Evaluation of Optical Grid Scheduling through Dimensioning

C. Develder (1), M. De Leenheer (1), B. Dhoedt (1), P. Demeester (1)
1 : Department of Information Technology (INTEC), Ghent University - IBBT,
G. Crommenlaan 8 bus 201, BE–9050 Gent, Belgium - chris.develder@intec.ugent.be

Abstract *Optical Grids promise cost and resource efficient delivery of (distributed) services. We propose an optical Grid dimensioning methodology, and use it to evaluate the effect of Grid scheduling algorithms on the dimensions of such Grids.*

.

Introduction

Optical Grids promise to offer cost and resource efficient delivery of network services with possibly high data rate, processing and storage demands. Apart from (re)designing the architecture of a flexible optical layer, delivering the Grid promise implies answering fundamental questions, including dimensioning and routing/scheduling algorithms [1].

This paper focuses on optical Grid dimensioning, which fundamentally differs from dimensioning "classical" optical networks: (i) *Anycast routing paradigm*: A Grid job does not care where it is executed; (ii) *Burst starvation:* bursts can be lost not only because of network contention, but also through lack of Grid resources; and (iii) *Advance reservation*: Jobs may be announced relatively long in advance. Here, we focus on (i) and (ii).

Related work on dimensioning Grids is scarce. In [2] analytical ILP and heuristic approximations are used to cater for excess load. Other efforts assume that the fraction of jobs (originating at a particular site) going to a given computational Grid site is known, thus fixing a priori the arrival rates of jobs at each job execution site. In this paper however, to incorporate (i), we assume flexible scheduling strategies.

Optical Grid Network Architecture

Grid networks will benefit from optical technology, but whether to adopt Optical Circuit Switching (OCS) or rather Optical Packet/Burst Switching (OPS/OBS) is debatable. Depending on the ratio signalling time/job transmission time, OCS can be acceptable [3]. For small jobs, rather complex grooming/aggregation at the OCS edges will be required. As job data size reduces and/or latency-sensitivity increases, OBS will be more efficient [4]. Another advantage of OBS [5] is its ease in dealing with highly dynamic traffic patterns (both in space and time). The methodology proposed here can be used for both OBS and OCS choices.

Dimensioning Optical Grids

The problem we will solve is the following:

Given:

– A graph representing the network topology (nodes representing Grid sites and switches, links the optical fibers interconnecting them),
– The arrival process of jobs originating at each site,
– The job processing capacity of a single server, and
– A target maximum job loss rate

Find:

– The amount of Grid servers at each site, and
– The amount of link bandwidth to install,
– While meeting the maximum job loss rate criterion.

We take an iterative dimensioning approach, first calculating the amount of server sites needed, and subsequently deriving the inter-site job rates, hence bandwidth. Backed by real world Grid measurements, we will assume Poisson job arrivals [6].

Here, we do not take into account buffering: if at job arrival no free server is found, the job is lost. Thus, assuming Poisson arrivals (mean arrival rate λ), and exponentially distributed job processing times, we use the ErlangB formula to calculate the total number of servers n required to achieve a maximal loss rate L. To place the n servers among the N sites, we consider three strategies:

(i) *unif*: uniformly distribute the servers among all Grid sites (put n/N at each site);
(ii) *prop*: distribute the servers proportionally to the arrival rate at each site (if λ_i is the job arrival rate at site i, then put $n\,\lambda_i/\lambda$ servers at site i);
(iii) *lloss*: try and achieve the same "local loss rate" at each site, i.e. use ErlangB to calculate n_i as the number of servers to install locally at site i to achieve loss rate L, and install $n \cdot n_i/(\Sigma\, n_i)$ servers.

The scheduling algorithm decides where a job is executed. All scheduling approaches studied here will always choose a local server (i.e. at the job arrival site) if one is free. The approaches only differ in electing a remote server for job execution:

(i) *rand*: randomly choose a free server (i.e. among K free servers, each has 1/K chance);
(ii) *SP*: the closest free server in terms of hop count is chosen, thus striving to minimize network usage;
(iii) *mostfree*: choose a free server at site S, where S is the site with the highest number of free servers, in an attempt to avoid overloading sites and thus limiting non-local job execution.

Case Study

We performed a case study on a European network topology with 37 nodes and 57 bidirectional links. The job arrival rates at each site were chosen randomly (each rate λ_i was with 30% chance uniformly chosen in [1,15] and 70% from [30,60]).

The first criterion to judge the scheduling and dimensioning strategies by is the amount of jobs, taken over all sites, that is processed locally, shown in Fig. 1. Note the relatively low fraction of locally processed jobs, due to the absence of buffering and the high resource load (scaling the arrival rates down to 90%, we achieve ~70% local processing).

As intuitively expected, the *prop* and *lloss* strategies (placing more servers at sites where more jobs originate) achieve higher local processing rates. From the variation on local processing rates over all sites (see Fig. 2), we learn that *lloss* achieves its aim of equalizing local processing rates, esp. for the *mostfree* scheduling strategy.

From the scheduling perspective, *mostfree* confirms our intuition by achieving the highest local processing rates. Still, the difference with the other ones is rather limited. *SP*, by its deterministic order in choosing sites for remote processing, systematically (over)loads the same servers, thus achieving the lowest local rates.

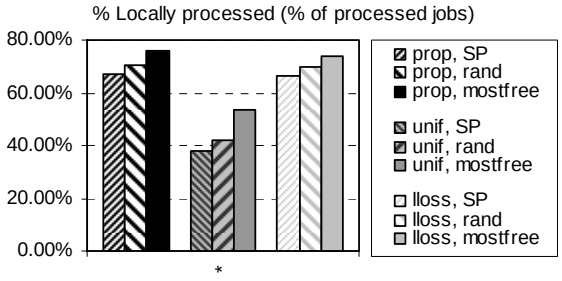

Fig. 1. Fraction of jobs that are processed locally (i.e. at originating site).

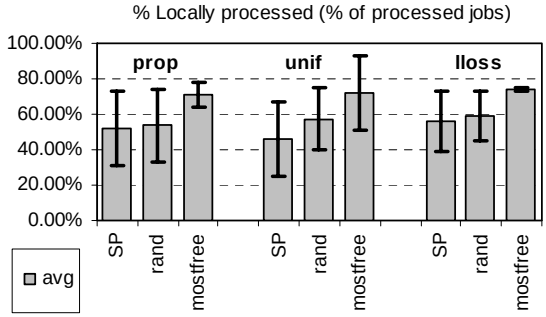

Fig. 2. Local processing fraction averaged over all sites, error bars indicate stdev.

The last step in the dimensioning process is determining link bandwidths. Using the site-to-site job rates, either an OBS or OCS network can be appropriately dimensioned using conventional methods, e.g. using the ErlangB formula to calculate the number of wavelengths on each link. (In this particular study using shortest path routing, the amount of wavelengths for OCS is a factor 5 higher.)

In Fig. 3 we present the total amount of jobs crossing each link. As expected, the *SP* scheduling achieves the lowest network load, by minimizing the path length that jobs have to cross. *Mostfree* obviously achieves lower network loads than *rand* due to its higher local processing rates, but by ignoring the network topology never comes close to *SP*. Note the striking impact of choosing an appropriate scheduling strategy: relative differences are bigger than comparing different dimensioning approaches.

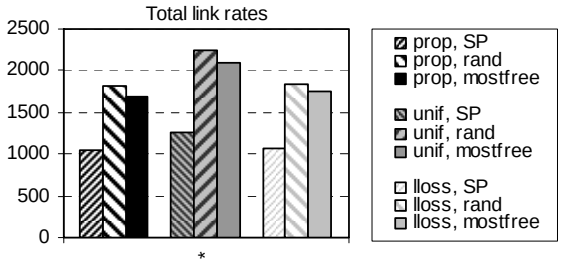

Fig. 3. Total link rates, i.e. number of jobs per time unit crossing each link summed over all links.

Conclusions

We outlined a dimensioning approach, calculating both server site capacities and network dimensions for optical Grids, to compare site dimensioning strategies and job scheduling algorithms. The impact of the scheduling mechanism on the required bandwidth is striking, and non-optimal job allocation in terms of processing resources (*SP* has lowest local processing %) is more than compensated by optimizing network use. Concerning dimensioning strategies, we found that *prop* leads to the cheapest network (lowest bandwidth), slightly outperforming the *lloss* strategy (which pays a price for achieving more local processing fairness amongst different sites).

Acknowledgements

This work was partly supported by the EU through the IST Project Phosphorus (www.ist-phosphorus.eu). C. Develder is a post-doctoral fellow of the Research Foundation – Flanders (FWO–Vlaanderen). M. De Leenheer thanks the IWT for his Ph. D. grant.

References

1 D. Simeonidou, et al., “Dynamic optical network architectures and technologies for existing and emerging Grid services,” J. Lightwave Techn., vol. 23, no. 10, Oct. 2005, pp. 3347- 3357.

2 P. Thysebaert, et al., “Using divisible load theory to dimension optical transport networks for Grid excess load handling,” Proc. OFC 2005, Anaheim, CA, USA.

3 F. Fahramand, et al., “A Multi-layered approach to Optical Burst-Switched based Grids,” Proc. WOBS, Boston, MA, USA, Oct. 2005.

4 M. De Leenheer, et al., “A view on enabling consumer oriented Grids through Optical Burst Switching,” IEEE Commun. Mag., vol. 44, no. 3, Mar. 2006, pp. 124–131.

5 C. Develder, et al., “Delivering the grid promise with optical burst switching,” (invited), Proc. OBS Workshop COIN-NGNCON, Jeju, Korea, 8 Jul. 2006.