Optical Networks for Cost-Efficient and Scalable Provisioning of Big Data Traffic

Krzysztof Walkowiak  
Wrocław University of Technology, Wybrzeże Wyspianskiego 27, 50-370 Wrocław, Poland,  
krzysztof.walkowiak@pwr.wroc.pl (Corresponding author)

Mirosław Klinkowski  
National Institute of Telecommunications, 1 Szachowa Street, 04-894 Warsaw, Poland

Michał Woźniak  
Wrocław University of Technology, Wybrzeże Wyspianskiego 27, 50-370 Wrocław, Poland

Wojciech Kmiecik  
Wrocław University of Technology, Wybrzeże Wyspianskiego 27, 50-370 Wrocław, Poland

Abstract

This article shows how recent advances in optical networks can be utilized to improve big data processing by cost effective and scalable provisioning of high-bandwidth connectivity for big data traffic in backbone networks and consequently tackle the current problems related to big data processing in distributed environment including cloud computing. We focus on two optical technologies, namely currently the most popular Wavelength Division Multiplexing (WDM) and a new emerging approach of Elastic Optical Network (EON). The performance metrics related to both: optical network resources (network CAPEX/OPEX cost, power consumption optical spectrum usage) and computing resources (response time and OPEX cost) are investigated in pan-European and United States networks based on the example big data application related to hyperspectral image processing. Results indicate that provisioning of big data traffic directly in the optical layer with the use of the EON concept provides the best results in most of examined criteria.

Keywords: big data, optical networks, cloud computing

1. Introduction
Recently, we have been experiencing an explosion of interest around big data both in the commercial and in the research communities. According to many experts, it is expected that analyzing big data will become a crucial competition, underpinning new waves of productivity growth, innovation, and consumer surplus. Gartner defines big data as “high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” \(^1\). It should be underlined that big data cannot be perceived only from the perspective of the size of an individual data set, since it is rather a collection of various types of data that is available online to requesting users from a wide range of different sources including: state and local governments, NGOs (Non-Governmental Organizations), media, companies, healthcare institutions, social networking services (e.g., Facebook, Twitter), research units, etc. The ability to acquire and analyze extraordinary amounts of data has the potential to provide a deep impact on almost every aspects of human life, that is, big data seems to be able to deliver direct benefits in many areas that aid society.

Big data is a very broad concept that embraces of numerous types of data. Some of the applications of big data can be categorized as local what means that results of data processing are needed only locally (e.g., traffic conditions, combat crime, company internal finance data, etc.). On the other hand, there are also big data applications that yield results requested in a much larger area covering countries or continents, e.g., meteorology and environmental data. In this paper, we focus on the latter case and investigate the possibility to use optical networks for cost-efficient and scalable provisioning of big data requests. We especially spotlight a new optical approach – namely Elastic Optical Networks. However, the present-day most popular optical approach – that is Wavelength Division Multiplexing – is also addressed. We consider a common, two-layer network architecture with optical network in the lower layer and packet-switched network in the upper layer. We address and compare two scenarios of big data traffic provisioning in the network: service network provisioning, where the packet-switched network layer provides delivery of the traffic related to big

\(^1\) [www.gartner.com/it-glossary/big-data/](http://www.gartner.com/it-glossary/big-data/)

2
data and optical network provisioning, where the big data traffic is delivered directly in optical layer with the use of short-lived lightpaths. Besides theoretical discussion on both approaches, we report results of simulation run with assumptions following from real networks and big data applications with a special focus on hyperspectral image processing. The performance metrics we address are related to both optical network and computing resources. In the former case, we report results related to CAPEX and OPEX costs and spectrum usage. In the latter case, we present results concerning response time and OPEX processing expenditures based on electrical energy costs. To the best of our knowledge, this work is the first to investigate the potential benefits of using Elastic Optical Network approach to facilitate provisioning of bandwidth demanding big data traffic.

2. Big Data

We are living in a digital world surrounded by various data from numerous sources. Each of enterprises collects huge amounts of valuable data, for which manual analysis is virtually impossible. The market-leading companies realize that smart analytic tools capable of interpreting collected data could lead to business success. Moreover, applying of machine learning techniques to extract hidden, valuable knowledge from the huge, usually distributed databases is supposed to be advantageous. The growing amount of the above repositories has caused a rapid development of distributed data mining tools which are able to interpret data effectively. Huge collections of diverse data available online are currently known under the term of big data. However, it should be underlined, that big data is not just a massive amount of raw data, but rather a group of various information and intelligent processing methods that can provide an added value benefits for all users.

Hyperspectral image processing is an interesting example of a big data application in the context of networking issues. According to J.P.Ferguson “If a picture is worth 100 words, a hyperspectral image is worth almost 1000 pictures”2. Hyperspectral imaging is nowadays a fast-growing and promising approach which has numerous potential applications in diverse areas, such as agriculture, food processing, mineralogy, chemical imaging etc. Hyperspectral sensors are capable of collecting

---

2 J.P.Ferguson, An Introduction to Hyperspectral Imaging, Photonics and Analytical Marketing Ltd
information as a pool of hundreds images which are associated with a particular range of electromagnetic spectrum, for the same area on the surface of the Earth. This spectral high-resolution is necessary to make detailed thematic maps (as deforestation image prepared on the basis of NASA’s Landsat images [Spr2012]) of remote sensing data by means of spectral classification of different materials expected in the sensed scene. This richness of information may even allow efficient content based exploration of remote sensing databases [Gra2012]. The volume of data generated by these multiple efforts is astounding, e.g., Geosynchronous Imaging Fourier Transform Spectrometer (GIFTS) which is the part of NASA’s New Millennium Program [Sha2006], [Hua2001]. The GIFTS uses a combination of Large area Focal Plane Arrays (LFPA’s), and a Fourier Transform Spectrometer (FTS) providing a spectral resolution of 0.6cm⁻¹ for a 128×128 set of 4km foot-prints every 11s [HUA2001]. It is anticipated that the GIFTS Level-0 data rate is about 55Mbps or about 1.5Terabyte per day. Such data stream should be divided and sent to computing systems where the particular hyperspectral imaging algorithms (as detection algorithm [Man2002]) on the packages are run [Liu2011].

One of the fast growing area of hyperspectral imaging is related to UAV (unmanned aerial vehicle), commonly known as drone. The drone controlling applications need very fast response and quick computing systems, e.g., to locate appropriate landing sites [Coombes2013], [Fitzgerald2005] or to analyze the images and build desired models. The first group of applications usually tries to locate the landing point (i.e., to answer the question if the surface of the analysing area is plane enough) by calculating the geographic characteristics of the potential area on the basis of a mathematical model to describe the mapping between actual land plane and the plane in the image captured by drone’s camera [Cesetti2010], [Li2013]. Such an approach requires large calculations, which could be speed up in well-designed computing environment. The second exemplary application is a DroneMapper system, which is a cloud based imagery processing service that turns 2D aerial images into high resolution geo-referenced Digital Elevation Models, Digital Surface Models, Orthomosaic Maps, 3D...
Point Clouds and more. The imagery is processed on the Amazon EC2 cloud computing cluster resulting in quick turn around and high precision.

The enormous amount of data related to hyperspectral image processing needs large computing resources to run the algorithms and obtain the results. In order to provide fast processing with low costs, distributed computing systems like cloud computing or Grids are required. From this perspective, an important challenge emerges, namely: how to efficiently send data to processing nodes (e.g., data centers, clusters) and afterwards to clients with simultaneously minimizing the costs in respect of the required performance metrics like response time. To answer this challenge, the sites involved in hyperspectral image processing or other big data applications must be connected through an optical network, since the optical fiber is the only transmission media capable of transferring immense data on long distances in short times.

3. Optical Networks

Optical fiber networks play an important role in communication networks since they provide an infrastructure for the transport of aggregated IP traffic. Nowadays, optical transport networks are based on the wavelength division multiplexing (WDM) technology and implement wavelength switched optical network (WSON) architectures enabling all-optical transmission and switching of data streams of 10Gb/s, 40Gp/s, and, recently, 100Gb/s rates. Such networks operate within rigid/fixed frequency grids and with single-line-rate transponders making use of single-carrier modulation techniques. Although the WDM technology provides high transport capacities, still the main drawbacks of current WSON architectures are their low spectral efficiency, lack of adaptability to heterogeneous bandwidth demands, and low optical path (lightpath) scalability in terms of carried bit-rate.

During last couple of years, the research in optical networking has experienced significant advances, which among others have been governed by developments of spectrally-efficient modulation techniques and new functional optical components which benefit from both fast electronic signal processing and silicon photonic integrated circuit technologies. It is broadly expected that with the
advent of new networking capabilities and demanding network services the evolution of transport networks will lead towards more flexible mixed-line-rate (MLR) and elastic optical network (EON) architectures [Gerstel2012]. The foreseen evolution path of optical transport networks comprises:

- The application of advanced modulation formats, such as phase-shift keying (PSK) and quadrature amplitude modulation (QAM) in MLR networks, allowing for distance-adaptive lightpath provisioning.
- Elastic access to optical spectrum resources within flexible frequency grids, enabling further bandwidth and bit-rate scalability beyond 100Gb/s.
- The use of multi-carrier modulation techniques, such as optical orthogonal Frequency Division Multiplexing (O-OFDM), allowing for highly-granular and elastic bandwidth provisioning in terms of carrier allocation. In O-OFDM, a high rate data stream is split into a number of low rate data streams that are transmitted simultaneously over a number of subcarriers.

Consequently, the next generation of optical transport networks will utilize network resources more efficiently and, at the same time, they will provide network connectivity adaptively, according to bandwidth demands and transmission path characteristics. Eventually, each of the above listed components will bring the improvement in spectral efficiency and, by these means, higher throughput and better exploitation of network resources.

The main innovation of EON with respect to conventional WSON (WDM) is the provisioning of sub-wavelength granularity for low-rate transmission and super-channel connectivity for accommodating ultra-high capacity client signals within a common network. The EON allows to allocate flexibly appropriate-sized optical bandwidth, by means of contiguous concatenation of optical spectrum, to an end-to-end lightpath and according to traffic demand. There are two key components in EON architectures, namely, bandwidth-variable transponders (BVTs) and bandwidth variable wavelength cross-connects (BV-WXCs). The role of BVTs is to adapt the client data signal to be sent to/received from the optical network with just enough frequency resources. Concurrently, BV-WXCs allow to
create an optical routing path through the network by switching transmitted signals within their frequency bandwidth to appropriate switch output ports. EON architecture implementations and proof-of-concept EON experiments have been reported in the literature, e.g., see [Casellas2012]. To meet the new requirements of flexible spectrum allocation in EON, ITU-T has revised the G.694.1 recommendation and included the definition of a flexible WDM grid.

Big data is closely related to grid computing - in both cases the processing and storage resources are distributed geographically and the network is responsible for providing the connectivity between these sites. Optical networks and, in particular, MLR and EON solutions, have been already appointed to play an important role in grid and cloud applications [Develder2012], [Klinkowski2013]. Similar arguments as in [Develder2012], concerning flexible bandwidth provisioning and scalable networks resources, can be also used in the context of dynamic and bandwidth-demanding big data applications. Additionally, such applications may result in varying and unpredictable patterns of network traffic, both in temporal and geographical domains, due to huge data transfers occurring in random time moments. As a result, a conventional network design approach based on provisioning of permanent lightpath connections operating at fixed bandwidth rates and with bandwidth over-provisioning for accommodating temporal peaks of traffic may be costly and not efficient. Such EON-enabled solutions as bandwidth adaptation of existing lightpaths and dynamic lightpath provisioning were shown to be effective for supporting varying traffic demands and large data transfers [Velasco2013].

4. Big Data Networking

The universal workflow of bandwidth demanding big data processing in optical networks can be described as follows. First, we assume that the client which sends a big data request is located far from the place where the source of (raw) data is placed. Therefore, the connection between those two entities cannot be established by the means of a local area network (LAN) or a metropolitan area network (MAN) and as a result a wide area network (WAN) must be applied. An additional element of the system are computing nodes (data centers) connected to the network that are able to process
the data in a short time and with low cost. We focus on requests that concern data of very large size (e.g., hyperspectral image processing [Liu2011], [Spr2012], genomics [Schadt2010]) and thus the network capacity is the key resource to serve the requests in acceptable time.

In the considered workflow, the input data is transferred from the source node (where the data is created or stored) to a computing node (data center) that processes the data. Next, the output data (results of computations) is sent from the computing node to the client node. If more than one computing node is available, a scheduling system can optimize the allocation of requests to data centers in order to minimize the response time, energy usage, operation costs, etc.

The data processing system – described above – is located on the top of a WAN. We assume a following two-layer network architecture: packet-switched network (PSN) over an optical network (ON). In the lower layer, optical paths (lightpaths) are established over the physical transparent optical network (comprising OXCs connected through optical fibers). The upper layer is based on a packet-switched approach (e.g., MPLS) and the connections (e.g., LSPs) are routed over the logical topology of lightpaths (the virtual optical network). The proposed architecture is relatively simple, however it corresponds to the most popular approaches (e.g., MPLS over WDM) and according to many predictions it will be the prevailing approach of network deployment in near future [Contreras2012].

We assume two possible scenarios of big data traffic provisioning in the considered network architecture. The first scenario – called service network provisioning – assumes that the traffic related to data analysis is carried in the PSN layer using the existing virtual topology of lightpaths, i.e., the network is overprovisioned and some extra capacity is left in order to carry big data traffic according to new requests. Consequently, the lightpath topology is fixed and it is not possible to establish new lightpaths on demand. However, the limited resources of network capacity can significantly increase the system response time to unacceptable values regarding big data applications requiring very fast reaction. The second scenario – referred to as optical network provisioning – enables a dynamic establishing of short-lived lightpaths totally dedicated to carry the big data traffic. In more detail,
when a new request of data analysis arrives, the control plane of the network can decide to establish a new lightpath in the optical domain to connect directly the source node to the computing node and/or next connect the computing node to the client node. This possibility enables transmissions with very high bandwidths (for instance up to 400 Gb/s in EONs) and in consequence should provide very short response time.

In Fig. 1, we show an illustrative example of the network architecture and request provisioning scenarios. Node A.2 stores the source data, node B.2 hosts a data center that provides the processing (computations) and the client that requests the results (output data) is located at node D.2. Service network provisioning scenario is presented in Fig. 1a, while the optical network provisioning scenario is illustrated in Fig. 1b. In the former case, the data traffic between elements of the system (data repository, computing node, client) is provisioned in the PSN with relatively small bandwidth (e.g. 1 Gb/s) what follows from limited resources of overprovisioned capacity. In contrast, in the optical network provisioning scenario, two additional lightpaths are established to serve the big data traffic and the lightpaths can be allocated with up to 400 Gb/s of capacity.
Fig. 1. Scenarios for provisioning of big data requests in two-layer network architecture: (a) service network provisioning scenario (b) optical network provisioning scenario.

In order to positively respond to the new service context following from growing interest on analyzing big data, the networks must meet the following requirements:

- **Throughput** – defined as the volume of data (bit-rate) that can flow through a network expressed in bps. The throughput is an important parameter that in a large extent influences the response time of the big data analysis. To illustrate this issue let’s consider a simple example. The client requests to process input data of 100 GB. Notice that to deliver 100 GB of data, the network needs 2, 8 and 800 seconds for 400 Gb/s, 100 Gb/s and 1 Gb/s connection, respectively. The concept of virtualization of optical networks can be utilized to provide a relatively cost effective approach to build an overprovisioned service network on a top of existing optical network [Pages2012], [Jain2013]. Moreover, the new EON concept provides significant improvements in terms of bandwidth and elasticity of capacity spectrum usage in comparison to currently most popular WDM network [Gerstel2012].

- **Scalability** – the network must allow elastic on-demand delivery of huge amount of data according the changing demands, what requires an automated connectivity control in order
to enable dynamic use of the network resources and to enhance the network configuration. The recent concept of Software Defined Network (SDN) seems to be the key element that is expected to make the networks programmable, easily partitionable, and virtualizable [Jain2013].

- Resilience and security – since some results of big data analysis are of large significance (e.g., public security, tornado warnings, earthquake warnings, etc.), the network involved in data delivery must be supported with robust protection and security mechanisms. The currently offered resilience and security solutions in both PSN and ON layers should be sufficient to protect the big data traffic.

- Low cost – reducing both capital expenditures (CAPEX) and operating cost (OPEX) is a key requirement for most of customers and delivering more bandwidth at a lower cost is a significant challenge to network operators and service providers. Because of the size of big data traffic, the required capacity can be only provisioned in the optical layer. According to our recent results shown in [Klinkowski2013], the use of EON architectures significantly improves metric like cost, power consumption, and spectrum usage when compared to the conventional WDM approach.

- Energy efficiency – the cost of energy is one of the main contributors in operational expenditures in the computer systems and networks. Moreover, a lot of consideration is being focused on “green” ICT solutions to answer the challenges following from the threat of energy crisis and environmental protection issues. The big data analysis consumes most of the energy in: transport network, data centers (processing and analysis) and storage systems. In the case of optical networks – according to [Klinkowski2013] – the WDM needs from 36% to 49% more energy than EON. Moreover, the high bandwidth provided in optical networks enables easy and flexible scheduling of big data tasks in sites that provide lower energy cost or green energy, without significant deterioration of parameters like response time or throughput.
5. Case Studies

In this section, we present exemplary results of simulations run to show main advantages of big data traffic provisioning in optical networks. Using real network topologies, traffic patterns generated according to Cisco predictions\(^3\), and assumptions following from big data applications like hyperspectral image processing [Liu2011], we report the key performance metrics related to provisioning of big data traffic in optical networks, that are: network throughput, OPEX and CAPEX cost, optical spectrum usage, and response time.

5.1 Simulation Setup

In the simulations we use two real network topologies: a) a European Nobel-EU backbone network (Euro) that includes 28 nodes and 82 directed links (Fig. 2a) and b) a United States backbone network (US) that consists of 26 nodes and 84 directed links (Fig. 2b) [Orlowski2010]. The data centers are located in different nodes of the network. For each network, the international traffic is carried through three interconnection points to other networks (e.g., locations for submarine cable landing stations). The locations of both interconnection points and data centers were made according to data available at http://www.datacentermap.com/.

\(^3\) Cisco Visual Networking Index, Cisco Global Cloud Index, http://www.cisco.com/
We analyze two alternative optical transport network scenarios, namely:

- **WDM** - a wavelength switched optical network operating within a fixed 50GHz ITU-T grid and allowing for mixed-line-rate transmission with nowadays available, fixed 10 Gb/s, 40 Gb/s, and 100 Gb/s WDM transponders. The transmission distance limits of transponders are equal to, respectively, 3200 km, 2300 km, and 2100 km (as in [Klinkowski2013]).

- **EON** - an elastic optical network operating within a flexible 6.25GHz ITU-T grid and with BV-Ts implementing the PDM-OFDM technology. Multiple modulation formats selected adaptively
between BPSK, QPSK, and m-QAM, where m belongs to \{8, 16, 32, 64\} are considered. Here, the spectral efficiency is equal to 1,2,\ldots,6 \text{ [b/s/Hz]}, respectively, for these modulation formats and PDM stands for Polarization Division Multiplexing, which allows to double the spectral efficiency. Three types of BV-Ts are used, namely, with 40 Gb/s, 100 Gb/s, and 400 Gb/s capacity limits, and with bit-rate adaptability of 10 Gb/s granularity. A guard band of 12.5GHz between neighboring lightpaths is assumed.

In both analyzed scenarios, we assume that the transmission range is extended by means of regenerators, which are applied whenever necessary.

In the simulations, the traffic patterns are generated according to two scenarios. In the former one – called Cisco – the traffic is created under the forecast of Cisco shown in “Cisco Visual Networking Index” and “Cisco Global Cloud Index” reports for year 2014 similarly to [Klinkowski2013]. The second scenario of traffic generation is a Uniform approach, i.e., the overall network traffic is uniformly divided to each origin-destination node pair, thus all demands in the network have the same capacity.

5.2 Network Throughput

The first goal of simulations is to show how increasing the network throughput will impact performance of optical network in terms of spectrum usage and cost. The traffic demands are provisioned in the optical transport network according to the following assumptions. For each demand, the transponder with the lowest capacity limit, but exceeding the demand volume, is selected. In addition, in EON, the modulation level for which a given performance metric was minimized is selected. As performance metrics we use: (i) optical spectrum resource usage in THz and (ii) network cost. The cost includes the following elements: CAPEX cost of equipment (transponders, regenerators) and 1 year OPEX cost related to fiber leasing. All costs are presented in Euro in the current prices. We take cost assumptions as in [Klinkowski2013].

In Fig. 3, we show main optical network performance metrics (i.e., spectrum usage, CAPEX/OPEX cost and power consumption) as a function of network throughput. The key observation is that for all
reported performance metrics, traffic patterns, and networks, the EON approach outperforms the WDM approach. The largest gap is spotted for spectrum, what can be easily explained by the large flexibility of optical grid usage in EONs in comparison to WDM. Thus, EON seems an appropriate direction for evolution of optical networks from the perspective of bandwidth demanding services including big data analysis. We can also see that for the Uniform traffic pattern the spectrum usage increases in a step-wise way with the increase of network throughput, what directly follows from the uniform growth of demand volumes at the same time for all connections in the network. Analysis of the network cost shows that with the increase of the throughput the cost grows mostly in a linear way.

Fig. 3. Spectrum usage, CAPEX/OPEX cost and power consumption for Euro and US networks as a function of network throughput (a) spectrum usage for Euro, (b) spectrum usage for US, (c) cost for Euro, (d) cost for US, (e) energy for Euro, (f) energy for US
5.3 Capacity Overprovisioning

Now we will present how the capacity overprovisioning influences the performance of the optical networks. The general assumptions of the simulations are as previously. The traffic matrix in the optical layer is a full mesh to provide a direct connection between each node pair in the virtual layer for the packet-switched layer. We assume that the overall traffic is set to 20Tbps and the capacity for each node pair demand is increased (overprovisioned) with a 10 Gb/s granularity. In Fig. 4, we report the spectrum usage, cost and power consumption as a function of capacity overprovisioning. We can easily notice that in terms of the spectrum usage, EON does not require much more additional spectrum when the overprovisioned capacity increases, while WDM starting from 70 Gb/s of overprovisioned capacity requires about twice more spectrum. The cost grows almost linearly with the capacity overprovisioning increase, and for 100Gb/s the cost is larger by about 60% comparing to the not overprovisioned network. It should be underlined, that in all reported cases EON outperforms WDM.
5.4 Response Time and OPEX Processing Cost

The last goal of simulations is to report the performance metrics related to processing of big data requests: response time and OPEX processing cost. The response time is defined as the whole time needed to response to a big data request and includes the following elements: execution time of the scheduling algorithm (50 ms), optical connection setup (50 ms for WDM and 100 ms for EON), delivery of input data and output (calculated according to the selected provisioning scenario and the size of data), computing time (calculated according to the number of available CPUs and speedup).

Since the main element of the OPEX cost for data centers is energy, we assume that the cost of using of 1 CPU in time of 1 second is proportional to the cost of electrical energy. Therefore, we present results of OPEX processing cost in uniform units. The current energy prices that we use in simulations for Euro and US networks are based on data provided in Eurostat (http://epp.eurostat.ec.europa.eu) and US Department of Energy (http://www.eia.gov/electricity/), respectively.

The scheduling algorithm takes into account two performance metrics: response time and OPEX cost with a special tradeoff parameter called response time tolerance (RTT) expressed in percentage. When a new request arrives, the minimum possible response time is calculated considering all processing nodes. Next, the algorithm finds the cheapest processing node (in terms of the OPEX cost) but assuring that the response time will be not greater by more than RTT comparing to the minimum possible response time. For instance, if RTT is 60%, the scheduler selects for the request the most
cost effective computing node among computing nodes that guarantee the response time not greater than 1.6 of the minimum response time.

We examine four provisioning scenarios of big data traffic, namely:

- 1 Gb/s connections in the packet-switched network (named as PS(1)).
- 10 Gb/s connections in the packet-switched network (named as PS(10)).
- 100 Gb/s connections in the optical network using WDM (named as WDM(100)).
- 400 Gb/s connections in the optical network using EON (named as EON(400)).

For each tested network (Euro and US) we assume there are 11 DCs (Fig. 2) that can be used to process big data requests. Available CPUs are assigned to the DCs proportionally to populations and GDP (Gross Domestic Product) of the node. We generate a set of 1000 tasks described by: source node, destination node, amount of input data, amount of output data, processing time on 1 CPU and speedup. The detailed data of each task is generated randomly, but with accordance to the examples of hyperspectral image processing requests shown in [Liu2011]. The tasks (requests) arrive one by one, following a Poisson process.

![Figure 5](image-url)  
Fig. 5. Response time and OPEX cost for Euro and US networks as a function of number of CPUs (a) RTT=0% for Euro, (b) RTT=60% for Euro, (c) RTT=0% for US, (d) RTT=60% for US; ‘lines’ and ‘bars’ correspond to OPEX costs and response times, respectively.
In Fig. 5, we show the performance of various big data provisioning scenarios with RTT=0% and RTT=60%. Surprisingly, we can notice that for RTT=0% with the increase of processing power (number of CPUs), the processing OPEX cost slightly grows. This follows from the fact that the if RTT=0%, then the algorithm considers the response time as the only objective. In consequence, when additional computing resources are available in the system, a relatively larger number of tasks is assigned to more expensive nodes what provides faster execution of scheduled tasks. We run the scheduling algorithm for various values of RTT parameter: 20%, 40%, ..., 200%. The most interesting tradeoff between the response time and OPEX cost is obtained for RTT=60%. On average, in comparison to reference RTT=0% for RTT=60% the response time is increased by 18%-26% depending on the provisioning scenario and network, while the OPEX cost is decreased by 23%-34%. Moreover, in opposition to RTT=0%, RTT=60% provides a decrease of the OPEX cost with the increase of the processing power – comparing 50 000 CPUs against 10 000 CPUs, the gain is 18%-24% for various provisioning scenario. It means that the additional goal to minimize the OPEX costs while controlling the response time tolerance benefits from additional computing power resources.

6. Conclusions

We studied the main advantages of optical transport networks for provisioning of big data traffic. We discussed two optical network approaches WDM and EON in the context of big data networking. We proposed and compared two scenarios of big data traffic provisioning in the network: service network provisioning and optical network provisioning. Moreover, we presented the key network requirements that must be provided according to a new service context following from big data applications. Next, using realistic scenarios, we run a range of simulations to report performance of various provisioning scenarios of big data traffic. We showed that the use of EON architectures significantly improves performance metrics for both network domain (cost, energy and spectrum usage) and processing domain (response time).

Acknowledgment
The work was supported by the Polish National Science Centre (NCN) under Grant DEC-2012/07/B/ST7/01215. The work of Mirosław Klinkowski was supported in part by the FP7 project IDEALIST (Grant Agreement no. 317999).

**Literature**


