Design of Multi-Granular Optical Networks

(Invited paper)

M. De Leenheer, J. Buysse, C. Develder, B. Dhoedt, and P. Demeester

Dept. of Information Technology, Ghent University – IBBT, Belgium
Tel: +32 933 14939, Fax: +32 933 14899, E-mail: marc.deleenheer@intec.ugent.be

Recent years have demonstrated the limited scalability of electronic switching to realize transport networks. In response, all-optical switching has been identified as a candidate solution to enable high-capacity networking in the future. One of the fundamental challenges is to efficiently support a wide range of traffic patterns, and thus emerges the need for equipment that is both practical and economical to construct and deploy. We have previously proposed the use of multi-granular optical cross-connects (MG-OXC), which support switching on both the wavelength and sub-wavelength level. To this end, the MG-OXCs are equipped with cheap, highly scalable slow switching fabrics, as well as a small number of expensive fast switching ports. This work summarizes our earlier work to demonstrate that a small number of fast switching ports suffice to support a wide range of traffic requirements, and that multi-granular optical switching can offer cost-benefits on a network-wide scale.

1. Introduction

Optical networks have a proven track-record in long-haul, point-to-point networking, where large amounts of data are transported in a cost-effective way. An enabling technology is Wavelength Division Multiplexing (WDM), as it allows multiple signals (wavelengths) concurrent access to a single fiber. However, interest is growing to use optical networks in edge and even access networks (e.g. Fiber To The Home or FTTH), mostly because of the predictable performance of photonic technology (i.e. High bandwidth, low latency). A major issue is O/E/O (optical/electronic/optical) conversions in the network, because the speed of electronic processing can not match the bandwidths currently offered in the form of wavelengths of 40 Gbps and higher. For this reason, most current research is focusing on all-optical networking solutions.

As of today, it is possible to create all-optical networks through the use of circuit-switched paths, which essentially reserve one or more full wavelengths between end points. For instance, Lambda Grids are a general term to refer to Grid applications making use of wavelengths (i.e. lambdas) to connect high-performance computing sites over an optical network [1]. However, novel applications are appearing which demand a much more fine-grained access to bandwidth capacity, as is demonstrated for instance in consumer Grids [2]. In such a scenario, data sizes become smaller, since aggregation of multiple data sources is much harder, and the bandwidth utilization would drop dramatically if full wavelengths were used by these applications. Consequently, the network must support reservation and allocation of bandwidth on a sub-wavelength scale. In this paper, we propose a generic multi-granular optical switch

architecture, which supports both circuits (wavelength level) and bursts (sub-wavelength level).

This paper will investigate the specific details of realizing optical networking solutions that allow efficient transfer of both large (circuit) and small scale (burst or packet) data sizes. This is useful for future optical (possible Grid) deployments, which must support a new and emerging generation of distributed network-based applications that combine scientific instruments, distributed data archives, sensors, computing resources and many others. Each application has its own traffic profile, resource usage pattern and different requirements originating in the computing, storage and network domains [3]. Dedicated networks do not offer sufficient flexibility to satisfy the requirements of each application type, nor are they economically acceptable. Hence it is vital to understand and redefine the role of networking, to support applications with different requirements and also offer service providers a flexible, scalable and cost effective solution. A dynamic optical network infrastructure with the ability to provide bandwidth granularity at different levels is a potential candidate. In this way, the network can adapt to application requirements and also support different levels of Quality of Service (QoS). However, care must be taken to devise a solution that remains scalable and cost effective.

More specifically, we define a *multi-granular optical switched network* as a network that is able to support dynamic wavelength and sub-wavelength bandwidth granularities with different QoS levels. As such, the network will support the three basic switching technologies in WDM networks; optical circuit switching (OCS), optical packet switching (OPS) and optical burst switching (OBS). In order to support these switching approaches, optical switching fabrics with speeds on the millisecond range down to nanosecond range must be considered. As we will demonstrate in the following section, OCS can utilize millisecond switching technologies efficiently, whereas this switching speed causes bandwidth inefficiency and unpredictability for the performance of OBS. This is mainly caused by the high overhead incurred by large offset times required to configure slow switches. Consequently, fast switching fabrics should be introduced in the network.

The ideal solution would thus consist of deploying fast switches of large dimension; however current technology can only realize fast fabrics of limited scalability at a very high cost (for more details, a review of existing switching technologies is provided in [4,5]. Therefore, one possible solution is an optical cross-connect (OXC) which combines both slow and fast switching elements, with careful consideration of scalability and cost properties. Furthermore, users and applications can decide on slow or fast network provisioning, and additionally the network service provider can optimize bandwidth utilization by allocating wavelengths or lightpaths according to the traffic's switching needs.

In summary, the multi-granular OXC (MG-OXC) has a number of distinct advantages over traditional single-fabric switches:

- Bandwidth provisioning and switching capability at fiber, wavelength and sub-wavelength granularities;
- Agility and scalability of switching granularities providing a dynamic solution;
- Fast reconfigurability and flexibility on the electronic control of switching technologies;
- Cost-performance efficiency by offering an optimal balance between slow and fast switch fabric technologies.

The remainder of this paper is organized as follows. Section 2 further elaborates on the need for multi-granular switching, along with an indication of the most important challenges associated with the concept. Simulations are then used to evaluate a generic MG-OXC design for various traffic and design parameters in Section 3. A dimensioning study for the optimal design of a multi-granular optical network is presented in Section 4, while the concluding Section 5 summarizes our findings.

2. Problem Statement

The basic function of an optical switch is straightforward: create a connection between an input and an output port for each incoming data packet. The decision which output port a data packet should be directed to is usually made in a control unit available at each optical switch. This unit receives control information from each data transfer, which can be a reservation packet long in advance in the case of circuit switching, or a header prepended to the actual data in the case of packet switching. In this work, we assume data is sent in bursts (OBS), and control information is sent ahead of the actual data on a seperate control plane (i.e. out-of-band signaling). The time between the control packet and the actual data transfer is denoted by T_{offset}, and is the time available to the switch to reconfigure its internal cross-connections. Each switching fabric (see [4,5] for current technologies) is limited by its switching speed T_{switch}, and thus a data burst can only be switched successfully if $T_{switch} < T_{offset}$. T_{data} represents the length of the actual data in time, and thus is the time the switch's connection is in use. From this we can derive the *switch utilization*, i.e. the maximum fraction of time the switch is actually transferring data: $\frac{T_{data}}{T_{data} + T_{switch}}$.

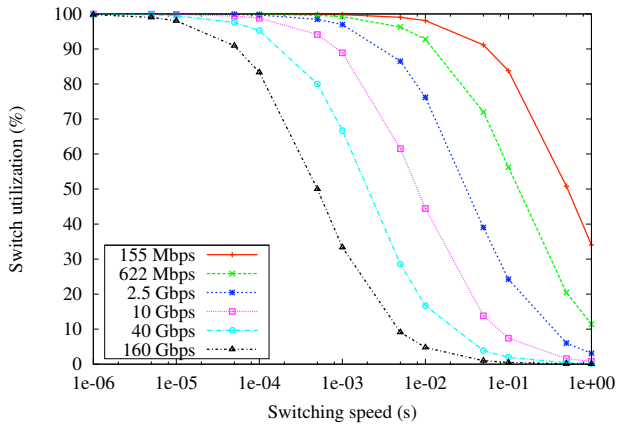

Figure 1: Upper bound for utilization of an optical switch for different switch speeds and bandwidths(data size is 10MB).

An illustration of this can be found in Figure 1, which shows the maximum utilization of an optical switch as a function of varying switching speeds. The data transferred has a size of 80 Mbit (10 MB), and the curves are shown for different link speeds. If we take, for instance, a switch speed of 10 ms (a representative value for micro-electro-mechanical systems or MEMS-based switches), we see that the switch utilization is 76% for a 2.5 Gbps link speed. This value drops to below 20% for 40 Gbps link speeds, and the situation clearly becomes worse for even higher bandwidths. Obviously, the same argument holds for a fixed bandwidth and decreasing data sizes. In contrast, a semiconductor optical amplifier or SOA-based switch can achieve nanosecond

switching speeds, and is thus much better adapted to support the full range of data sizes and bandwidths required for OCS, OBS and OPS.

The example shows that, to support very long data transfers (i.e. circuits), slow switching speeds are usually sufficient to obtain a high switch utilization, even for very high speed link rates. However, for smaller data transfers (burst or even packet sizes), high speed switching fabrics are required to achieve acceptable throughput in optical switching nodes. As current and emerging applications generate data according to very diverse distributions (both the data sizes and the instants of time at which the data is created), the idea emerged to integrate multiple types of switching fabric into a single optical switch. This concept is generally referred to as *multi-granular optical switching*, and becomes essential if a single, unified data plane needs to support a wide range of users and applications. This is especially true if complex grooming is to be avoided, which can be implemented in either a single-layer or multi-layer approach. The single-layer variant corresponds to the use of burst assembly algorithms, which have a negative effect on latency, while multi-layer grooming is less dynamic as multiple layers of control need to be activated before actual data transmission.

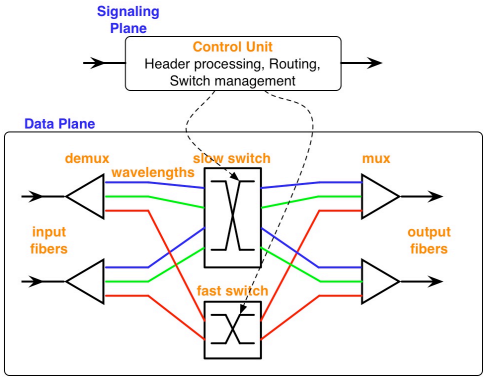

Figure 2: Multi-granular optical switch supporting wavelength and sub-wavelength switching.

Global network optimization not only depends on efficiency and utilization, but also on the feasibility to offer this technology in a cost-effective and practical way. Current optical switching technologies offer a broad range of switching speeds, but faster switching speeds generally have two distinct disadvantages: cost and scalability. For instance, MEMS switches have a typical switching time in the millisecond range, while it is technologically feasible to produce port counts of for instance 1000x1000. In contrast, SOA technology can only scale up to 32x32 port counts at very high cost, but at the same time can achieve switching speeds in the nanosecond range. Hence, cost-effectiveness is an important driver for hybrid optical switch designs requiring only a limited amount of expensive fast switching components. In response to this, Figure 2 presents the generic design of a multi-granular optical cross-connect (MG-OXC). The switch is composed of two separate switching fabrics, in order to support various application and QoS requirements on a common transport network infrastructure.

A final note is related to the practical realization of the MG-OXC, where several architectural choices remain an open research challenge. For instance, a sequential design (where the fast switching fabric is cascaded behind the slow fabric), allows reconfiguration of the fast wavelengths, at the expense of an increase in dimensionality of the slow switch. The design depicted in Figure 2 places the two switching fabrics in

parallel, and results in a slightly smaller slow switching matrix, but loses the reconfigurability of the fast wavelengths.

3. Node Performance Analysis

In this section, we summarize our results obtained in [6] where simulation analysis was used to provide insight in the behaviour of a single MG-OXC. The implementation allows us to evaluate an MG-OXC in a generic way, independent of architectural details. A comparison between MG-OXC and traditional, single-speed OXCs (slow only, fast only) is presented, and results are given for varying traffic load, fractions of slow/fast traffic and number of slow/fast ports available. The introduction of an MG-OXC in a network effectively creates a wavelength partitioning, by grouping wavelengths that are switched on the same type of switching fabrics. As such, an algorithm is required to assign generated traffic to a suitable wavelength partition, and the available wavelengths within a partition. This algorithm will be executed at the network's edge, thus before entering the all-optical data transport network. We refer the interested reader to [6] for more details. In the following, the assumption was made that only two partitions (corresponding to slow and fast) are introduced. Furthermore, all designs support 2 input and 2 output fibers, each fiber carrying 10 wavelengths. Neither wavelength conversion nor buffering capability is present in any of the switch designs. Each incoming data burst has a 50 % probability of choosing the first output fiber. The bandwidth of each wavelength is 10 Gbps, and traffic is generated according to a Poisson process with an average inter-arrival time of 15 ms. Data sizes follow an exponential distribution, with a varying average to establish the generated load. Because of the limited scale of currently deployed OBS networks there is no conclusive data available on a number of relevant traffic parameters. Thus, to control and evaluate the influence of different traffic types, the offset times between control packet and data are modeled as a 2-phase hyper-exponential distribution: $\alpha \cdot f_{slow} + \beta \cdot f_{fast}$. The pdf of the slow (respectively fast) traffic is an exponential distribution with average 100 ms (respesctively 10 ns). The slow switching fabric has a switching speed of T_{slow} = 10 ms, while the fast switch has T_{fast} = 1 ns.

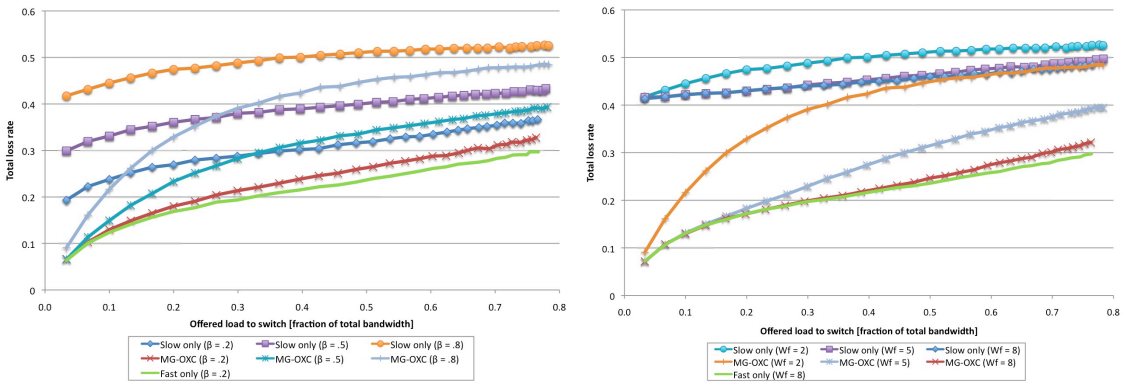

Figure 3: Higher fractions of fast traffic increase the total loss rate for fixed number of fast wavelengths (left), and higher number of fast wavelengths decrease the total loss rate for a fixed fraction of fast traffic (right)

In the first study, 2 wavelengths are available in the fast partition, while the remaining 8 are allocated for the slow partition, i.e. W_s = 8 and W_f = 2. Simulations were performed to evaluate the influence of the fraction of fast traffic for the three switch designs. The

resulting Figure 3 shows the total loss rate (i.e. ratio of dropped traffic to the offered load) for a varying offered load. First observe that for low loads, the relatively high loss rates can be attributed to the fraction of fast traffic which has an offset time lower than the fast switching speed. Then, an increasing fraction β of fast traffic causes higher loss rates, since the number of fast switching ports remains fixed (0 for the slow only, 2 for the MG-OXC). This does not apply to the fast only design (only shown for β = .2), whose performance is very similar for all fractions of fast traffic. Also, it is readily apparent that the MG-OXC outperforms the slow only design for all values of β. Another observation is that the MG-OXC offers loss rates similar to the fast only design, unless high fractions of fast traffic are generated (β = .5 and .8). This is not surprising considering the small number of fast switching ports available to the MG-OXC.

In the following study, the generated traffic consisted of 80% fast traffic (β = .8). Now, simulations focus on varying the number of slow/fast wavelengths in each partition (recall $W_s + W_f = 10$), and hence also the exact number of slow/fast wavelengths available to the MG-OXC. Figure 3 shows the total loss rate for a varying offered load, where one can immediately observe that an increased number of fast wavelengths results in a lower loss rate. That this result holds even for the slow only designs, is due to the simulation setup: the initial switch configuration connects the top input and output fibers (and likewise for the bottom fibers), and traffic is generated with a 50% probability of choosing either output fiber. Consequently, more or less half of the traffic on the W_f wavelengths can be switched correctly, and this explains why increasing values of W_f reduce the total loss rate. As before, the MG-OXC can provide an overall improved loss performance compared to the slow only design (behavior of slow only and MG-OXC are similar only for high loads and a severely under-dimensioned fast switching block). For high numbers of fast wavelengths (W_f = 8), the loss rate of the MG-OXC approaches the performance of the fast only design. Note again that results of the fast only design are shown only for W_f = 8, as other values for W_s lead to very similar loss rates.

4. Network Dimensioning

Assume the network is composed of OXCs, capable of switching circuits or slow bursts on millisecond scale (slow MEMS switch), and fast bursts or packets on a nanosecond scale (fast SOA switch). Likewise, traffic is generated by clients requiring both fast and slow switching. This corresponds to OBS for fast traffic, where multiplexing of different bursts on a single wavelength is allowed, and OCS for slow traffic, where end-to-end lightpaths are reserved exclusively for the endpoints. The question arises how to dimension the network for a static traffic demand and a pre-determined fraction of fast and slow traffic. The main objective is to minimize the network's cost, given a price ratio of slow over fast port costs. Another objective is to reduce the cost of the cross-connect with the highest cost, and as such obtain reduced node complexity.

In previous work [6], we have proposed ILP formulations which exactly solve the aforementioned problem statement. These ILP models were implemented and solved through the use of the ILOG CPLEX library. All OXC design approaches are evaluated, including slow only, fast only, and both multi-granular (parallel and sequential) architectures. Results are obtained for a specific scenario, defined by the Phosphorus topology consisting of 13 nodes and 36 directed links [7]. The traffic demand matrix is fixed, and consists of uniformly generated traffic between all source-destination pairs with average load .05. The low traffic demands are established in order to maximize the influence of traffic grooming; observe that when shortest path routing is used for the

given topology, the maximum number of demands making use of the same link is 15. To reduce computational complexity, we only considered the 5 shortest paths for each demand; this suffices for the topology considered, as the maximum distance between any node pair is 4 hops. Results show the total network cost and highest node cost, for objective functions which return the minimal total network cost (objective-1) or the minimal highest node cost (objective-2).

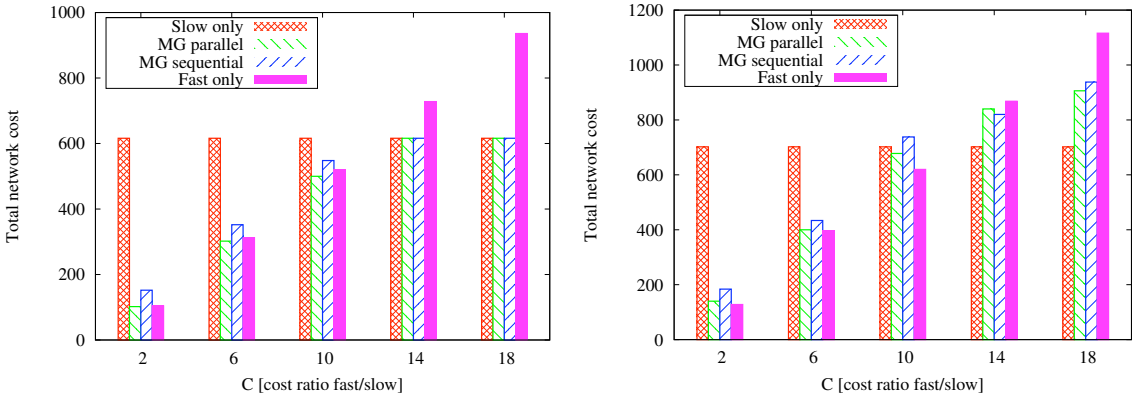

Figure 4: Total network cost when minimizing total network cost (left) or minimizing largest node cost (right)

Comparing the total network cost when minimizing either total network cost or highest node cost Figure 4, a number of interesting observations can be made. First, note that slow only returns constant network costs, due to its independence of cost ratio C. As expected, minimizing the highest node cost slightly increases total network cost when compared to objective-1 (observe the different Y-axis scales). Furthermore, MG sequential produces total network costs at least as large as MG parallel when minimizing network cost, although this is not the case when minimizing the highest node cost. Finally, for high values of C, the multi-granular approaches return identical results as the slow only design when using objective-1. In summary, significant cost savings are possible when using multi-granular optical switching, in comparison to slow only or fast only switching. Also, introducing reconfigurable fast wavelengths through the MG sequential design will only slightly increase total network cost.

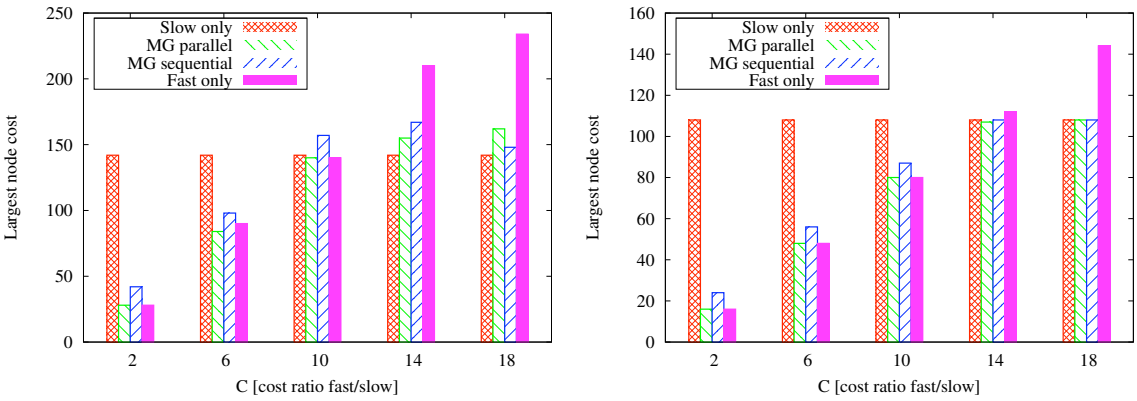

Figure 5: Largest node cost when minimizing total network cost (left) or minimizing largest node cost (right)

We now consider the highest node cost when minimizing total network cost or highest node cost Figure 5. Again slow only produces constant results, but lower values are

achieved by optimizing for objective-2 (again, note the different Y-axis scales). Observe that MG sequential returns highest node costs lower than MG parallel, only when minimizing the highest node cost. Multi-granular optical switching can thus clearly reduce the highest node cost, and consequently improve node complexity which is critical for scalability issues.

5. Conclusion

In this paper, we described the trend towards all-optical switching where data remains in the optical domain from source to destination. We indicated a number of problems related to supporting a wide range of applications and services on a single, unified optical transport plane. A possible solution has been identified in the concept of multi-granular optical switching, where OXCs integrate different switching fabrics to support switching at different bandwidth granularities. We studies the blocking behaviour of a single multi-granular switching node, and demonstrated the potential cost reductions on a network-wide scale.

Acknowledgements

This work was carried out with the support of the BONE-project (Building the Future Optical Network in Europe), a Network of Excellence funded by the European Commission through the 7th ICT Framework Programme, as well as the IST Phosphorus project. C. Develder is supported by the FWO as a post-doctoral fellow, J. Buysse received a Ph.D. scholarship from the IWT.

References

[1] T. DeFanti, C. de Laat, J. Mambretti, K. Neggers, B. St. Arnaud, “TransLight: A Global-Scale LambdaGrid for E-Science”, *Communications of the ACM*, vol. 46, no. 11, pp. 34-41, November 2003.

[2] M. De Leenheer, P. Thysebaert, B. Volckaert, F. De Turck, B. Dhoedt, P. Demeester, D. Simeonidou, R. Nejabati, G. Zervas, D. Klonidis, and M.J. O'Mahony, “A View on Enabling Consumer Oriented Grids through Optical Burst Switching”, *IEEE Communications Magazine*, vol. 44, no. 3, pp.1240-131, March 2006.

[3] D. Simeonidou, R. Nejabati, G. Zervas, D. Klonidis, A. Tzanakaki, M.J. O'Mahony, “Dynamic Optical Network Architectures and Technologies for Existing and Emerging Grid Services”, *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3347-3357, October 2005.

[4] S.J. Ben Yoo, “Optical Packet and Burst Switching Technologies for the Future Photonic Internet”, *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4468-4492, December 2006.

[5] G.I. Papadimitriou, C. Papazoglou, A.S. Pomportsis, “Optical Switching: Switch Fabrics, Techniques, and Architectures”, *Journal of Lightwave Technology*, vol. 21, no. 2, pp. 384-405, Februari 2003.

[6] G. Zervas, M. De Leenheer, L. Sadeghioon, D. Klonidis, R. Nejabati, D. Simeonidou, C. Develder, B. Dhoedt, P. Demeester, M. O'Mahony, “Multi-Granular Optical Cross-Connect: Design, Analysis and Demonstration”, *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, April 2009.

[7] S.Figuerola, N.Ciulli, M.De Leenheer, Y.Demchenko, W.Ziegler, A.Binczewski, “PHOSPHORUS: Single-step on-demand services across multi-domain networks for e-science”, *Proceedings of SPIE Asia-Pacific Optical Communications* (APOC), Wuhan, China, November 2007.