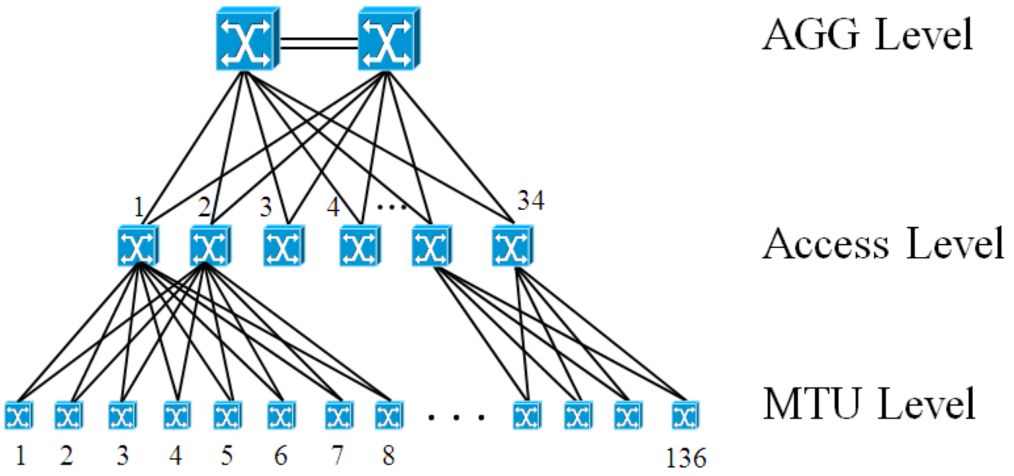


Figure 2.2.1.2-1: H-VPLS Metro model

In this network model, the lowest hierarchical level (named MTU for “Multi-Tenant Unit”) provides capillarity to the Core VPLS network formed by the other two hierarchical levels (namely access and aggregation levels) and concentrates traffic from multiple DSLAM/OLTs.

We will consider that the Core VPLS network in our use case serves the purpose of providing mesh connectivity between every network node, but no “end customers” (DSLAMs, OLTs…) will be connected at the first two levels. The Metro nodes at aggregation and access levels will be supported over some sort of optical transport (SDH, CWDM, DWDM rings…), depending on its availability.

MTUs will always be connected to a pair of access nodes by means of GbE/10 GbE optical links through diversified fiber paths. In this case either dark fiber or optical transport could be used.

As MTU nodes and their connections to the Core will be common to H-VPLS and OPST architectures, they will not be taken into account in this analysis.

We will initially consider the following number of nodes in the Metro area:

- **Access elements**: 2000 DSLAM/OLTs.
- **MTU** nodes: 136.
- **Access nodes**: 34.

- **Aggregator** nodes: 2.

For the access and aggregator levels, and based on commercial implementations, we will consider a high performance switch with 10 available slots for line cards and 100Gbps switching capacity per slot. The MTU level would be implemented with smaller size/performance switches, but it will not be considered in this study since, as already mentioned, the same element could be used for both H-VPLS and OPST architectures.

Video servers will be connected either at the access or at the aggregation level. The number of each server model and over dimensioning ratios will be dependent on the selected scenario.

The following user distribution (Table 2) has been designed for the study. There is a total of 20 administrative zones (A to T), each with 1 to 3 metro access nodes, that concentrate some traffic volume to/from a number of MTU nodes, the last level of the metro hierarchy.

Aggregation nodes will be placed at central city locations. Link distances to the VPLS Core are provided in the table, being the longest link slightly exceeding 50 Km.

Cost and power considerations

Based on TID internal data and [HED], the following values are considered for the model (Table 1):

		Relative Cost [cost units]	Power consumption [W]
Network resources	Chassis and common elements	1,4	250
	10GbE port	0,32	38 [HED]
	40GbE port	2	105 [HED]
Transport resources	10G transponder	Max: 1,4 (estimated)	50 [HED]
		Min: 0,7 (estimated)	
	40G transponder	Max: 5,6 (estimated)	200 (estimated)
		Min: 2,8 (estimated)	
IT resources	HCS (20G, 4000 users)	4	100
	MCS (10G, 2000 users)	2,4	75
	LCS (5G, 1000 users)	1,5	55

Table 1: Cost and power model

In the cost model, the cost unit is equivalent to the IT price for a capacity of 1000 users in a HCS server.

For the power calculations in the Transport layer, we consider an extra 20% accounting for the power overhead due to the optical switching, amplification, control, etc. along the optical paths.

Zone	Node ID	Distance to AG1 (Km)	Distance to AG2 (Km)	Population per node area (K)	Global user base (K)	Video user base (k)	Video demand at peak hour (K flows)	Average Internet BW at peak hour (Mbps)
A	1	20	30	146	34,696	4,626	1,850	5204,37
B	2	30	20	142	33,745	4,499	1,800	5061,79
C	3	20	30	124	29,468	3,929	1,572	4420,15
D	4	30	20	148	35,171	4,689	1,876	5275,67
E	5	35	38	80	19,011	2,535	1,014	2851,71
	6	32	36	74	17,586	2,345	0,938	2637,83
F	7	34	36	153	36,359	4,848	1,939	5453,90
G	8	34	40	143	33,983	4,531	1,812	5097,43
H	9	35	32	98	23,289	3,105	1,242	3493,35
	10	30	32	54	12,833	1,711	0,684	1924,90
	11	40	38	80	19,011	2,535	1,014	2851,71
I	12	30	34	86	20,437	2,725	1,090	3065,59
	13	28	32	66	15,684	2,091	0,837	2352,66
J	14	50	48	117	27,804	3,707	1,483	4170,63
K	15	30	34	75	17,823	2,376	0,951	2673,48
	16	35	38	74	17,586	2,345	0,938	2637,83
	17	32	36	62	14,734	1,965	0,786	2210,08
L	18	28	25	60	14,259	1,901	0,760	2138,78
	19	30	32	83	19,724	2,630	1,052	2958,65
M	20	36	42	40	9,506	1,267	0,507	1425,86
N	21	30	34	107	25,428	3,390	1,356	3814,16
O	22	32	38	83	19,724	2,630	1,052	2958,65
	23	40	46	50	11,882	1,584	0,634	1782,32
P	24	38	42	132	31,369	4,183	1,673	4705,32
Q	25	30	32	101	24,002	3,200	1,280	3600,29
	26	32	30	80	19,011	2,535	1,014	2851,71
	27	30	36	76	18,061	2,408	0,963	2709,13
R	28	35	40	112	26,616	3,549	1,420	3992,40
	29	34	38	60	14,259	1,901	0,760	2138,78
	30	30	36	68	16,160	2,155	0,862	2423,95
S	31	45	50	64	15,209	2,028	0,811	2281,37
T	32	30	35	85	20,200	2,693	1,077	3029,94
	33	30	32	82	19,487	2,598	1,039	2923,00
	34	34	32	75	17,823	2,376	0,951	2673,48

Table 2: Traffic matrix

Dimensioning, resiliency and scalability issues

Following the model shown in Figure 2.2.1.2-1: H-VPLS Metro model

, we will make the following assumptions:

1. Network nodes are located on exchanges with dual output manholes. All possible individual fiber cables comprising a logical link follow the same route. It is assumed that a failure in one logical link will affect all the fibers between the two link endpoints.
2. To optimize fiber utilization and performance, the following dimensioning rules are used:
    i. Deploy 40GbE interfaces when there is more than 20Gbps of remaining traffic per logical link. For instance, if there are 65Gbps in a given link, it will be composed by two 40GbE interfaces.
    ii. Deploy 10GbE interfaces when there is less than 20Gbps of remaining traffic per logical link. For instance, if there are 55Gbps in a given link, it will be composed by one 40GbE interface and two 10GbE interfaces.
3. Service will not be disrupted after a failure on any of the two logical links from an access node to the aggregation level. The surviving link (over a diversified path) is able to carry the whole service demand.
4. At any time, a single access node can fail or lose both links to the aggregation level. Access nodes will be arranged in pairs (every node has a “dual” one), so that one can fully protect the other.

    In case an access node fails or gets completely isolated, MTUs will redirect all traffic to their dual access node, since access nodes will be dimensioned to absorb this extra traffic (Figure 2.2.1.2-1: H-VPLS Metro model

).

    A total of 4 MTUs will be server by an access node and backed up by its dual node. This means an access “pair” is able to absorb the traffic from/to 8 MTU nodes.
5. One of the aggregator nodes could fail. The remaining aggregator will cope with all the traffic demands.
6. At any time, all the video streamers at one specific location (access or aggregator node) can fail or the network location can get totally isolated/unreachable. In this case, service requests will have to be attended using IT resources at any other location.

2.2.1.3. Network architectures: Metro Network model based on OPST

This model will be based on a flat architecture supported on a number of OPST rings. Each ring will comprise several metro access nodes, where MTUs are connected in a dual-home fashion. Opposite to the previous network model, all the ring nodes belong to the same network hierarchy and are liable to provide access to the video streamers. This is shown in Figure 2.2.1.3-1.

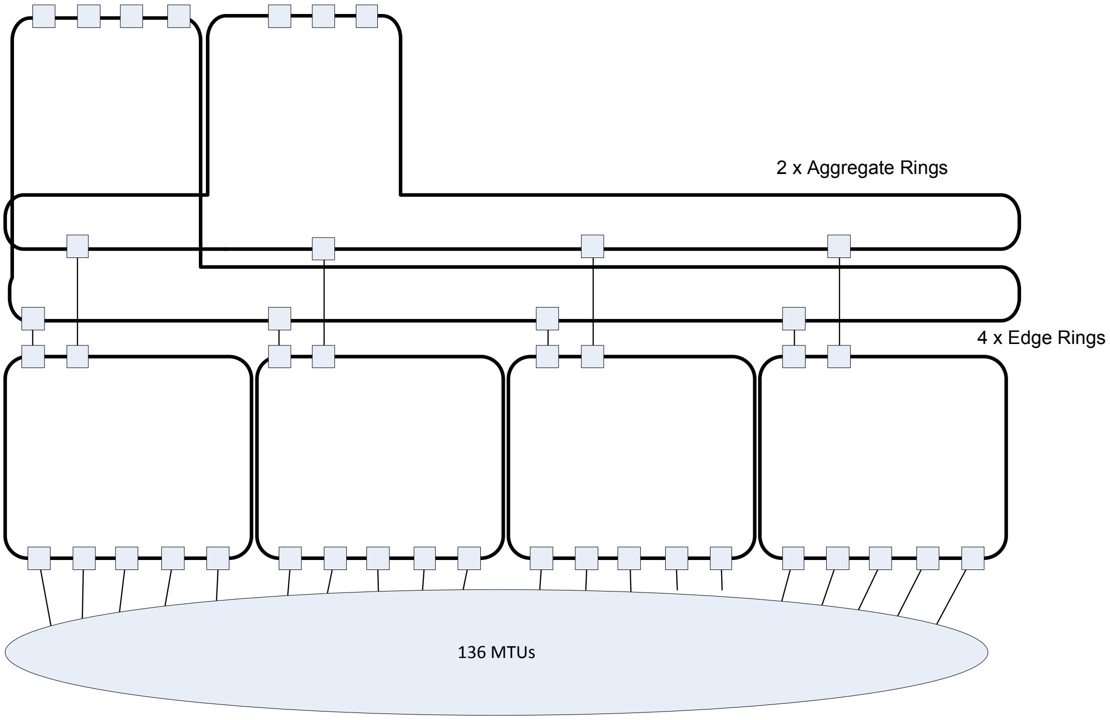


Figure 2.2.1.3-2- OPST Metro model

Cost and power considerations

The following values are considered for the model (Table 3):

		Relative Cost [cost units]	Power consumption [W]
Network resources	Chassis, common elements and Optical Transport	3.19	350
	Switching Fabric + 10G port	0,48	122.5
IT resources	HCS (20G, 4000 users)	4	100
	MCS (10G, 2000 users)	2,4	75
	LCS (5G, 1000 users)	1,5	55

Table 3: Cost and power for OPST model

In the cost model, the cost unit is relative to that used in the H-VPLS model to enable comparison. The OPST solution is a convergence of optical transport and Layer 2 switching, hence the Common elements include optical transport. The system is architected in such a way that adding 10GE ports, adds switching capacity to the entire ring, hence the price and power consumption of each 10GE port include a portion of switching.

OPST architecture

OPST comprises two independent distributed switches that use optical fibre as the medium for the switching fabric. These optical fabrics are arranged as two fibre rings with opposing direction of transmission. The dual optical fabrics enable two key advantages:

- 2x speedup in burst forwarding
- Provision of optical path redundancy for the protection mechanisms

From a logical perspective, the network topology is that of a non-blocking full mesh of interconnected locations as illustrated in Figure 2.2.1.3-3, i.e. there is full logical mesh connectivity that operates on a physical ring topology. Each of the Nodes illustrated below home a number of client interfaces that present the external traffic to the system. Configuration of traffic flows over the optical ring provides for any-to-any packet flow connectivity (with defined class of service behaviour) across the OPST structure.

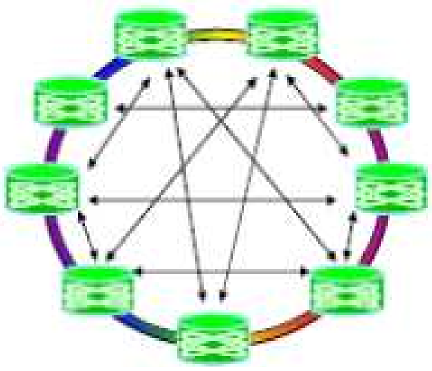


Figure 2.2.1.3-4-OPST Full Logical Mesh Any to Any Optical Path Topology

The convergence of burst mode photonics and distributed traffic scheduling within OPST creates the functional equivalent of a single Ethernet switch with distributed traffic interfaces – specifically applicable to Metro Area network requirements.

OPST is not an arrangement of Ethernet switches that are connected by burst-mode optical transmission interfaces. OPST is a fully converged networking platform, switching packets in the optical domain.

The optical fibre that connects the OPST Nodes around the ring circumference should be viewed as the ‘switching fabric’ of the OPST platform. The two optical fibre rings form two autonomous, redundant packet switching fabrics that use burst mode photonics to provide managed ingress and egress of packetised traffic flows.

Based on the industry standard ITU grid, wavelengths in the C-band can support up to 40 λ's, each at a nominal line rate of 10 Gbps. Therefore a nominal, non-blocking, full-duplex capacity of 2 x 400Gbps is created across the ring.

Each of the external facing client interfaces are assigned a dedicated wavelength to be used for the optical reception of traffic from other client interfaces equipped in other Nodes. Each client interface is capable of transmitting traffic to all other client interfaces by optical burst transmissions on the appropriate wavelength.

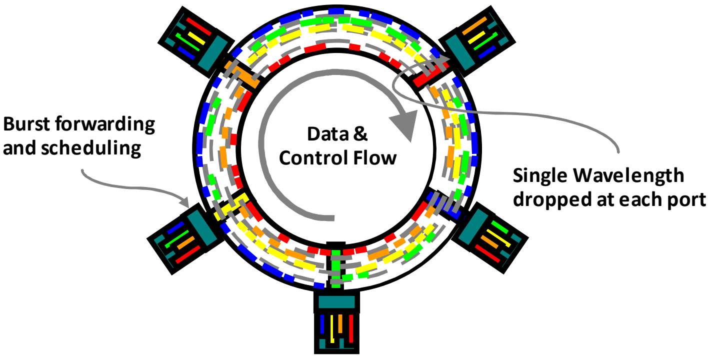


Figure 2.2.1.3-5-OPST Illustration of a wavelength routed system

A resilient configuration is created by utilising two fast tunable laser and burst mode receiver devices that form part of each client interface that in turn connect to both optical rings simultaneously. This forms the basis of the OPST load-sharing and protection mechanisms.

The rapid and reliable wavelength switching time of optical burst technology to create this new genre of OPST based network platform can be viewed as being analogous to a 'time share scheme' for lasers. A key factor of the mechanism for controlling the 'time share scheme' is that it is completely asynchronous and adaptive to the traffic requirements across the ring; i.e. nodes do not require fixed or reserved time slots in order to transmit to each other.

OPST technology will provide service providers with a lower total cost of ownership compared to any alternative technology whilst simultaneously providing the flexibility to meet the increasing unpredictable traffic growth and uncertainty of tomorrows complex non-hierarchical and 'any to any' services in a continuously increasingly competitive environment.

As a result of implementing a fully merged L0 to L2, OPST solution - providing aggregation, grooming, switching and transport within a single platform – service providers will be able to deliver fixed and mobile, residential and commercial, QoS and 'best effort' services across a single managed IP network, and single control plane, offering interoperability with existing MPLS core/service edge networks and existing access/edge access node.

Assumptions for the scenario

OPST model follows the same assumptions as the H-VPLS scenario.

2.2.1.4. Detailed definition of scenarios and results

A preliminary analysis showed that the required switching capacity for the aggregation level was higher than the maximum achievable with the 10-slot chassis model being considered in the study (i.e. 1 Tbps). Also, power dissipation needs if the whole node was filled with maximum capacity boards (100GbE) could be very high.

Therefore, in order to obtain a feasible solution, it was decided to separate the analysis of the Metro Network described in Table 2 into two separated Metro Area Networks (MANs): one with the first 18 access nodes and the second with the remaining 16 access nodes. By doing this, the switching capacity of the 10-slot chassis is sufficient for the traffic that the aggregate switches are processing, avoiding the necessity of multi-chassis configurations. Maximum power dissipations figures will be reduced as well, since now each aggregator node needs to cope with less than 500Gbps.

Overall network dimensioning

The analysis is based on a dimensioning rule where no specific content distribution policy has been defined for the video service deployment. This means that all the video streamers receive/provide all contents and all users can be served by its assigned video streamer even at peak hour. The goal is to determine the optimum quantity of servers and the most appropriate location to obtain the lowest cost and consumption of "the service". We assume here that an ideal load balancing mechanism is deployed to assign Video streamer resources to all of the end-users.

We consider three different approaches for the overall network and IT deployment: (1) IT at aggregation level, (2) IT at access level, (3) IT at aggregation and access levels. In these 3 study cases, there is an equal IT distribution. But we have also considered additional scenarios similar to (3) in which there is an unequal IT distribution between aggregation and access levels.

We will apply these approaches to both MANs, i.e. the one with 18 access nodes and the other with 16 access nodes.

First, we will dimension the base-line scenario just with Internet traffic. It is worth to mention that the internet servers are located in nodes 4, 8, 12 and 14 (according to the node IDs in Table 2) in the MAN with 18 access nodes. In the MAN with 16 access nodes, the internet sites are the nodes 21, 24, 28 and 32 (according to the node IDs in Table 2).

Baseline scenario

We will determine in the first place the cost and power consumption before Video services introduction. As mentioned before, only 10Gbps links will be considered.

According to Table 1, the total cost of this solution is the sum of the cost of the network resources (i.e. chassis and ports), since we do not have any IT servers.

Taking into account the location of the internet servers and the traffic generated by the corresponding user base in each access node (according to Table 2), a traffic matrix is obtained. From this demand, the dimensioning of the network nodes (i.e. aggregators and access nodes) and links (i.e. between ACC and AGG level and from ACC level to MTU level) is computed, considering all the previous assumptions in order to provide the appropriate level of protection and, therefore, guaranteeing the video service availability. As MTU level is not considered in the study, only the access side of the access-MTU links has been included in the cost/power calculations.

In Table 4, the network elements needed in each MAN (i.e. the one with 18 access nodes and the other with 16 access nodes) are shown.

	10GbE ports	Chassis
Number of elements (MAN with 18 access nodes)	284	20
Number of elements (MAN with 16 access nodes)	224	18

Table 4: Number of ports and chassis in each MAN for the baseline scenario

Adding 10GbE ports and chassis for both MANs, we obtain the total amount of network elements in the baseline scenario (**Figure 2.2.1.4-1**).

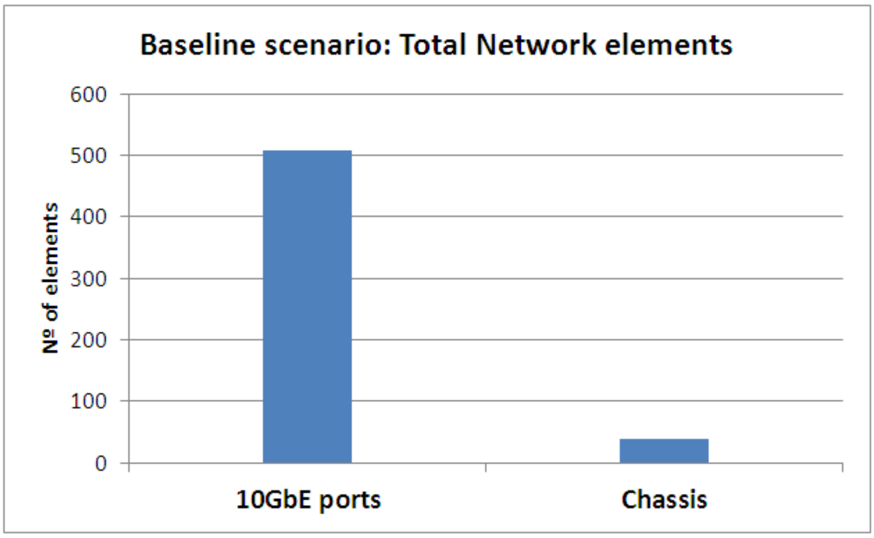


Figure 2.2.1.4-2-: Total Network resources needed in the baseline scenario

IT at aggregation level

Taking into account the two separated MANs, we initially consider that (for each MAN) we have just a couple of clusters of high capacity video streamers, each placed at an aggregator site. The overall cost of the service is driven, on one side, by the necessary capacity upgrade in the links of every access node (i.e. 18 or 16 access nodes) towards the two aggregators (i.e. 36 or 32 links), that have to carry all the video flows of their associated user base and, on the other side, by the required IT server capacity.

Every access to aggregator link is able to cope with all the traffic demand towards the servers, preventing service disruption after a failure in one of the two access-to-aggregator links. If, for example, access nodes 1 and 2 (dual access node) in Figure 2.2.1.4-3 (left – MAN with 18 access nodes) have a total traffic demand (i.e. internet and video demands) of T_1 and T_2 respectively, the links between them (access node 1 and access node 2) and the aggregators are provisioned for T_1+T_2 capacity. The same occurs for the rest of access node pairs (and for the MAN with 16 access nodes in Figure 2.2.1.4-4- right).

This means that for a total video demand of $B=T_1+T_2+T_3+T_4+\ldots+T_{17}+T_{18}$ (over the 36 links) there is a need to provision a *4B* capacity. A factor of 2 due to each pair of access node and dual protecting each other and the other factor of 2 because there are two links from each access node towards the aggregator nodes.

We can assume that aggregator and server failures will happen very seldom, while failures on access-to-aggregator links can happen more frequently due to constant civil works being done in urban areas. If we wanted to maintain a perfect load-balancing policy for the video servers usage, we should also consider upgrades to the inter-aggregator links. By doing this, traffic demands directed to server-A at aggregator-A could reach their destination when one/some access nodes lose their link to aggregator-A. We will not consider, for the sake of simplicity, this aggregator-to-aggregator link upgrade in this study.

However, we will consider the (non simultaneous) event of a complete aggregator or service location failure. This means that server resources need to be duplicated at every aggregator location so that we can allow for load variation (unbalanced) and redirect demand to server-B when access to aggregator-A links are degraded.

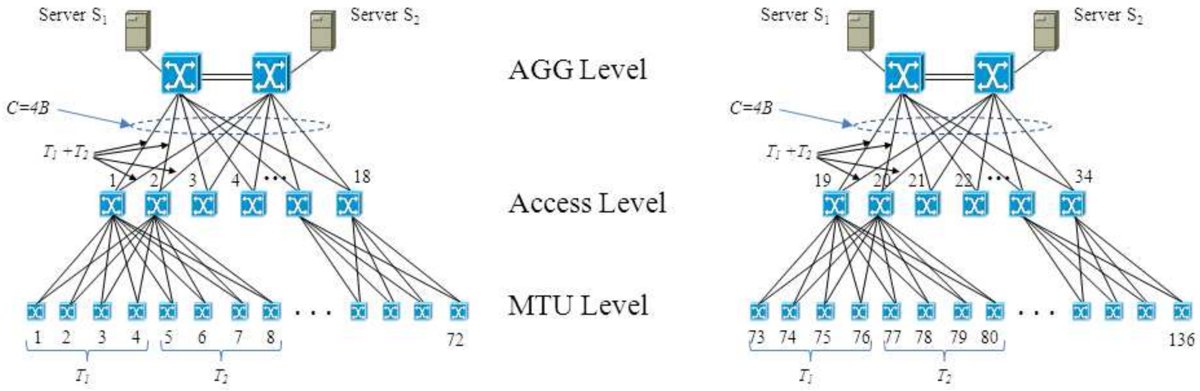


Figure 2.2.1.4-5-: IT servers at aggregation level

The network resources are computed from a traffic matrix that considers the location and traffic from the video servers.

Table 5 shows the number of 10 and 40GbE interfaces needed to deploy this infrastructure, assuming no initial deployment, that is, a Greenfield scenario. Again, we are dimensioning the links between ACC and AGG levels and from ACC to MTU level (the same applies for the following scenarios).

	10GbE ports	40GbE ports
Number of elements (MAN with 18 access nodes)	312	8
Number of elements (MAN with 16 access nodes)	272	0

Table 5: Number of ports for the scenario 'IT at aggregation level'

The number of chassis is the same as in the baseline scenario. The same will be consistent for the rest of scenarios within this study.

With regards to IT resources, since each video service location is protecting the possible failure or isolation of the other one, both need to be dimensioned to cope with all the video service users. Therefore, a relevant number of high capacity servers (HCS) per location is expected.

As in the MAN with 18 access nodes we have 56.464 video users (see Table 2) and a high capacity server (HCS) can offer video service to 4.000 users, 14 HCSs are required. The remaining 464 users can be served by a LCS (i.e. 1.000 users). Therefore, according to Table 1, the IT relative cost in each location is 14x4+1x1.5=57.5. Since we have two locations in this scenario, the IT relative cost is 115.

In the MAN with 16 access nodes we have 41.128 video users (see Table 2). Thus, 10 HCSs and 1 MCS are needed, whose cost is 42.4. As we have two locations, the IT relative cost is 84.8.

In summary, we have a total IT cost in this scenario (including both MANs) of 199.8 cost units.

IT at access level

Now, we suppose that we place IT resources on 4 access node locations in each MAN (**¡Error! No se encuentra el origen de la referencia.**). The provisioning of the links between the access nodes with no IT resources and the aggregators is the same as in the previous scenario (i.e. the sum of the traffic from their assigned MTUs and the traffic from the MTUs assigned to the dual access node). However, we need to upgrade the links from the aggregators towards the access nodes with IT resources.

In case one of these access nodes crashes, gets isolated or there is a partial/total unavailability of IT resources we assume that the traffic that was served at that location will be equally distributed among the other three IT sites. Therefore, these links need a provisioning factor of 4/3. As a consequence, the cost of the network resources will be higher than in the previous case.

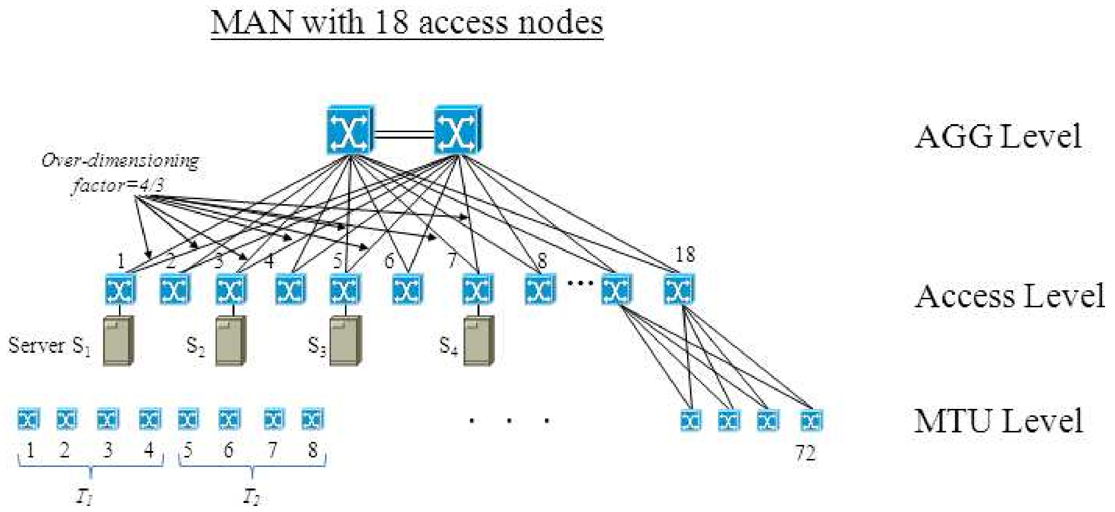


Figure 2.2.1.4-6: IT at access level (with 4 clusters of video streamers)

The number of the required 10 and 40GbE interfaces can be seen in Table 6.

	10GbE ports	40GbE ports
Number of elements (MAN with 18 access nodes)	268	40
Number of elements (MAN with 16 access nodes)	264	20

Table 6: Number of ports in the scenario 'IT at access level'

For the MAN with 18 access nodes, each location has to cope with 56.464/4=14.116 video users. The over-dimensioning factor for these servers is 4/3 to allow a complete IT site outage. Therefore, we need a capacity for 18.822 video users at each location.

As a HCS can serve up to 4.000 users, we need 4 HCSs per location and the remaining 2.822 users can be served by a MCS (i.e. 2.000 users) and a LCS (i.e. 1.000 users). The cost of the IT resources per location is then 4x4+1x2,4+1x1,5=19.9, giving a total IT cost of 79,6 units.

For the MAN with 16 access nodes, we make the same computation but taking into account that the number of video users per location is 10.282. Thus, 3 HCSs and 1 MCS are required, whose cost is 14.4, giving a total IT cost of 57.6 units.

The total relative cost of the IT resources in this scenario (comprising both MANs) is therefore 137.2 units.

After the analysis of the first two approaches, with IT distribution either at aggregation or access level, it becomes evident that from the cost perspective placing the IT resources at the upper level (i.e. aggregation) is the most beneficial option. This is mainly due to the lower transport costs in this solution and, in a smaller proportion, to a higher user concentration level that allows optimization of IT resources.

In the next study cases we will show how overall costs decrease as IT resources are progressively moved to the aggregator sites.

IT at aggregation and access levels

A. Equal IT distribution between upper and lower levels

In this scenario, we have again 4 video servers, but now two of them are located in the aggregator nodes and the other two servers are placed in the access level (Figure 2.2.1.4-4 shows this situation for the MAN with 18 access nodes). There is a similar picture for the MAN with 16 access nodes.

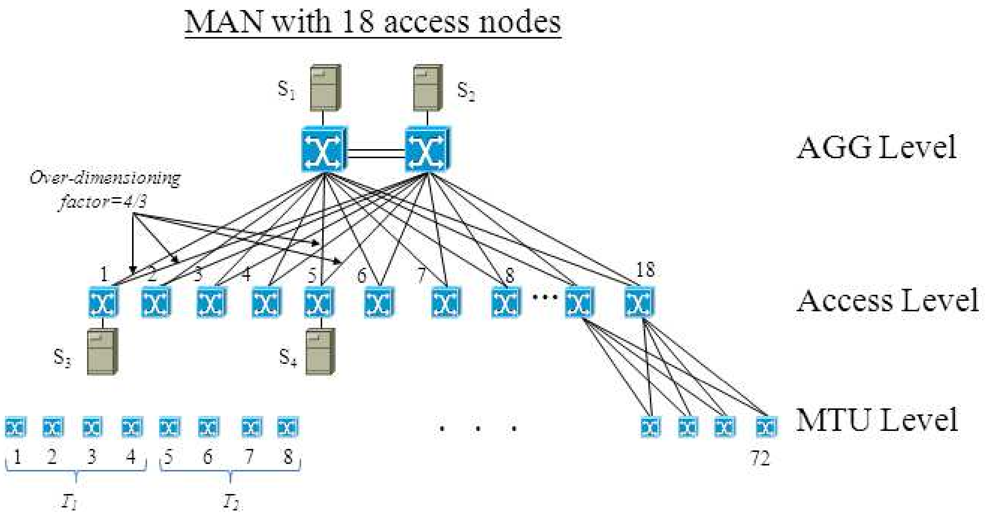


Figure 2.2.1.4-7: IT at aggregation and access level

As in the previous study case, in case of failure in a server location, the corresponding traffic is redirected to the other three IT sites in equal proportion. For this reason, the IT requirements are the same as in the previous study case (i.e. a relative cost of 137.2 units).

However, in this scenario, since we have two servers in the aggregation level, there are fewer links to be upgraded. Therefore, the cost of the network resources will be smaller.

Next, it is shown the number of 10 and 40GbE interfaces (Table 7).

	10GbE ports	40GbE ports
Number of elements (MAN with 18 access nodes)	288	24
Number of elements (MAN with 18 access nodes)	264	12

Table 7: Number of ports in the scenario 'IT at aggregation and access levels'

B. Unequal IT distribution between upper and lower levels

This last scenario is similar to the previous one. The only difference is that the IT resources are not equally distributed among access and aggregation level.

The video servers located in the aggregator sites have been over dimensioned with different factors (from 1.5 to 2.25) so that they can cope with a larger part of the traffic when any IT site is unavailable. The video servers placed at the access level, on the other hand, need to be dimensioned in a lower degree (from 1.25 to 0.875).

For every case, the links between the aggregator nodes to the IT locations at the access level will have to be dimensioned according to the IT capacity at that level, i.e. from 1.25 to 0.875. Thus, the network resources will be reduced but more IT resources will be needed.

The number and cost of the network and IT resources in each of these cases are shown in Table 8 and Table 9.

	10GbE ports	40GbE ports	10G transponders	40G transponders	IT resources		
					HCS	MCS	LCS
Case 1: x1.5 at AGG x1.25 at ACC	548	36	276	36	32	4	2
Case 2: x2 at AGG x1 at ACC	576	24	304	24	34	4	8
Case 3: x2.25 at AGG x0.875 at ACC	568	24	296	24	38	0	4

Table 8: Network, transport and IT resources in different examples of the scenario 'IT at aggregation and access levels – unequal IT distribution between upper and lower levels'

	10GbE ports cost [cost units]	40GbE ports cost [cost units]	10G transponders max. cost [cost units]	40G transponders max. cost [cost units]	IT cost [cost units]	Total cost w/o chassis [cost units]
Case 1: x1.5 at AGG x1.25 at ACC	175,36	72	386,4	201,6	140,6	975,96
Case 2: x2 at AGG x1 at ACC	184,32	48	425,6	134,4	157,6	949,92
Case 3: x2.25 at AGG x0.875 at ACC	181,76	48	414,4	134,4	158	936,56

Table 9: Cost of the network, transport and IT resources in different examples of the scenario 'IT at aggregation and access levels – unequal IT distribution between upper and lower levels'

As can be seen, as more IT capacity is allocated at the aggregation level, the cost of the IT resources increases but in a lower proportion than the reduction of the cost of the network plus transport resources. Therefore, it is checked that the overall network cost decrease as more IT capacity is allocated in the aggregation level.

OPST

The OPST solution is shown in Figure 2.2.1.4-5:

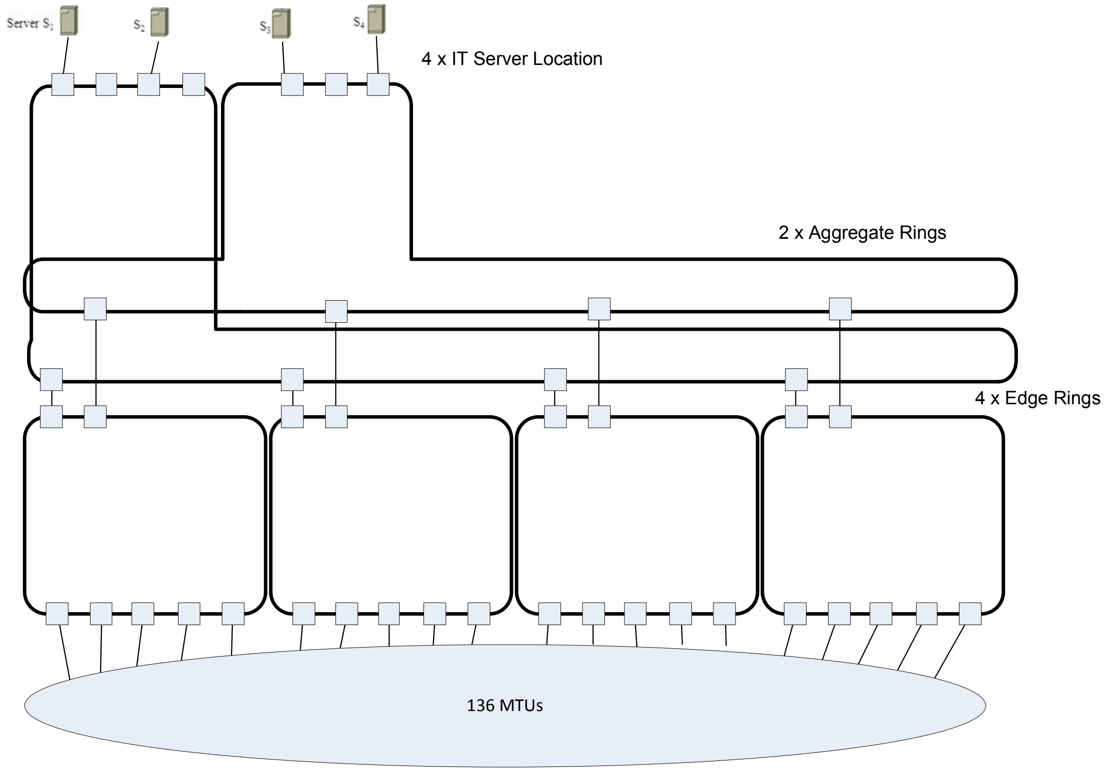


Figure 2.2.1.4-8: OPST Metro model

The total service source bandwidth for the VOD service and the BRAS peak loading for the entire network was calculated.

The peak load bandwidth for VOD and the peak load bandwidth for BRAS were converted to numbers of ports. The iVX8000 is a distributed switch with an optical fabric. This means that the port numbers must include a factor related to the optical burst scheduling algorithm and burst overhead. This factor increases the number of ports on each service to a maximum average loading of 6.4GE per port. This is used throughout the design.

Network resilience can be provided with a distributed switching system by placing the service sources in different sites then increasing the service source capacity by a factor such that if any one site is out of service the network will still be able to source to peak load capacity. If the capacity for peak loading is C_{peak} it can be shown that the capacity required for resilience is given by $C_{peak}(N/N-1)$, where N is the number of sites used to source the service to the network.. In the OPST design a number of sites must be chosen such that there is a balance between the reduction in required source capacity for resilience, the number of iVX8000 rings needed to provide load balanced service source to the network and the efficient use of the node capacity within each ring. For this design the number of sites chosen to source the services to the network was 4. The total failure of any one of these sites will have no impact on the peak demand for either VOD or internet browsing simultaneously peaking. The capacity increment required for this in terms of both the BRAS and VOD server capacity is 33% of the peak loading as compared to 100% for a 1 + 1 protected system.

The design has a service feed network comprising two service load balancing iVX8000 rings. There are 2 service feed sites on each ring with 6 nodes used in connecting the VOD servers and BRAS ports. The feed sites are arranged with 6,6 and 6 ports used on each node. This design is replicated on each ring over the 4 sites. Each iVX8000 ring from the service source sites connects to 4service feed sites. There are therefore 4 service feed sites on the A ring and 4 service feed sites on the B ring.

The MTUs connected to the service feed sites by 4 iVX8000 rings to give a total of 6 rings for the entire network. Each edge ring passes through 2 service feed sites, one from the A service load balancing service source ring and one from the B load balancing service source ring. One node from each edge ring is used to connect to the one node from the load balancing service source ring in each site.

This network has no single source of failure at node, port fibre or site level.

A site outage at the service source will have no impact on the service.

The maximum bandwidth per edge ring is significantly less than the port capacity and thereby the optical fabric forwarding capacity. This means that a single fibre break on any ring will have no effect on the ability of the network to provide continued peak load service.

A node or site outage at any of the service feed sites will cause a 50% reduction in peak service capacity for combined internet and VOD to the subscribers connected to the edge ring using that service feed site. This can be improved to 57% on the first two rings and 66% on the other two by the addition of extra ports on each.

The number of required 10GE interfaces and chassis are shown in next figure.

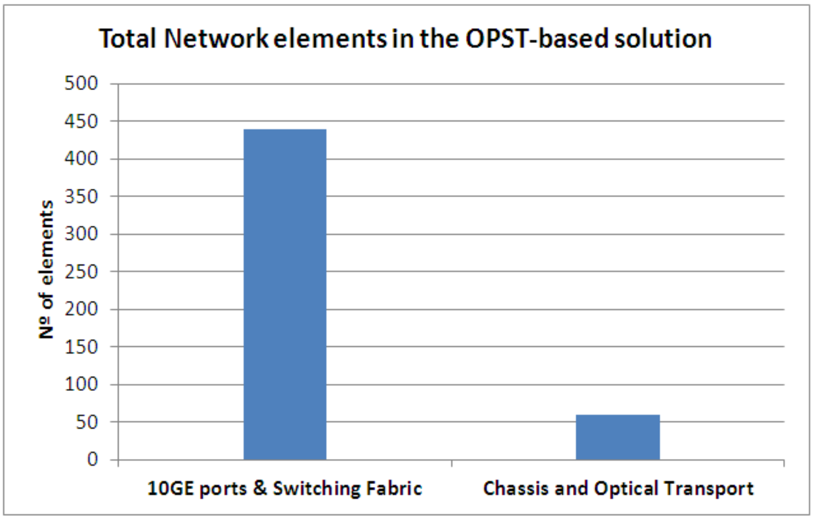
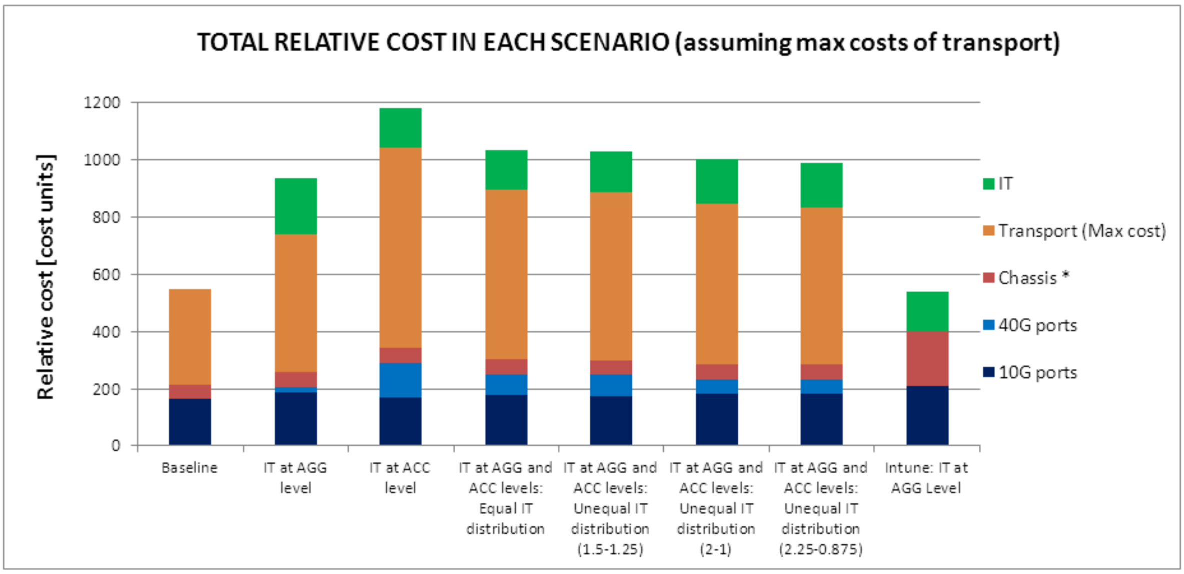


Figure 2.2.1.4-9: Network resources in the OPST scenario

Cost and power consumption comparisons

Once we have calculated the total network and IT resources for each study case, and added the cost of the transport for the two assumptions/estimations, we present a comparative analysis in terms of total relative cost, with a price breakdown per element.

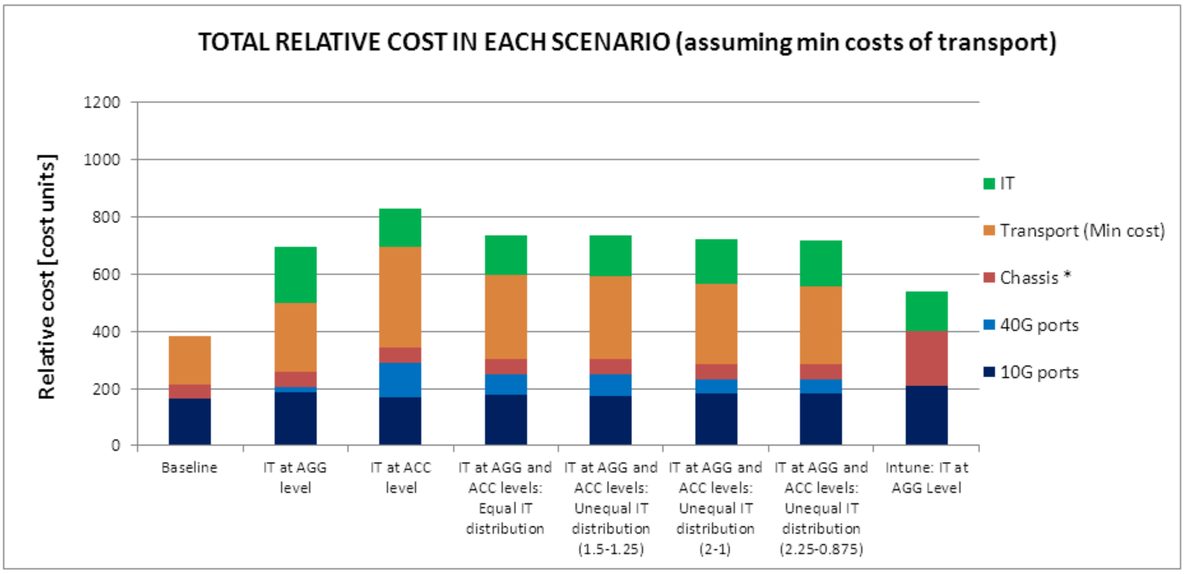
Considering the maximum cost of the transport, we obtain the following results (Figure below):



*Chassis on Intune solution includes transport

Figure 2.2.1.4-10: Cost comparison (Maximum transport cost assumption)

If we take into account the minimum cost of the transport, the results are the following figure:

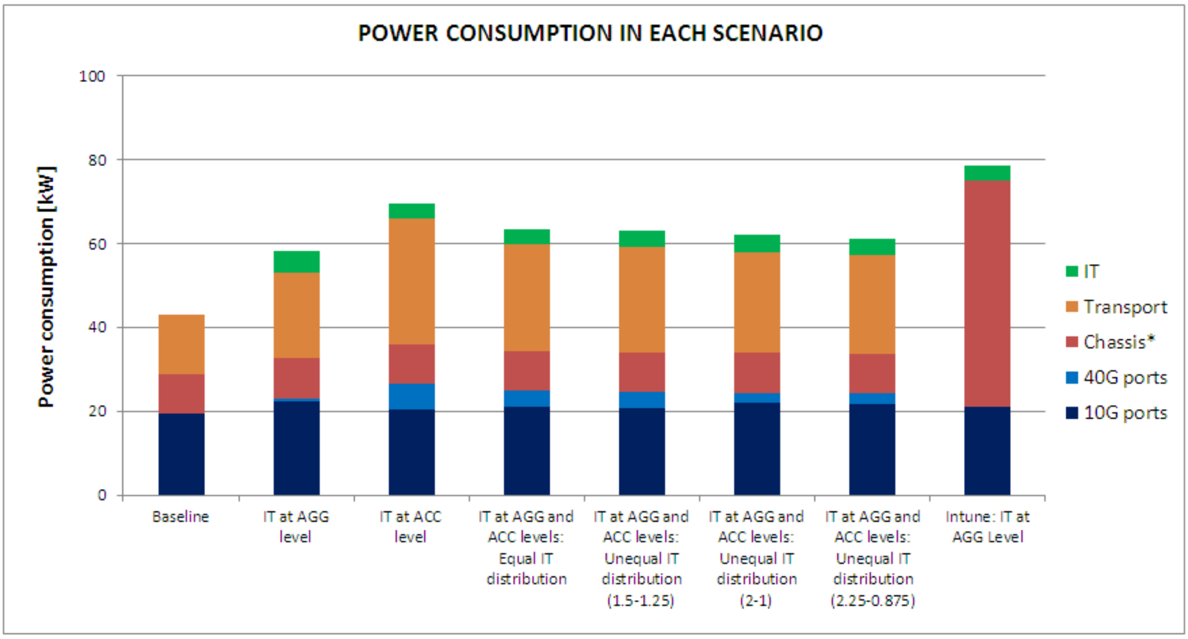


*Chassis on Intune solution includes transport

Figure 2.2.1.4-11: Cost comparison (Minimum transport cost assumption)

As seen in the figure, the OPST architecture is the most beneficial in term of capital expenditures. Within the H-VPLS architectures, the scenario *'IT at AGG level'* provides the required level of video service availability at the lowest cost.

An analysis of the total power consumption of both network and IT resources in each scenario has also been made. The results are presented in figure below:



**Chassis on Intune solution includes transport*

Figure 2.2.1.4-12: Power consumption comparison

In this analysis, a slightly higher power consumption is observed for the OPST architecture. We have to point out, though, that power figures used correspond to the very first OPST implementation and up to a 20% improvement on power consumption is expected in a further release. This means that OPST technologies will be advantageous in terms of overall network cost and recurrent power expenditures.

Conclusions

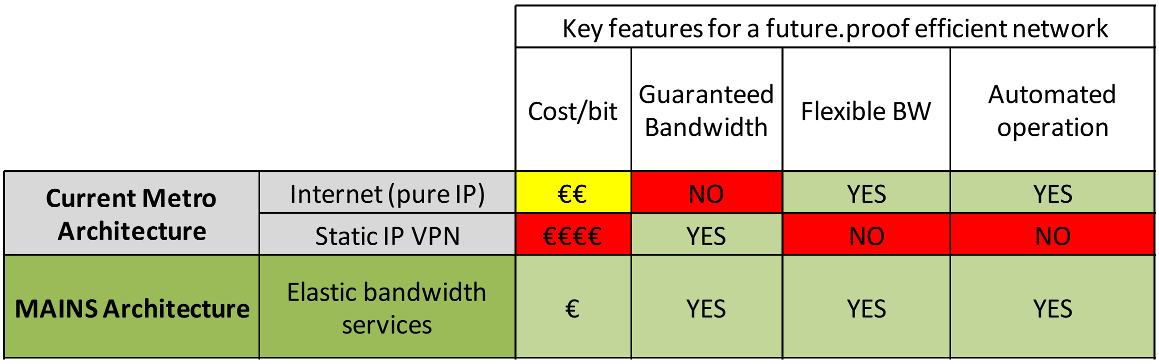
The comparative analysis presented in this section shows significant cost advantages (up to 42%) of MAINS architecture with respect to common H-VPLS for Unicast Video service distribution.

Additional advantages are also envisaged with respect to service availability (e.g restoration times <50ms for the OPST platform) as well as system scalability for increased service penetrations or traffic growth rates. Further OPST is expected to provide operational advantages in terms of reduced equipment count and configuration simplicity due to its integrated control plane. Service configuration is expected to be simpler, without the need for a hierarchical management system, whilst disaster recovery faster and more manageable due to the distributed ring architecture.

With regards to energy efficiency, potential savings are expected in further OPST implementations, as the reduced number of OEO conversions with respect to electronic aggregation (H-VPLS) makes OPST inherently more power efficient.

2.3. Elastic VPNs

MAINS architectures would also enable new Elastic VPNs services which are not available. As demonstrated in the industrial performance analysis (see section4). Subwavelength switching technologies can provide guaranteed bandwidth on demand from Mbps to tens of Gbps.



		Key features for a future.proof efficient network			
		Cost/bit	Guaranteed Bandwidth	Flexible BW	Automated operation
Current Metro Architecture	Internet (pure IP)	€€	NO	YES	YES
	Static IP VPN	€€€€	YES	NO	NO
MAINS Architecture	Elastic bandwidth services	€	YES	YES	YES

Figure 2.3-1:Rationale behind Elastic VPNs

Some potential use cases for elastic VPNS are the interconnection of datacenters with BoD for data intensive transmissions or the modification of guaranteed bandwidth between datacenters and access nodes according to dynamic traffic patterns.

2.4. Flexible Data Centre deployments

There are more than 400 cities around the world today with more than 1 million inhabitants, while the population of 30 metropolitan areas exceeds 10 million people. Such dense metropolitan areas are responsible for the majority of the traffic growth. According to recent traffic forecasts [1], by 2015 the global data consumption will cross the Zettabyte (1021 bytes) threshold. The largest portion of this data is some form of video content. However, next generation cloud services are also attracting a lot of attention and in the future they will progress toward commercial offerings. Combined with mobility, such cloud services can easily generate traffic flows capable of exceeding typical current networks capacity [3], even at a very low penetration rate.

These trends lead the telecommunications industry to the realisation that cost-effective scalability is a crucial characteristic for future network architectures.

In addition to scalability, there is huge uncertainty around the level and pace of the growth as well as around the location for placement of the service platforms. New services, content-rich applications and mobility are only the top of the list of factors that can easily and instantly swing the traffic pattern in a network.

The location that content will be provided from, i.e. the location of the data centres is also subject to conjecture. Network operators are heavily investing into data centres to maintain profitability in an environment where third-party content is increasing their cost, without impacting their revenues at the same extent. The ownership of the network infrastructure provides operators with a competitive advantage through control over performance that they are looking to monetise. This suggests that the source of content will migrate from the Internet - its current location - closer to the user, into the national backbone, metro head-end and/or even into the metro access sites. Designing a network that costeffectively provides high performance for delivering services with unpredictable traffic patterns is a major challenge to network operators.

In this section we analyse the scalability and durability of next-generation networks with an emphasis on the economics. Incremental network deployment, data centre location, network architecture, service mix, traffic and subscriber take-up are all taken into consideration. Furthermore, next generation networking technologies like sub-wavelength optical packet switching and IPoDWDM/IPoROADM are compared to understand the cost implications.

In the remainder of this case study Section II describes the modelling approach and assumptions used. Section III presents the techno-economic comparison of the modelled architectures. Finally, Section IV concludes the case study.

2.4.1. Network and data center modelling

Two network architectures are modelled and compared to analyse their characteristics at scale. A subscriber base of 1 million is provided with a set of services. Each service is allocated bandwidth per subscriber. The models calculate the aggregation, switching, grooming and transport resources required in the network to deliver the services when the data centre is placed in the core, at the metro hub and distributed out in the metro. When the resources are derived, the models calculate, using typical unit pricing, the capital cost of the network. The models were run at five points in time (T1,..,T5) that represent growing uptake of services. The resulting cost output was used to compare an optical sub-wavelength packet forwarding technology called Optical Packet Switch and

Transport (OPST) solution with an IPoDWDM one for each placement of the DC. OPST merges switching and transport into a single layer and thus provides a subwavelength optical packet forwarding platform that can aggregate, groom, switch and transport packets in a uniquely efficient manner [5]. A next-generation cloud service over OPST case study is presented in [3], where a brief summary of the OPST technology and equipment is also available. On the other hand, IPoDWDM is a reasonable solution for carriers to reduce investment, while absorbing traffic increment [4]. Traffic breakdown and growth is based on internal Telefónica sources and aligned with the surveys in [1] (see Figure 2.4.1-1:). The traffic is divided into a number of services and 5 points in time are chosen to illustrate growth. These time points are not necessarily annual, but could represent quarterly or monthly change as new services grow. Between T1 and T5 the traffic gradually grows from 300Gbps to 7Tbps to reflect a significant network scaling scenario. The time span of this scaling is variable for each network, depending heavily on local market dynamics, operator strategy and success. The Cisco survey [1] is the basis for the percentages of each service type in the overall matrix, whereas the absolute numbers of subscribers and network architectures are based on data provided as a guideline by Telefonica.

A. IPoDWDM Architecture

In the IPoROADM architecture (Fig. 1, bottom) traffic is collected by IP access routers from Passive Optical Network (PON) head-ends, called Optical Line Terminals (OLTs). It is then aggregated by hub routers located at the metro head-end. A dynamic Wavelength Switched Optical Network (WSON) [7] built with ROADMs is used in the backbone to transport the traffic to core routers, which on turn are connected to Internet exchange points [8]. Routers with 2Tbps non-blocking switching capacity are used in the hub and core. Each router has 20x100G slots that are populated with 10x10G port cards. When the capacity of a router is exceeded, router matrices are built. Router matrices can be placed in the hub or in the core. All of the relevant combinations of router matrix and DC placement are analysed. In the metro access sites the IP access routers have 8x10G slots for a total of 80Gbps non-blocking switch fabric capacity. Client-facing slots are filled with 4x10G port cards, displaying a 4:1 oversubscription. This is feasible, because it is assumed that the average utilization of the OLT 10G uplinks is 25%. Therefore, 4 OLT uplinks can be served by 1x10G access router slot. The network-facing access router slots are filled with non oversubscribed 1x10G port cards

B OPST Architecture

Multiple OPST rings can be connected using MPLSTP (Multiprotocol Label Switching – Transport Profile) or Intune's Three-Stage Switch (3SS) solution [5]. The 3SS consists of three sets of OPST rings. The first set of rings collects traffic from metro access sites. The third set provides connectivity to the backbone network. Finally, the second or middle set of rings provides interconnection for the first and third sets. Flows are managed through the 3SS using intersection controllers. There is one intersection controller for each middle stage ring that makes flow control through the entire 3SS automatic and transparent to the operator. From a service provisioning perspective, the 3SS appears as a single L2 switch, while it comprises multiple optical platforms interconnected on fibre rings from an operations and maintenance perspective. In the comparative OPST architecture (Fig. 1, top) a 3SS carries out the aggregation, grooming, switching and transport functions in the metro network. The 3SS substitutes all of the routers and ROADMs used in the IPoROADM solution for the metro.

C. Data Center Location

Data centre location can have a significant impact on the cost and performance of a network. In this study different configurations are analysed, where the data centre is placed in the core, in the hub or embedded/distributed in the metro access. Currently most of the data centre traffic is coming from the Internet. However, with increasing service uptake and lower latency requirements for key cloud applications, the data centres in the metro hub and out into the metro can better respond to the performance and cost requirements. All of these DC location considerations are addressed in the models and discussed in detail in Section III of this case study, in conjunction with the modeling outputs.

It should be noted that as the data centre location moves closer to the user, it must change form to become a distributed data centre in order to retain homogeneity from the viewpoint of a user. In other words, smaller data centres located for example in more local switching offices would exist ensuring that no user is penalised for their location. The data centre is not moved closer to any one group of users, thereby creating a disadvantage to another group of users.

D. Service Mix

Eight services types of residential, mobile and business type are modelled. These services are High Speed Internet (HSI, e.g. email, browsing), Internet television (IPTV, e.g. live television), Video on Demand (VoD, e.g, Content Delivery Network (CDN) or Youtube video streaming), Residential Cloud (e.g. Virtual PC, machineto-machine and other next generation services), Peer-to-Peer (P2P, e.g. file sharing), Mobile Backhaul (e.g.

LTE,UMTS, GPRS, GSM), Business Broadband (employee browsing) and Private Cloud (e.g. corporate cloud, includes private lines). For each service an average bandwidth per subscriber was assumed. The traffic is categorized by destination, meaning that traffic can arrive to the metro access sites from the Internet, from DCs located in different places of the network, local metropolitan area traffic and multicast traffic. Multicast traffic is assumed to be initiated from the DCs and form a separate category for the particular way it is handled in the network. Average traffic breakdown per service over time is shown in Figure 2.4.1-2

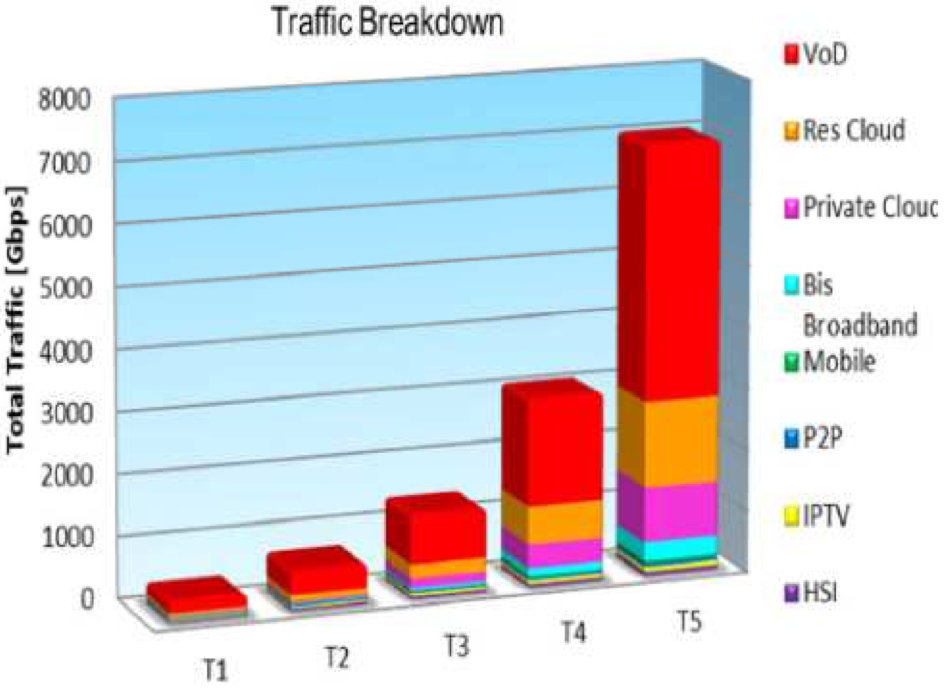


Figure 2.4.1-3: Modelled traffic evolution per service

Each service is also associated a percentage of subscribers. This percentage represents the subscriber take-up, which is also an expression of the popularity of each individual service (see Figure 2.4.1-4). The random values are generated according to a uniform distribution within each individual range.

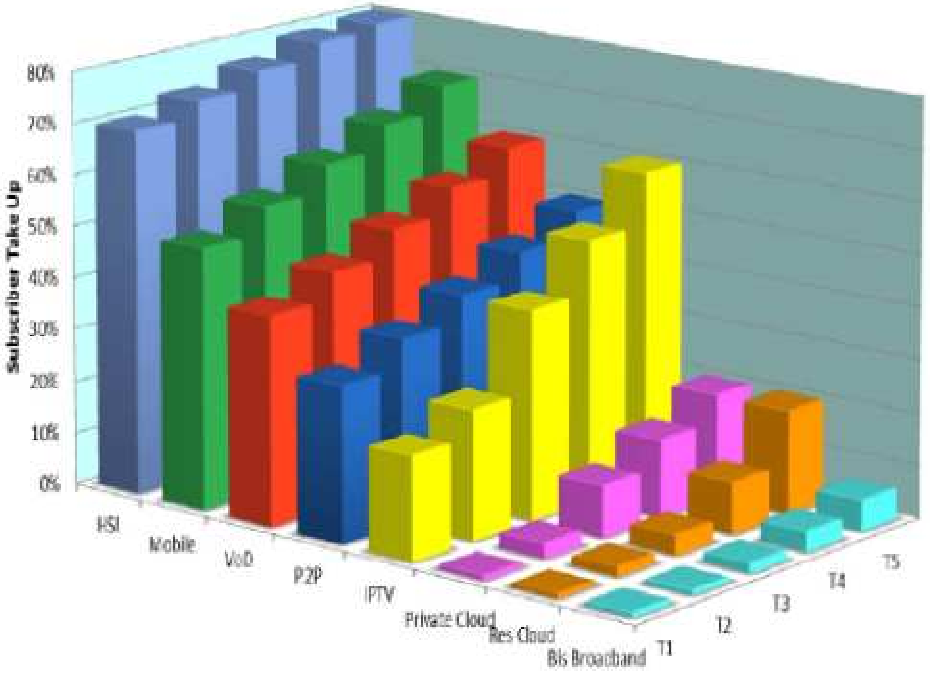


Figure 2.4.1-5: Modelled subscriber take-up evolution per service

Approximately 25 different pricing values were used in the model, but the key comparative pricing is the average cost of a 10GigE OPST port as 1.2 times the price of an average 10GigE IP Router port cost. This relative pricing variable was validated against current industry pricing.

Savings are achieved because of the inefficient use of Router ports per solution, e.g. up to 50% of Router ports in some scenarios are interconnecting Routers together rather than delivering services.

Technoeconomic analysis

The most important results of this study are related to the scalability of a network solution as a function of data centre (DC) location. Figure 2.4.1-6 presents the CapEx (i.e., capital expenditure) of all meaningful solutions for DC and switching capacity locations. The most frequently used solution today is represented by the “IP Core – DC in Core” curve, whereby most of the traffic is backhauled to core routers to which the DCs are also connected. Most often the DCs are located in a remote network in the Internet. From a router port requirement perspective, however, this is similar to having the DCs located in the national core network. For this reason these subcases are addressed together.

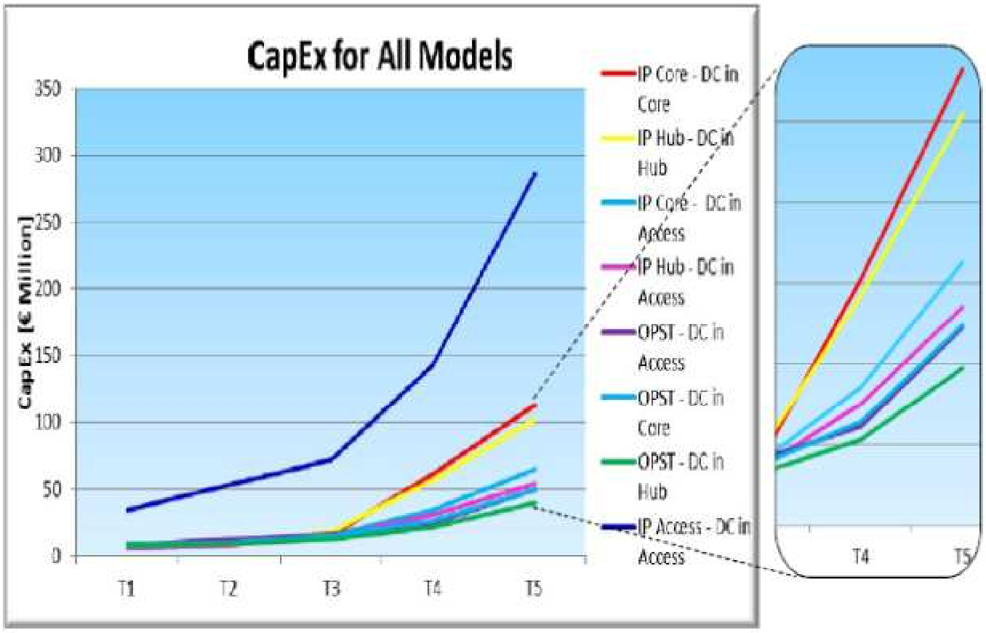


Figure 2.4.1-7: CapEx for all variations of DC and switch capacity locations

In the “IP Core – DC in Core” solution huge routing and grooming capacities are built in the core network. Traffic collected from client interfaces on the client edge of the metro network is aggregated by further switches and/or routers and then groomed in the core to ports dedicated to different services operated on DC servers. This is the most expensive solution (apart from “IP Access – DC in Access”, presented later) due to the large number of expensive core router ports required for routing and grooming the traffic. Such high costs fall on the network operators, while at the same time the revenues obtained from content services (e.g., VoD, social media) go to the content providers (i.e., not to the network operators). This asymmetry of the cash flow has a negative impact on the profitability of the network operators. To address the cash flow problem, network operators are working on bringing the service platforms into their networks, either by building their own service platforms or by providing premium services for content providers through creating local distribution points for third-party content. Also, by moving the content closer to the enduser, a significant amount of transit traffic is eliminated from the core network, while the service performance (e.g., delay, jitter) improves. This can be achieved by placing DCs in the hub and switching the bulk of the traffic in hub routers (see “IP Hub – DC in Hub” in Figure 2.4.1-8). However, router port requirements remain relatively high. Therefore, although the performance and cash flow symmetry improves, the solution CapEx is still high.

Further improvement on the CapEx and/or performance can be achieved by moving the DCs even closer to the end-user through distributing them out to metro access sites. Interconnection of the sites can be achieved by performing the routing function a) in the core (“IP Core – DC in Access”) or hub (“IP Hub – DC in Access”); or b) by building out direct optical connections (i.e., optical mesh) between the metro access sites (“IP Access – DC in Access”). As shown in Figure 2.4.1-9, these different options have significantly diverging CapEx implications. Although IP core and hub solutions achieve important CapEx reductions, all traffic is switched in the core or hub, respectively, which implies multihop connections and, hence, higher latency and jitter. On the other hand, the “IP Access – DC in Access” full optical mesh solution provides single-hop optical connectivity between the metro access sites, which implies the lowest possible delay and jitter. However, such optical connectivity requires a very expensive optical layer as well as a high number of L2/L3 ports. This provides by far the most expensive solution (see Figure 2.4.1-10). Therefore an important decision needs to be made in the IPoROADM solutions between providing high performance or low cost as high performance at low cost is not possible in these networks.

The collapsed switching and optical layers of OPST provides a solution to this cost/performance problem. Figure 2.4.1-11 shows that OPST provides the lowest CapEx network independently of the location of the DC. The “OPST – DC in Core”, “OPST – DC in Hub” and “OPST – DC in Access” curves provide significant CapEx savings for all these scenarios, which makes OPST a durable solution that scales cost-effectively with the increasing traffic demand. The reason for this is that OPST is a distributed switch based on a full and dynamic optical mesh of connectivity. This uses shared optical paths between ports allocating resources in real time to where and when they are needed, thereby optimizing both switching and transport port costs simultaneously. Furthermore, the OPST switch’s utilization of this full mesh of direct optical connectivity ensures single-hop forwarding between ports thereby inherently minimizing latency.

Thus OPST supports services that require stringent performance guarantees in a cost effective manner. CapEx advantages are of the order of 150%, e.g. the IPoWDM solution is 1.5 times more expensive the DC is located in the core or hub (see 2.4.1-12). When the DC is distributed in the metro access, IPoROADM solutions can close the CapEx gap with OPST but this can only be achieved by compromising performance (“IP Core – DC in Access” and “IP Hub – DC in Access” in 2.4.1-13).

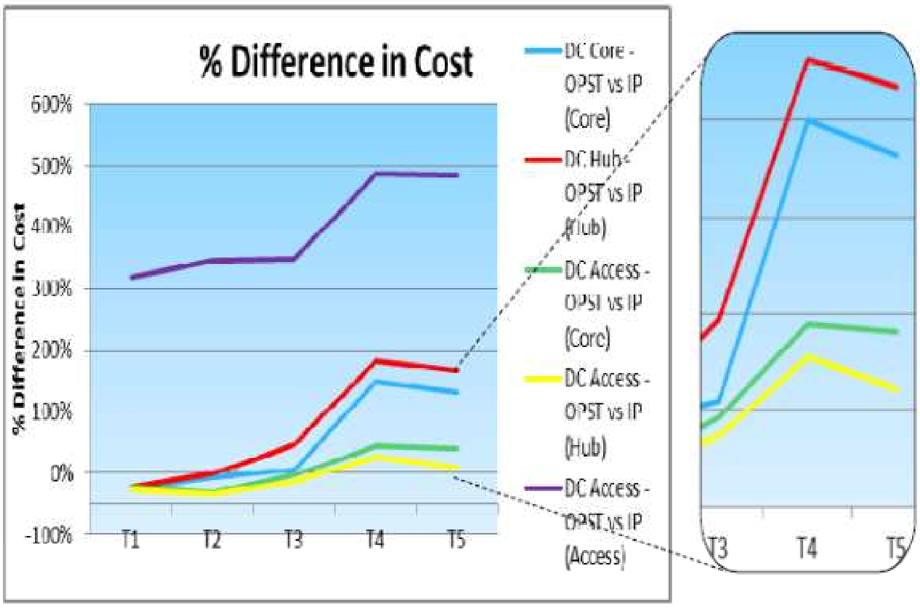


Figure 2.4.1-14: Relative CapEx for all models

However, as the importance of delay and jitter guarantees increase, the CapEx moves higher up toward the “IP Access – DC in Access” curve of 2.4.1-15, being bound by a configuration that uses a full mesh of fixed optical connections between the metro sites. The CapEx of this latter solution is almost 500% higher (5 times) than that of the OPST one.

As an ancillary output of the modelling performed, the power consumption (Figure 2.4.1-16) and rack space (Figure 2.4.1-17) requirements were also quantified. These two important operational expenditure (i.e., OpEx) drivers display similar trends to that of the capital costs. At scaled traffic requirements the power consumption advantages provided by OPST are at the order of 100%-300% for the most of the scenarios. Also in this case, if OPST performance is emulated over an IPoROADM solution this latter consumes up to 800% more power.

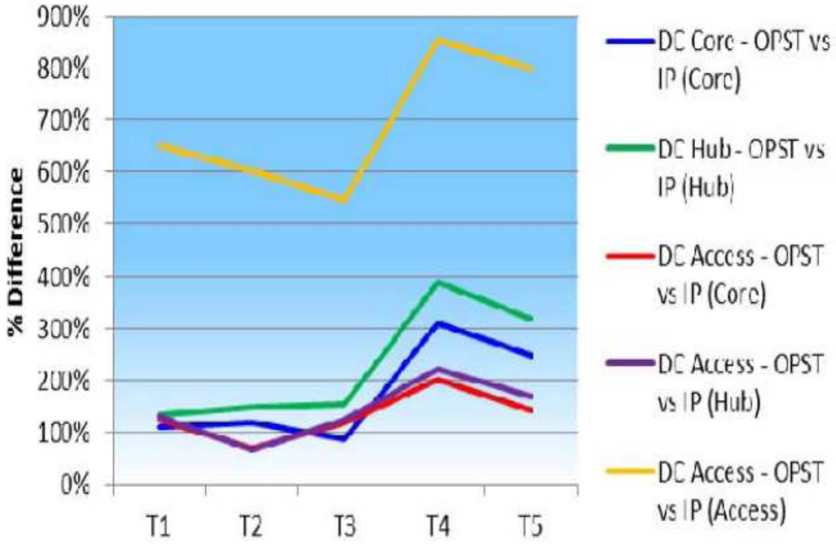


Figure 2.4.1-18: Relative power consumption. Positive percentiles indicate that OPST enables power savings compared to the different IP0DWDM configurations.

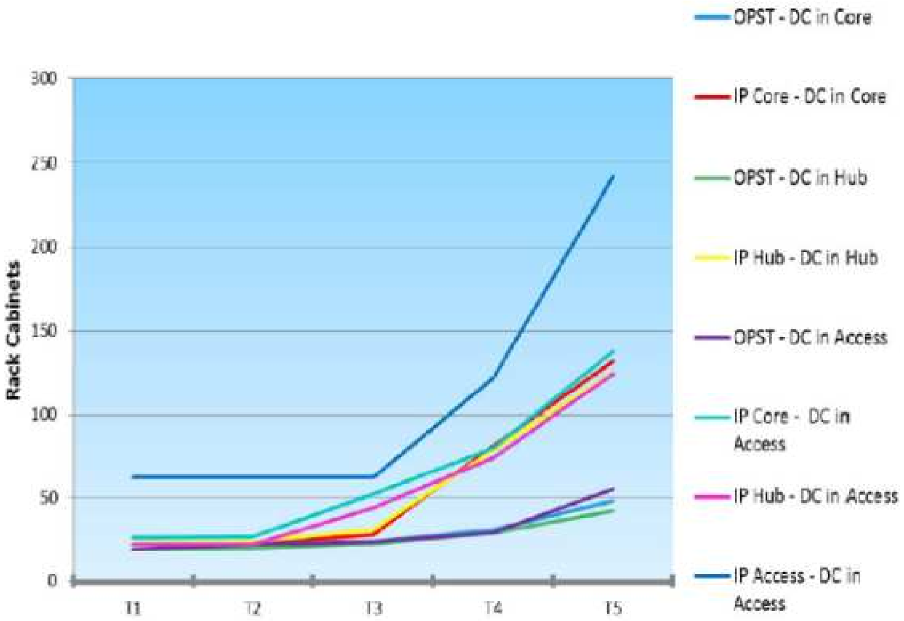


Figure 2.4.1-19: Rack space requirements. The 3 OPST solutions are the 3 bottom curves, showing lower rack cabinet requirements independently from the scenario configuration

The rack space requirement gap between the two compared solutions is also quite significant at scale (i.e, at T4 -T5). With the DC in the core and hub and routing capacity focussed either to the core or hub, the gap gets up to 80 rack cabinets at a total of 130. This translates into 150% more cabinets used for the IPoROADM solution. In the case of an optical mesh in the metro network that connects a distributed DC (to emulate OPST performance) up to 500% more rack cabinets are needed for the IPoROADM solution. Thus, beyond the CapEx savings, significant OpEx reductions can also be achieved with a sub-wavelength optical packet switching technology.

This analysis is based on a chosen static traffic pattern upon which a staged growth pattern is imposed over 5 time periods. An alternative approach using random variations on the traffic patterns is presented elsewhere [16], using a Monte Carlo simulation [17] to gain visibility into impact of variable network conditions on the economic performance. The results presented in that simulation further validate and support the findings of the work presented here.

2.5. Intra-Datacenter Networks

Datacenters are the main structures to host server side application and services, enabling the rapid growth of cloud based storage and computation. Modern datacenters consist of thousands of servers as individual computing units, which are arranged in racks and clusters in massive scales [19,20]. These servers are effectively networked to provide scalable storage requirements and computational systems over distributed platforms.

Intra-datacenter [20] network interconnection provide high data rate lines preferably using optical fibres among the servers to enable bandwidth aggressive applications with tight delay and jitter requirements of parallel and distributed computing in massive scales.

Support for any to any server communications with very high data rates and dynamicity is the ideal interconnection pattern for intra-datacenter communications [20]- The requirement of high connectivity among servers is currently approached by deploying hierarchies of L2/L3 switches in different levels within and between datacenter racks and clusters [21]. This leads to deployment of ample number of fiber links and connections for intra-datacenter networking with not much concern about spectral efficiency [20].

However the non efficient use of links and spectrum inevitably incurs higher operation and maintenance cost in the longer term.

MAINS subwavelength architectures such as OPST and TSON are perfectly adapted to the connectivity requirements for intra-datacenter networking, since they introduce higher efficiency while preserving the effectiveness of the communications.

This section in particularly focused on a line card specifically designed in MAINS for intra-data center networking.

This design implements FPGA based network onchip reconfigurable line card supporting 10GEthernet and TSON Tx and Rx. The line card operation of Ethernet or TSON is supported by the flexible and reconfigurable Architecture on Demand (AoD) [24] optical transport layer built as the network off-chip. It can support high capacity (e.g. > 5 Gbps per 10G NIC) and ultra-low latency (<10 μs) services on Ethernet over WSON or lower capacity (< 5 Gbps) services to be statistically multiplexed over TSON.

Network on and off Chip (NooC)

The novel concept of the NooC design addresses the rigid and static network configuration of intra-datacenter interconnections. As displayed in Fig.2.5 1(a) (right side), the proposed solution presumes installing the Network-on-Chip line cards capable of Ethernet and TSON (sub-wavelength) on the servers (or PCI-E to Ethernet and TSON), and replacement of a typical Ethernet rack switch with a more flexible all optical switch supporting both (sub) wavelength switching for TSON and Ethernet, which is built as Network-off-Chip solution. The Network-on-Chip design is implemented using a High Performance Xilinx V6 FPGA, which operates as a bi-functional line-card supporting 10GEthernet, and 10G TSON transport interfaces with up to 8.6 Gbps throughput. It enables hitless switching between Ethernet and TSON operation, so the server can deploy either upon request with no interruption (no packet loss) to the running services. Using TSON the network interface card can transmit sub-wavelength data employing time-shared statistical multiplexing; it can support as low as 100Mbps with equal step size over 10G links, offering highly efficient traffic aggregation. Using Ethernet transmission, the server will have a dedicated wavelength to transport its data over wavelength switched optical network (WSON).

On the other hand, the off-chip design provides flexible node and network architecture, following the AoD approach. In AoD, an optical backplane, e.g. 3D MEMS switch, interconnects a pool of components and enables reconfiguration of the transport layer by configuring internal cross connections. The coordinated operation of the on and off chip networking enables bandwidth flexible, efficient and transparent cross connection within datacenters.

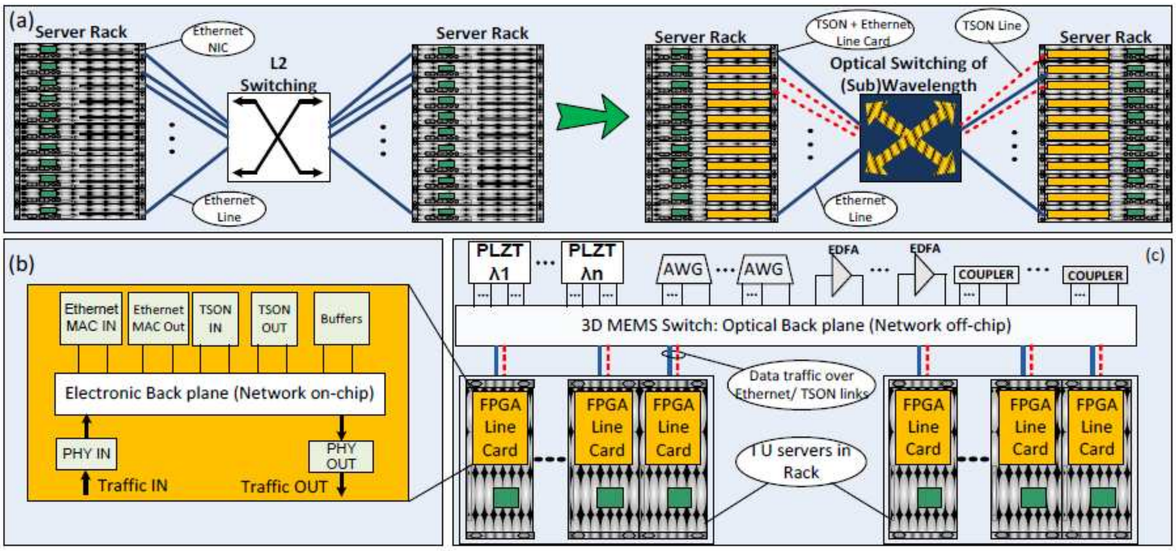


Figure 2.5-1: (a) Network on-off chip for intra datacenter, enabling transparent Ethernet and TSON transport. (b) Network on-chip implementation with modules attached to an electronic back plane. (c) Network off-chip where optical components are selected for Ethernet or TSON connectivity paradigms.

Network on-Chip design

Network-on-chip design is an approach to design the communication subsystem between sub-modules. Several sub-systems for separate functions of Ethernet RX, TSON RX, Ethernet TX and TSON TX have been developed in a modular manner. Using a deployed internal electronic backplane switch, each of the implemented sub-systems can be selected to enable the desired service. Different sub modules of the network on-chip are displayed in Fig.2.5 2(b). The same figure illustrates how with a select-and-place approach using different subsystems of TSON TX and RX, Ethernet MAC and buffers, the line card is able to operate as Ethernet to TSON or Ethernet to Ethernet interface. The implemented on-chip design allows a hitless switch over between TSON and Ethernet modules, by incorporating Ethernet or TSON related blocks in the operation. The electronic switch over is dynamically controlled by a Server that is able to update the to FPGA mode (ETH or TSON) LUT. The selection of any of the combinations, for example Ethernet-TSON-Ethernet, is supported by the reconfiguration of the transport plane as well.

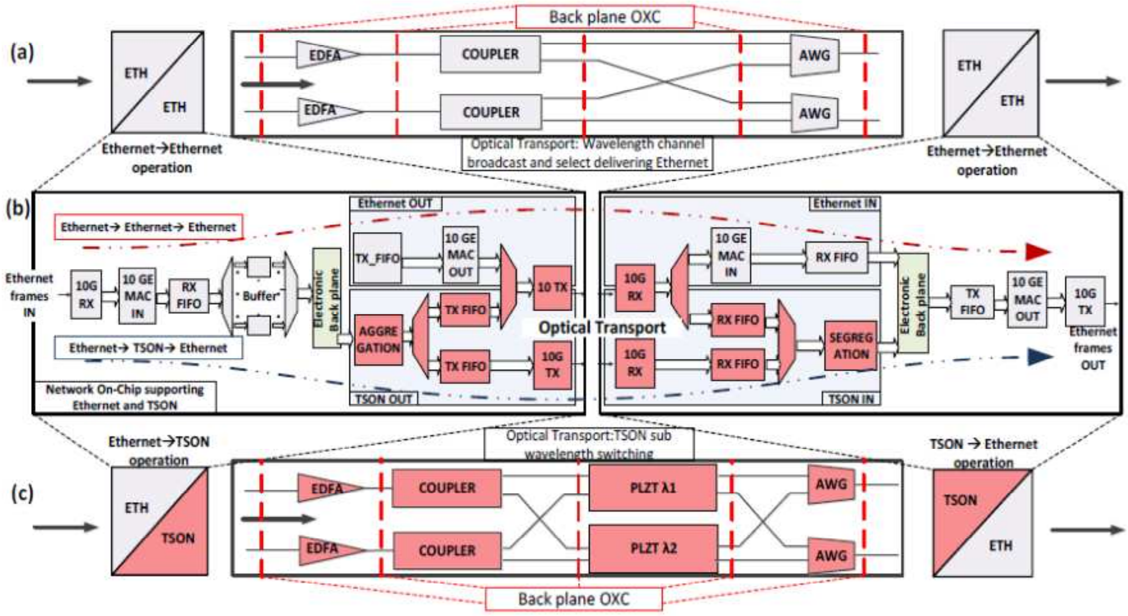


Figure 2.5-2: (a) network off-chip for optical transport of broadcast and select delivering Ethernet (b) network on-chip for building TSON and Ethernet interfaces (c) network off-chip for TSON operation

The Network-off-Chip is implemented with an AoD node [25] which is able to implement the required optical node architecture for Ethernet transport, or to re-configure it to support TSON aggregation and switching in 20 msec (optical backplane reconfiguration time). For transparent transport of Ethernet channels, wavelength switching can be adopted using broadcast and select topology by choosing couplers and AWG components and the back plane in-between as in Fig.2.5 2 (a). For TSON operations fast optical switches (PLZT) need to be involved as well, as shown in Fig. 2.5 2(c), in order to multiplex time-slice data sets from different servers and map them on specific time slices over one or few wavelengths. Fibre switching is also available by using the ports on the back plane MEMS switch.

Performance analysis of MAINS chipset for intra-datacenter networking

The FPGA-based network on-chip experimental results were measured using the Anritsu MD1230B data analyzer and shown in Fig.2.5 3. In the top 2.5.3 figures, at point A the operation mode is switched from TSON to Ethernet, and at point B it is switched from Ethernet to TSON. Fig.2.5.3.a shows the measured throughput result when switching between TSON and Ethernet (Input frame size 64B). The bit rate step difference is because of the limited TSON maximum throughput. Fig.2.5.3.b shows how the received bit rate changes when switching between TSON and Ethernet. When switching from TSON to Ethernet, the bit rate increases for a short duration because the data in the TSON aggregation FIFO is transmitted immediately without waiting for a time-slice allocation. Also, when switching from Ethernet to TSON, the bit rate decreases for a short duration as it is necessary to wait for at least one time-slice to store the data for aggregation. Fig.2.5.3.c shows that the number of received frames matches the number of transmitted frames, which proves the hitless feature of the design. Fig.2.5.3.d shows the theoretical and experimental Ethernet and TSON throughput. The TSON delivers 87.96% of the maximum measured Ethernet throughput (8.68 Gbps). Fig.2.5.3.e shows a comparison of the measured ultra-low latency results for Ethernet (<6.7 μs) and very-low latency for TSON (<160.6 μs). TSON latency depends on the transmitted bitrate due to the delay introduced by data aggregation and time-slot mapping. Experimental jitter results for TSON are shown in Fig.3.f. 87.5% of the data transmitted in TSON is received within 2μs. For Ethernet, with input stream of 1500B frame size, 100% of the frames are received in 2 μs.

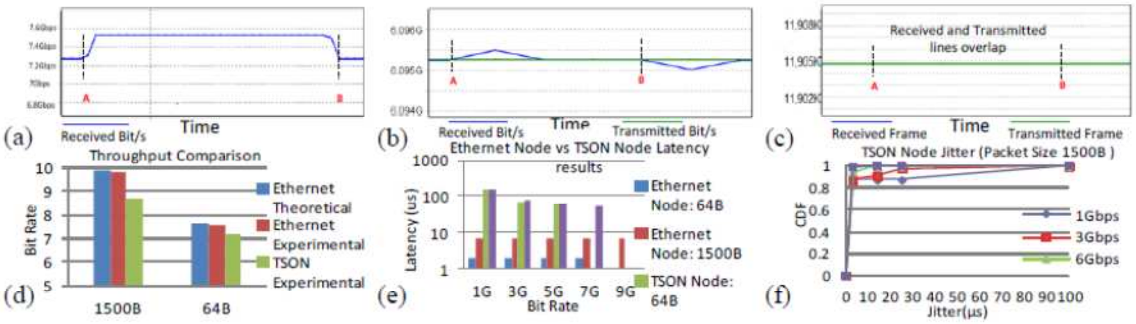


Figure 2.5-3: FPGA-based TSON-Ethernet switching NoC results. a) Throughput, b) received bitrate, c) received frame changes; d) Throughput comparison; e) Ethernet node vs TSON node latency comparison; f) TSON node jitter.

2.6. Cost effective Virtual PC services

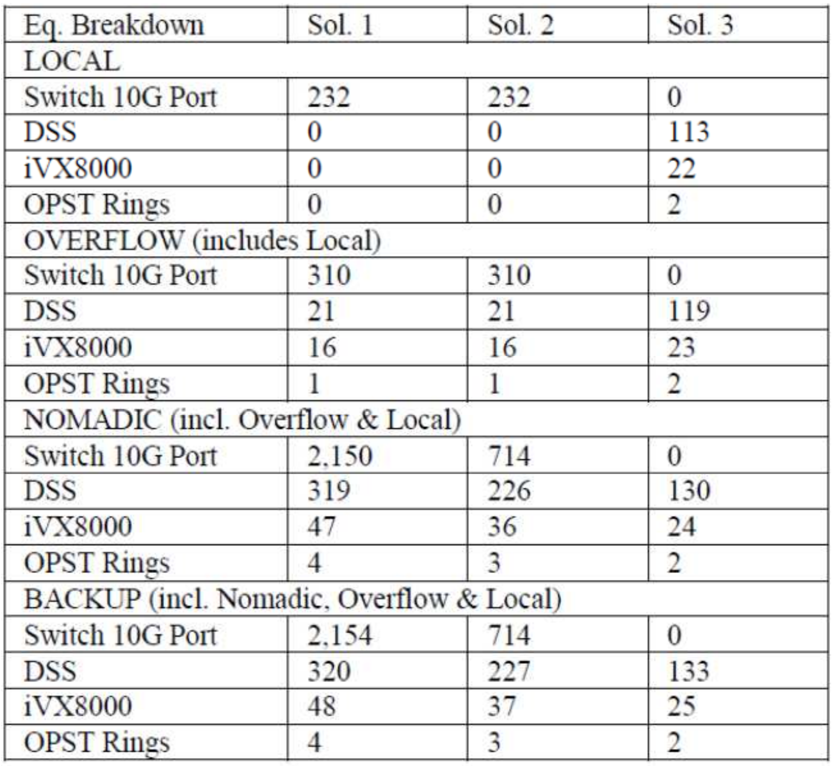
As demonstrated in D1.2 and [3], MAINS architecture is specially adapted to nomadic Virtual PC applications requirements. Three network solutions for supporting a next-generation Virtual Personal Computer service were modeled:

- Solution 1: Local Switch, OPST Ring and Relocation of User Data to the closest server
- Solution 2: Local Switch, OPST Ring and Direct Remote Connection to a centralized server
- Solution 3: No local switch, OPST Ring and Direct Remote Connection to a centralized server

The equipment breakdown for forecast period 2013 for the different traffic types and the 3 solutions studied is presented in Table II. With local traffic only, Solution 1&2 require 232 10G ports each on the local switches. Since there are no local switches in Solution 3, local traffic is also handled by OPST equipment.

Although overflow traffic increases Solution 1&2 equipment requirements significantly compared to Solution 3, the major difference is caused by the nomadic traffic that affects the switch port requirements for Solution 1 to increase 7 times while the number of DSS modules grows 15 times. This is due to the fact that to move the huge amount of user data to the visiting location in a short time will require a lot of bandwidth.

TABLE II. CUMULATIVE EQUIPMENT REQUIREMENTS PER TRAFFIC TYPE FOR 2013



Eq. Breakdown	Sol. 1	Sol. 2	Sol. 3
LOCAL			
Switch 10G Port	232	232	0
DSS	0	0	113
iVX8000	0	0	22
OPST Rings	0	0	2
OVERFLOW (includes Local)			
Switch 10G Port	310	310	0
DSS	21	21	119
iVX8000	16	16	23
OPST Rings	1	1	2
NOMADIC (incl. Overflow & Local)			
Switch 10G Port	2,150	714	0
DSS	319	226	130
iVX8000	47	36	24
OPST Rings	4	3	2
BACKUP (incl. Nomadic, Overflow & Local)			
Switch 10G Port	2,154	714	0
DSS	320	227	133
iVX8000	48	37	25
OPST Rings	4	3	2

According to our analysis, CAPEX requirements for MAINS proposed architecture are 20% of the cost of the current network architecture approach, revealing a new and counter-intuitive network design.

3. Industrial performance validation

A test bed resembling a realistic deployment scenario was set up at TID premises in order to validate the OPST platform as an infrastructure for Telco services delivery.

The key features of the platform that were verified are the following:

- Distributed Bandwidth Allocation (DBA) and fairness.
- Rapid Response to service demands at the demand point.
- Operational flexibility at the core.
- Simultaneous grooming and aggregation.
- Flexibility of Core location within metro.

The above mentioned capabilities allowed to demonstrate the feasibility of deploying multiple services over the OPST network: Virtual Machine Transference, Video Streaming and Internet traffic between others.

3.1. System configuration

The system configuration consisted of a four node, dual plane ring (A, B), where up to ten end-point (10 GbE) can be flexibly distributed according to every test needs.

Next figure below shows a high level schematic of a setup with 5 end-points on the ring.

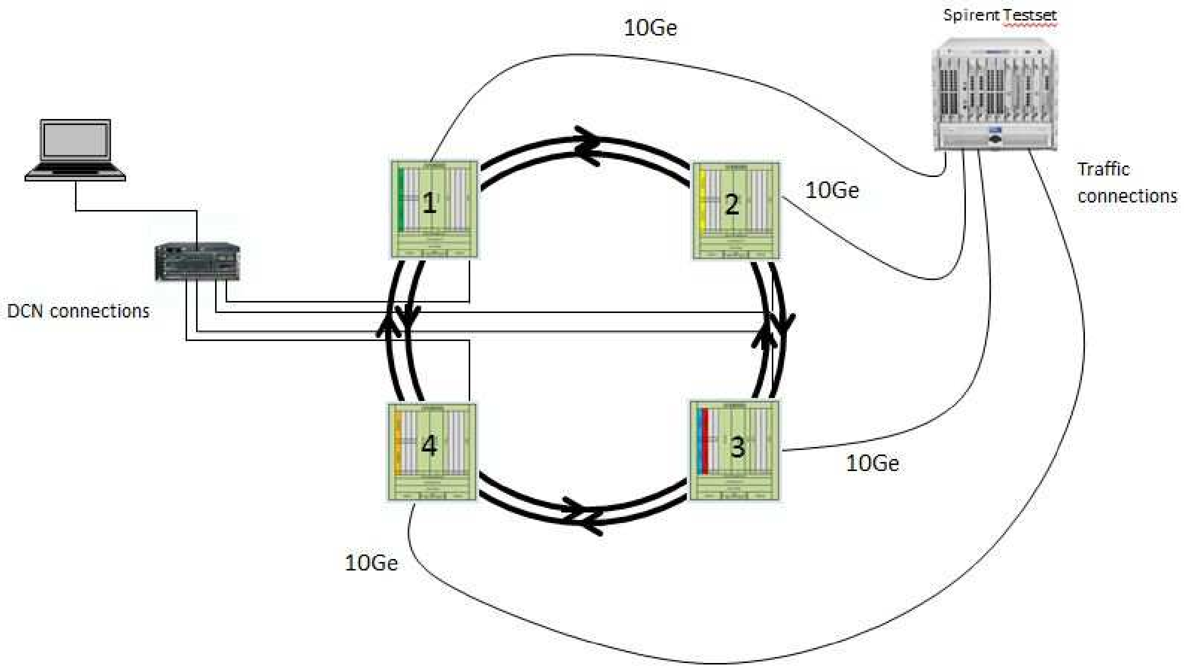


Figure 3.1-1: System configuration

Under normal conditions the system is working on the “A-plane”. Fibre spools of different lengths were used between every pair of nodes for the “A-plane”, with corresponding dispersion compensation fibres at the inter-stage of the EDFAs used for that ring direction. As for the “B-plane”, short length fibre cables were used for node interconnection (i.e “0” distance between them) and thus no compensation fibre modules were installed.

The following tables show the node distances for every plane.

A-Plane

A-End	Z-End	Distance (km)

Node 1	Node 2	40
Node 2	Node 3	55
Node 3	Node 4	30
Node 4	Node 1	45

B-Plane

A-End	Z-End	Distance (km)
Node 1	Node 4	0
Node 4	Node 3	0
Node 3	Node 2	0
Node 2	Node 1	0

A picture showing the four OPST nodes (two nodes per rack) and the fibre spools used for the ring links is included below. An Spirent Test Center platform with ten 10GbE interfaces was used to emulate different traffic conditions over all possible source-destination pairs.

Figure 3.1-2: Installed setup

3.2. Traffic and service tests

The platform provides EVCs (Ethernet Virtual Connections) from an ingress port to an egress port of the ring, located on the same or on different nodes. Each EVC may have up to 4 traffic classes. Each traffic class has a set of parameters associated with it that may be provisioned when the EVC is set up. These are:

- Committed Information Rate (CIR) in bits per second
- Committed Burst Size (CBS) in bytes
- Extended Information Rate (EIR) in bits per second

- Extended Burst Size (EBS) in bytes

In addition the system is configured with a bounded delay for each traffic class. These delay bounds are given in the table below.

Traffic Class	Delay Bound
TC7	5ms
TC5	10ms
TC3	15ms
TC0	N/A

Traffic classification is based on the priority code bit (PCP) carried on the Ethernet frames. The default PCP to Traffic Class mapping is as follows:

- PCP7 -> TC7
- PCP5 -> TC5
- PCP3 -> TC3
- PCP0 -> TC0

Traffic classification and treatment

These tests aimed at verifying the current behaviour of the system regarding bandwidth allocation and resource control at both provisioning and “under operation” stages.

At the provisioning stage, it was verified that the system consistently allows (or denies) resources for the services according to the overall capacity and considering TC delay bounds and flow parameters (CIR, CBS, EIR, EBS).

On the other hand, “under operation tests” start with a valid EVC configured to have all 4 traffic classes, each configured with a different set of parameters. The EVC was tested with a variety of flows that have burst size and bandwidth specified for each traffic class. The results were according to expected:

- Delay bounds were respected for classes TC7, TC5, TC3.
- Policing behaviour at the Ethernet size (guaranteed CIR/CBS, remarking up to EIR/EBS, dropping above EIR/EBS) worked as expected.
- Dropped rates for the different TC classes whenever aggregate demand is above maximum system capacity are consistent: TC0>TC3>TC5>TC7.

Service rollout

These tests were done to emulate several network configurations where some specific nodes are designated as service head ends (or aggregators) an some others as edge or access nodes. A different number of EVC are established along the ring, with a number of end-points ranging from 5 to 10.

For the initial service configuration, 5 end-points (NPC) were deployed, with one of the nodes acting as ring head end. From the head end, several flows were configured and different profiles (Traffic class and parameters like CIR, CBS, EIR…) were assigned to each one. The following figure shows the initial configuration, where vlan identifiers (or EVCID) like "1x0" were used for "best effort" traffic flows while "x0" vlan identifiers were used for all the other profiles.

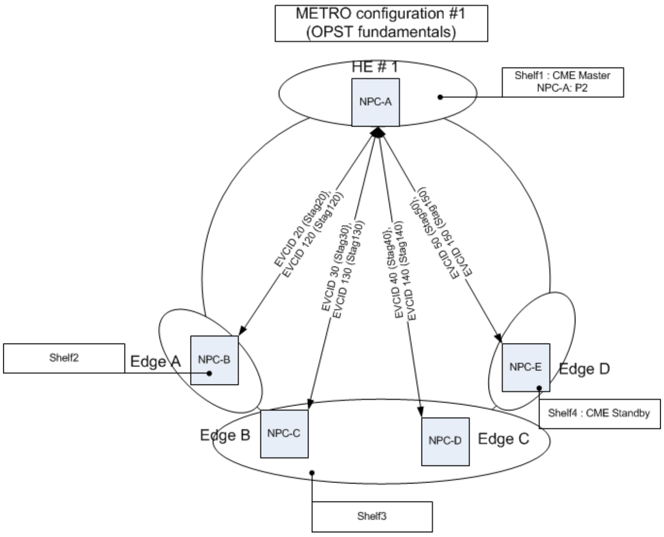


Figure 3.2-1: Initial service configuration

Further service configurations were validated in order to check some additional scenarios:

- With two ring head-ends in response to a relevant increase on a service demand.
- With three ring head-ends, to emulate a situation where service demands go even higher.
- With three ring head-ends and another end-point acting as a cache node, like in a typical CDN scenario where a popular service is moved closer to the end-user.

The total number of configured flows across the ring keeps increasing with every metro configuration, thus stressing DBA procedures.

The following pictures show some of the service configurations that have been tested. The system behaviour was as expected and no issues or limitations were observed in any case.

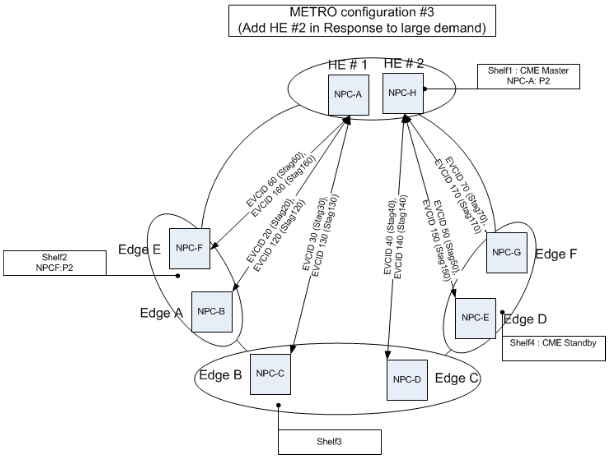


Figure 3.2-2: Two head-end scenario

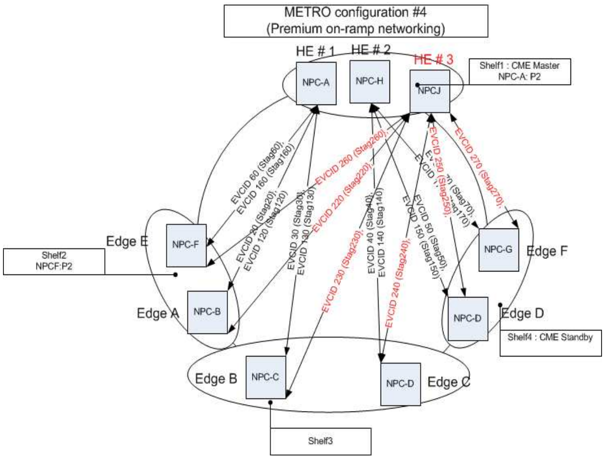


Figure 3.2-3: Three head-end scenario

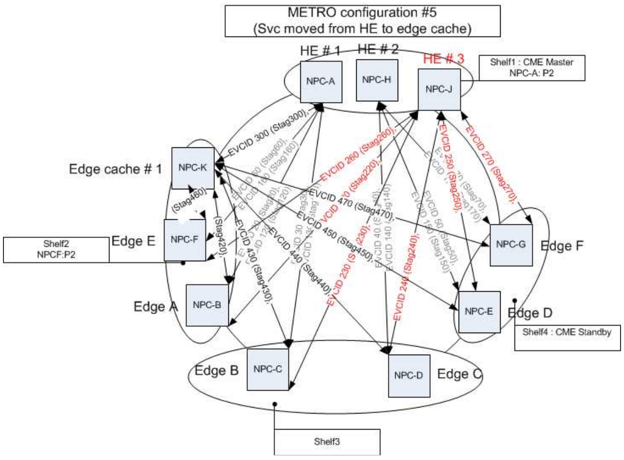


Figure 3.2-4: Three head-end scenario + caché

3.3. Service Integrity tests

These tests take a special relevance for Telecom operators, as they need to fully ensure network reliability and stability under common operations.

First set of tests deals with system behaviour under different card removal/reinsertion/fail situations, focusing both on the impact on traffic and the detection of the failures through the raisin /clearing of system level alarms. Obtained results were satisfactory in every tested scenario, recovery times being bound (whenever applicable) to around 30 ms. It was also shown that a redundant connection can be configured for ring management purposes, whereby a connection to one specific node can be designated as “active” and a different one as “backup” for the whole ring.

A complete failure of one node (power down) was tested, verifying that affected services are switched to the alternate plane (B) and the flows originating/ending on the node are stopped. Whenever the node is powered up again, after some restart interval the flows (going into and out of the node) become active and all the ring flows switch back to the configured (A) main plane.

A second set of tests deal with the Bi-directional Path Switched Ring (BPSR) protection scheme of the system. Recovery times were in the order of 30 msecs.

Finally, the system was stressed by configuring the maximum number of EVPL services supported by the system (4000) and carrying traffic over all of them, with an overall load below the one that can be accommodated on the ring.

Service integrity test results were satisfactory, no issues were found at the provisioning stage and no frame drops were observed.

4. Conclusions

Industrial application scenarios

Scalable multiservice networks

Current all-IP architecture costs depend significantly on traffic growth: the higher the traffic, the higher the network costs. Consequently, any cost increase is reflected on the service provider’s margins. MAINS analysis demonstrates that IP off-loading over multi-granular photonic switching technologies enables up to 42% CapEx savings. Such savings are achieved through the removal of most transit routers from the network as their routing, aggregation and grooming functions can be fulfilled in the multi-granular photonic network. Moreover, the case study discusses other benefits enabled by the proposed architecture, specifically, significant cost reductions for video and cloud services.

An all-IP solution requires a high initial investment because the common equipment dominates the CapEx, whereas the CapEx of the multi-granular OPST solution follows the actual traffic demand in the network significantly closer. Consequently, OPST proposes a future-proof network infrastructure investment solution as it yields greater benefits as more video, mobile and next-generation cloud services are added to the network.

Next-generation network design is always based on a series of assumptions about user behaviour and the resulting traffic demand. However, past experience shows that user behaviour forecasting implies a significant level of inaccuracy and, hence, risk for network infrastructure investment. MAINS studies were focused on the quantification of this risk through economically modelling two nextgeneration network architectures, based on IPoDWDM and sub-wavelength packet switching. The results show that the flexibility of a sub-wavelength packet switched network reduces the risks stemming from forecasting uncertainty by 500% for 99.8% of the 1 million uniformly distributed configurations compared. Moreover, the sub-wavelength solution reduces capital expenditure by 150% on the long term. This demonstrates that there are indeed ways for using optical packet switching efficiently for building cost-optimized, futureproof next-generation networks.

CDN optimization

A comparative analysis between H-VPLS and OPST technologies for video streaming services has been performed. A densely populated suburban scenario, where CDN delivery nodes serve high quality video to a realistic number of residential subscribers has been designed for the study.

The aim is at finding the overall (IT and network) cost and power consumption obtained with both technologies/architectures. The analysis shows there are relevant cost benefits when an OPST architecture is deployed (up to 42% cost saving with respect to the most economical H-VPLS based solution), which also brings additional advantages such as a lower usage of fiber resources or smaller restoration times. Although no power consumption advantages have been obtained from the study (power figures used correspond to the very first OPST implementation), up to a 20% improvement on power consumption is expected in a further release. This means that OPST technologies can potentially reduce energy expenses, optimizing overall (capital and operational) network costs.

Intra and inter-datacenter networking

Optical sub-wavelength packet forwarding technologies scale cost-effectively under increasing traffic requirements and provide a future-proof solution for next-generation services and networks independently from the future location of data centres in the network. Finally, services with stringent performance requirements are supported much more cost effectively (in terms of CapEx, power consumption and rack space) and without compromise when compared to IPoROADM technologies.

Besides metro network architectures, MAINS is also proposing a novel Network on-andoff Chip paradigm for intra-datacenter optical networks. The proposed approach uses programmable linecards capable of operating in 10GEthernet or in TSON sub-wavelength switching mode and a flexible optical switching/transport layer based on AoD to perform wavelength switching (for Ethernet) or wavelength and time-slice switching (for TSON). Experimental results demonstrate hitless switchover between Ethernet and TSON, very high throughput for TSON (8.68Gbps - 87.96% of maximum Ethernet), ultra low latency (<6.7 μs for Ethernet and <160.6 μs for TSON) and <2 μs jitter for both.

Industrial Performance analysis

MAINS platform has been validated in a complex test bed at Telefónica I+D premises, resembling realistic network scenarios and emulating service deployment evolutionary steps. Operational issues dealing with both network creation and service provisioning have been checked, while the performance values obtained have been according to expected for every service/s configuration. Platform resiliency has been extensively tested as well, in order to ensure an adequate response under most common failure scenarios is obtained.