University of Castilla-La Mancha

A publication of the

Computing Systems Department

Network-aware Peer-to-Peer Based Grid Inter-Domain Scheduling: A performance evaluation

by

Agustín Caminero, Omer Rana, Blanca Caminero, Carmen Carrión

Technical Report **#DIAB-09-04-2** April, 2009

This work has been jointly supported by the Spanish MEC and European Commission FEDER funds under grants “Consolider Ingenio-2010 CSD2006-00046” and “TIN2006-15516-C04-02”; jointly by JCCM and Fondo Social Europeo under grant “FSE 2007-2013”; and by JCCM under grants “PBI08-0055-2800” and “PII1C09-0101-9476”.

DEPARTAMENTO DE SISTEMAS INFORMÁTICOS
ESCUELA POLITÉCNICA SUPERIOR
UNIVERSIDAD DE CASTILLA-LA MANCHA
Campus Universitario s/n
Albacete - 02071 - Spain
Phone +34.967.599200, Fax +34.967.599224

Network-aware Heuristics for Inter-domain Meta-scheduling in Grids: A performance evaluation

Agustín Caminero[1], Omer Rana[2], Blanca Caminero[1], Carmen Carrión[1]

[1]Computing Systems Department. Escuela Politécnica Superior

Universidad de Castilla-La Mancha. 02071 - Albacete, SPAIN

{agustin, blanca, carmen}@dsi.uclm.es

[2]Cardiff School of Computer Science.

5 The Parade, Cardiff. CF24 3AA. UK.

o.f.rana@cs.cardiff.ac.uk

April 14, 2010

Abstract

When the QoS requirements of a user's job cannot be fulfilled through local resources, then the job may be forwarded to another domain. It therefore becomes necessary for each domain to maintain information about others, to faciliate such decision making. To be effective within large distributed systems (such as Grids), this must be done in a scalable and efficient manner. A proposal for meta-scheduling jobs between different administrative domains is presented and evaluated. The proposal is based on techniques already used for *peer-to-peer* systems, and integrated into a Grid Network Broker (GNB) framework, which performs meta-scheduling within an administrative domain. A key emphasis within this work is to utilize network metrics between domains, in addition to the computational/data capability available.

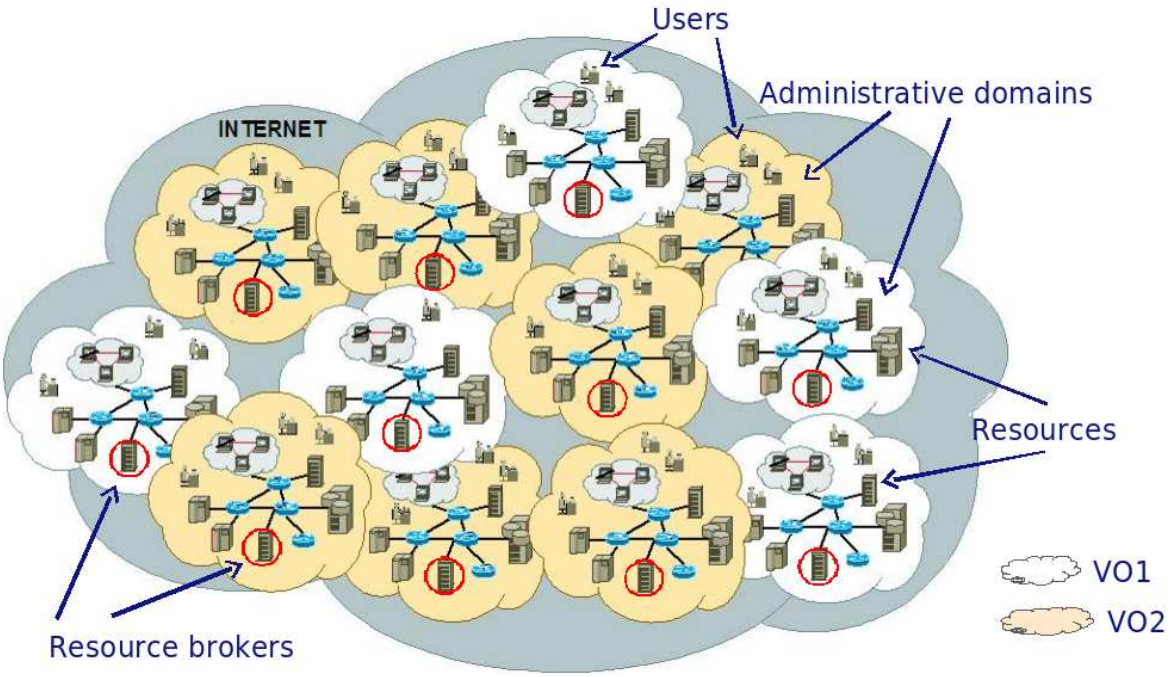


Figure 1: A Grid, made of several administrative domains.

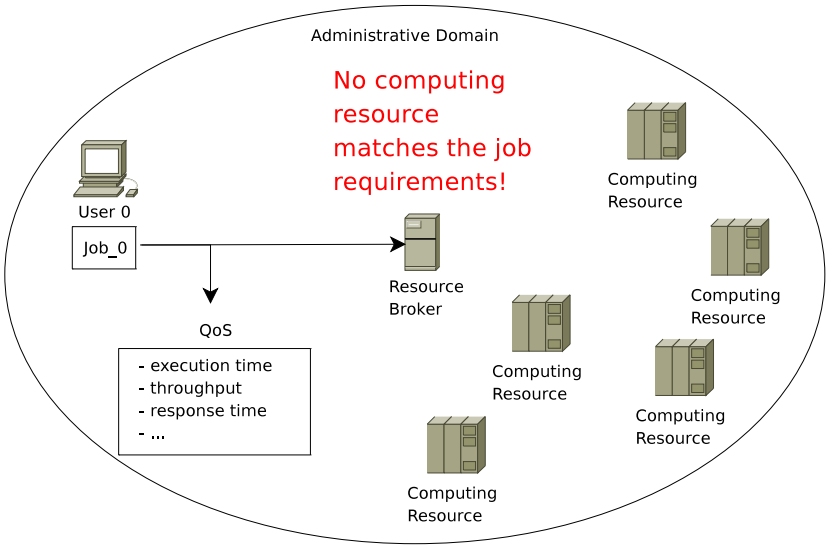


Figure 2: Match-making between job requirements and computing resources.

1 Introduction

Grid computing applications often require the use of resources that are managed by different organizational domains, each of which may keep their independence and autonomy [15]. As

illustrated in Figure 1, such sharing of resources across organizational boundaries leads to the formation of a *Virtual Organization* [42]. Hence, jobs belonging to a user may need to be executed in a computing resource from a different administrative domain, as shown in Figure 2. A user wishing to execute a job with particular *Quality of Service* (*QoS*) requirements, such as execution time or response time, must contact a resource broker in order to get a computing resource fulfilling those requirements. It now becomes necessary to consider an alternative administrative domain, if local resources cannot be found to fulfill these QoS requirements.

Our proposed heuristic is intended to manage QoS in a Grid system, and it is specially concerned with the interactions between administrative domains when performing the meta-scheduling of jobs to computing resources. It is implemented in an entity called *Grid Network Broker* (GNB), first presented in [8], and which has been extended in this paper to perform inter-domain meta-scheduling. The heuristic utilizes Peer-2-Peer (P2P) ideas centered on query routing, for identifying suitable neighbouring domains which may contain the required resources.

The paper is structured as follows: Section 2 reviews current proposals on network QoS in Grids, and the lack of attention paid to inter-domain relations. Also, existing proposals for inter-domain meta-scheduling are revised. Section 3 explains our proposal of inter-domain meta-scheduling. Section 4 provides an evaluation, demonstrating the usefulness of our work, and Section 5 provides conclusions and guidelines for future work.

2 Related work

The provision of QoS in Grids has been addressed by several research projects, among others [32] [7] [2] [1] [35] [36] [16] [29].

Regarding inter-domain relations, GARA [32] is difficult to scale, since users (or a broker acting on his behalf) has to authenticate himself with all domains. On the other hand, NRSE [7] is able to automatically negotiate a multi-domain reservation by communicating with its counterpart on the remote network, on behalf of its client. Reservations across multiple domains are made using two NRSEs, one at each end (improving on GARA's limitation), but it relies on the assumption that the core network is over-provisioned. Besides, NRSE is only aimed at performing network reservations, not meta-scheduling of jobs to computing resources.

Interactions between different administrative domains have been studied, among others, in [5] [6] [13] [10] [18] [24] [41], but they are mainly concerned with security issues, not scheduling. Furthermore, the combination of Grid computing with P2P has been studied, among others, in [34] [39] [40] [20] [37] [38]. Among them, Talia et al. [34] propose a P2P protocol for efficient invocation of Grid Services, and an architecture for resource discovery that adopts a P2P approach to extend the model of the GT3 information service. They propose a modified Gnutella discovery protocol – *Gridnut* – which makes it suitable for OGSA Grids. In particular, Gridnut uses appropriate message buffering and merging techniques to make Grid Services effective as a way to exchange messages in a P2P fashion.

Xion et al. [39] develop an algorithm for finding services in a P2P Grid. To use it, Grid resources are at first aggregated into a *GridPeer*. Then, when a Grid resource is needed, a genetic algorithm is used to find the closest GridPeer. Then, to find accurate resources within the GridPeer found, it applies the Ant algorithm. Similarly, Xu et al. [40] presented a framework for the QoS-aware discovery of services, where QoS is based on feedback from users. Gu et al. [20] proposed a scalable aggregation model for P2P systems to automatically aggregate services to support distributed application delivery, which satisfy user specified QoS guarantees.

Also, given the scenario where no suitable computing resource is available in the local administrative domain, a major issue is choosing the neighbor domain to which the query will be resubmitted. The proposal of the e-Protein Project [28] is called *Job Yield Distribution Environment* (*JYDE*) [27]. One of its components is the *Grid Distribution Manager* (*GriDM*, or *DM*), a P2P system that performs inter-domain meta-scheduling and load balancing above the intra-cluster schedulers like SGE, Condor, etc. On the submission server, GriDMs form a P2P network and attempt to balance the load across them. GriDM works by constantly checking the lengths of the wait queues at each site. When a queue on a particular site falls below a threshold, new permits are issued for that site, so that more jobs can be submitted to that site. The aim of this strategy is to keep every CPU at every site running jobs, and to keep a few jobs waiting at each site at any time, but not so many that it would hinder the DM's ability to make meta-scheduling decisions [27]. Thus, network QoS provision cannot be considered as one of the aims of this proposal.

Gnutella [17] uses flooding, requiring each peer to forward the query to all its neighbors. Every query has a *time-to-live* (TTL), which is decremented each time a peer receives a query. When the TTL reaches 0, the query will be rejected, and the user informed of the rejection. When one of the peers accepts the query, it also informs the user. Due to the fact that the number of queries increase each time they are forwarded by a peer – many different peers may accept the same query. In this case, the job will be executed in the peer whose answer reaches the user first.

DIANA [3] performs global meta-scheduling in a local environment, typically in a LAN. In DIANA, a set of meta-schedulers are used that work in a P2P manner. Each site has a meta-scheduler that communicates with all other meta-schedulers on other sites. DIANA has been developed to make decisions based on global information. This makes DIANA unsuitable for a realistic Grid testbed, such as the LHC Computing Grid [23], which has

around 200 sites and tens of thousands of CPU (for a map showing real time information, see [19]).

Assunção et al. [12] provide an architecture for the inter-networking of *islands of Grids*, which identifies and proposes an architecture, mechanisms, and policies that allow the inter-connectivity of Grids, and allows Grids to grow in a similar manner to the Internet – referred to as the *InterGrid*. The proposed InterGrid architecture is composed of Gateways responsible for managing peering arrangements between Grids.

3 Inter-domain meta-scheduling

The architecture presented in this work provides meta-scheduling of jobs to computing resources in different administrative domains. When a user queries the GNB for a computing resource to run a job, the GNB will proceed with a selection procedure. If there is a suitable resource in the local domain, the job will be allocated to that resource – alternatively, a resource in another domain may be required – requiring the GNB to determine which domain should be chosen.

Figure 3 shows the *intra-domain* meta-scheduling architecture. When the GNB of a domain receives a job to be scheduled, and no suitable computing resource exists locally, the GNB chooses one of the neighbor domains, and forwards the query to it. Apart from the GNB, each domain has other entities, such as a resource monitor (for instance, Ganglia [25]), a bandwidth broker (BB, such as [33]), and a *Grid Information Service* (GIS, such as [14]). A number of assumptions are necessary for the effective deployment of such an architecture. The *first* assumption is that each domain must be capable of providing the resources it advertises, i.e. when a domain publishes that it has, e.g, X machines with Y speed, those

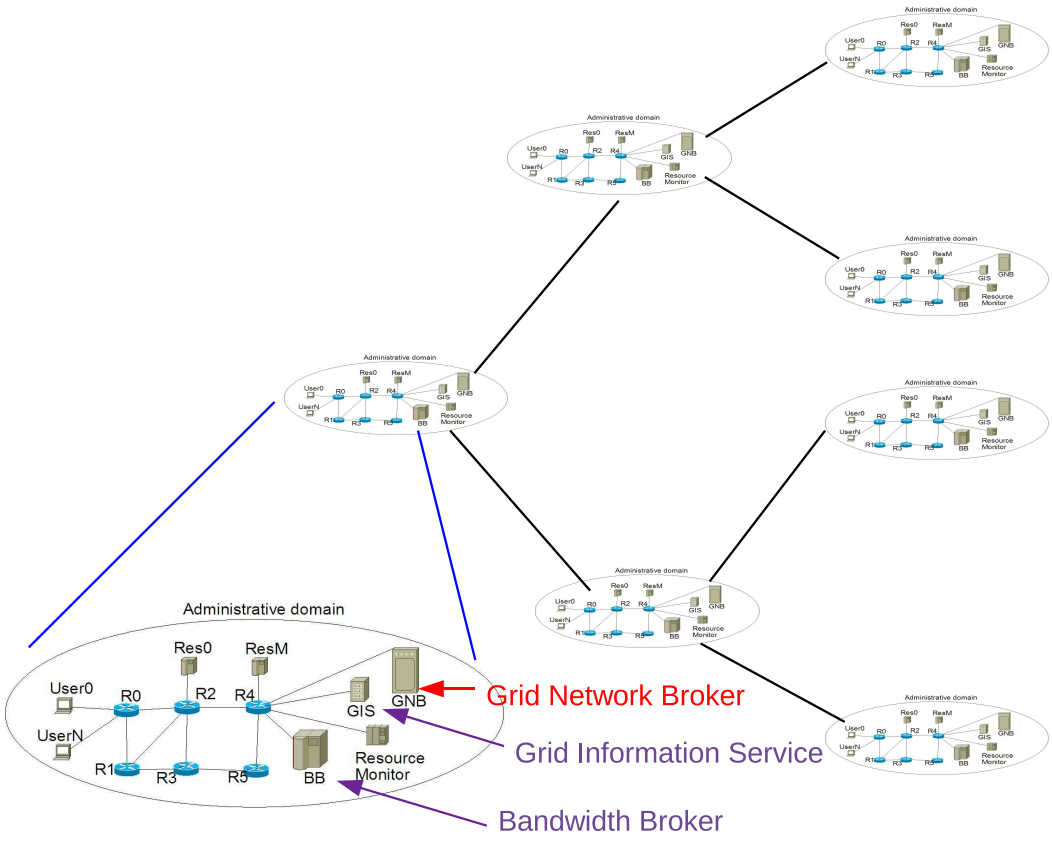


Figure 3: Inter-domain meta-scheduling architecture.

machines must be available *within* the domain, and conform to the advertised specification. Hence, a domain must not contain a pointer to machines held in other domains, but should be able to offer these machines locally. Specifying a pointer to machines held elsewhere is not useful to us, because the effective bandwidth and the number of hops of the network path from the current domain to each neighbor is needed. This path will be used by the job during its transmission. The *second* assumption is that the resource monitor should provide exactly the same measurements in all the domains. Otherwise, no comparison between resources available within different domains can be made. Besides, when there are more than one network paths from one domain to another, *Border Gateway Protocol* (BGP) [31] will decide which is the optimal path.

The concept of *Routing Indices* (RI) [11] is used in order to forward queries to neighbors that are more likely to have the required resources. Forwarding decisions use the local RI value of neighbouring domains, rather than selecting neighbors at random or by flooding the network by forwarding the query to all neighbors. RI will be explained the next.

3.1 Routing Indices

Routing Indices (RI) [11] were initially developed for document discovery in P2P systems, and have also been used to implement a Grid information service in [30]. The goal of RIs is to help users find documents with content of interest across potential P2P sources efficiently.

RI are used to make query forwarding decisions between domains in our system, and to avoid the need for flooding the entire network. The RI represent the availability of data of a specific type in the neighbor's information base. A version of RI called *Hop-Count Routing Index* (*HRI*) [11] is used, which considers the number of hops needed to reach a datum. This implementation of HRI calculates the aggregate quality of a neighbor domain, based on the number of machines, their power, current load and the effective bandwidth of the link between the two domains, as described in equation (1).

$$I_p^l = \left(\sum_{i=0}^{num\_machines_p} \frac{max\_num\_processes_i}{current\_num\_processes_i} \right) \times eff\_bw(l, p) \quad (1)$$

where I_p^l is the information that the local domain l keeps about the neighbor domain p; $num\_machines_p$ is the number of machines domain p has; $current\_num\_\_processes_i$ is the current number of processes running on the machine; $max\_num\_\_processes_i$ is the maximum number of processes that can be run on that machine, and will be explained later

on; $eff\_bw(l, p)$ is the effective bandwidth of the network connection between the local domain l and the peer domain p, and is calculated by considering measurements obtained by SNMP [26], as pointed out in [8]. Predictions on the values of the current number of processes and the effective bandwidth can be used, for example, and calculated as pointed out in [8]. As it can be seen, the network plays an important role when calculating the quality of a domain.

When using this equation, we aim at stressing the fact that both computing and network capability are equally important, and both parameters must be considered in order to decide about the quality of an administrative domain.

The $max\_num\_processes_i$ metric is is used to determine how powerful a particular machine is. It is calculated by considering the speed of the CPU and the amount of memory it has. Equation (2) shows the actual formula used.

$$max\_num\_processes = k_1 \times \frac{memory}{max(memory)} + k_2 \times \frac{cpu\_speed}{max(cpu\_speed)} \quad (2)$$

In Equation (2), k_1 and k_2 are two weighting constants that show the importance of each normalized parameter (memory and CPU speed) when calculating the maximum number of processes. Also, $k_1 + k_2$ represents the maximum number of processes we would like to have in the best of our machines. That is, if we take the machine with the fastest CPU and the machine with the lagest memory, $k_1 + k_2$ the maximum number of processes in that machine. This is done in order to allow local administrators to set limits on the use of resources. The maximum memory and CPU speed must be propagated between peers, so that all the peers share the same values for them.

This equation has been chosen because the capability of a computing resource depends on CPU speed and memory size, been both parameters very important. This is, the more

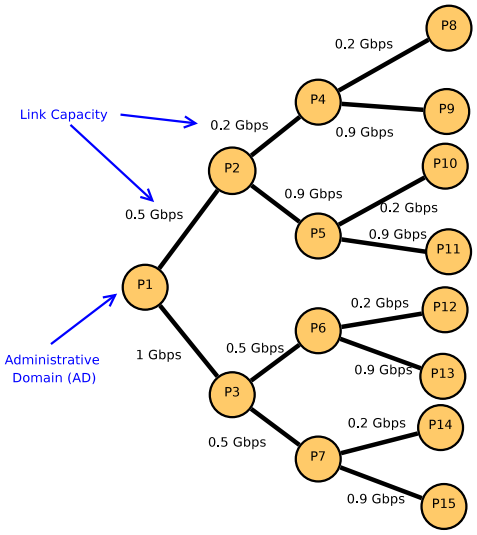


Figure 4: Peer-to-peer relations between several administrative domains.

memory a machine has, the more processes can be executed at the same time in that machine. Besides, the fastest a CPU is, the sooner processes will be executed.

Equations (1) and (2) show why the two assumptions mentioned before are needed. As the effective bandwidth between domains is needed in Equation (1), it is important that a domain correctly report it's resource capabilities. Otherwise, the actual links used to transmit the job could not be accurately characteized. The second assumption requires that the domains must report the same monitoring metrics (such as CPU speed, current load and effective bandwidth), as otherwise no comparison could be made between domains.

HRI have been used as described in [11]: in each peer, the HRI is represented as an $M \times N$ table, where M is the number of neighbors and N is the horizon (maximum number of hops) of our Index: the n^{th} position in the m^{th} row is the quality of the domains that can be reached going through neighbor m, within n hops. As an example, the HRI of peer P_1 are provided in Table 1 (for the topology depicted in Figure 4), where $S_{x.y}$ is the value for peers that can be reached through peer x, and are y hops away from the local peer (in this case, P_1), and calculated as in Equation 3. Hence, $S_{2.3}$ represents the quality of domains which can reached through peer P_2, whose distance from the local peer is 3 hops.

Table 1: HRI for peer P_1.

Peer	P_2	P_3
1 hop	$S_{2.1}$	$S_{3.1}$
2 hops	$S_{2.2}$	$S_{3.2}$
3 hops	$S_{2.3}$	$S_{3.3}$

Table 2: Detailed HRI for peer P_1.

Peer	P_2	P_3
1 hop	$I_{P_2}^{P_1}$	$I_{P_3}^{P_1}$
2 hops	$I_{P_4}^{P_2} + I_{P_5}^{P_2}$	$I_{P_6}^{P_3} + I_{P_7}^{P_3}$
3 hops	$I_{P_8}^{P_4} + I_{P_9}^{P_4} + I_{P_{10}}^{P_5} + I_{P_{11}}^{P_5}$	$I_{P_{12}}^{P_6} + I_{P_{13}}^{P_6} + I_{P_{14}}^{P_7} + I_{P_{15}}^{P_7}$

$$S_{x.y} = \begin{cases} I_{P_x}^{P_l}, \text{ when } y = 1 \\ \sum_i I_{P_i}^{P_t}, \forall P_i, d(P_l, P_i) = y \wedge d(P_l, P_t) = y - 1 \wedge d(P_t, P_i) = 1, \text{ otherwise} \end{cases} \quad (3)$$

In Equation 3, $d(P_x, P_i)$ is the distance (in number of hops) between peers P_x and P_i. $S_{x.y}$ is calculated based on the distance from some local peer. When the distance is 1, then $S_{x.y} = I_{P_x}^{P_l}$, because the only peer that can be reached from local peer P_l through P_x within 1 hop is P_x. Otherwise, for those peers P_i whose distance from the local peer is y, the information that each peer P_t (which is the neighbor of P_i) keeps about them has to be added. Hence, the HRI of peer P_1 will be calculated as shown in Table 2.

3.2 Goodness function

In order to use RIs, a key component is the *goodness function* [11]. The goodness function is needed to decide the quality of each neighbor domain. This is done by considering

the quality of peers that can be reached through each neighbor, and their distance from the local peer. In other words, for each direct neighbor of the local peer, the goodness function will decide how good each of them is. This is done by considering the HRI and the distance between neighbors.

For example, consider the topology depicted in Figure 4. If peer P_1 needs to forward a job to one of its neighbors, it will have to decide between P_2 and P_3. So, P_1 will apply the goodness function to both of them, and one of them will be chosen. When applying the goodness function to P_2, the quality of peers that can be reached through it (namely P_4, P_5, P_8, P_9, P_{10}, and P_{11}) will be considered. In the same way, the quality of P_3 depends on the quality of P_6, P_7, P_{12}, P_{13}, P_{14}, and P_{15}. This is done by means of the HRI, since it keeps information on the peers that can be reached through each neighbor peer.

Imagine that the best resources belong to peer P_6, and all the resources belonging to the other peers are overloaded. In this case, P_1 would choose to forward the job to P_3, because although P_3 does not have a suitable resource, it is closer to P_6 than P_2.

Our goodness function can be seen in Equation (4), where p is the peer domain to be considered; H is the horizon for the HRIs; and F is the fanout of the topology. As [11] explains, *horizon* provides an upper bound on the distance (number of hops) searched; hence, peers whose distance from the local peer is higher than the horizon will not be considered. Meanwhile, the *fanout* of the topology is the maximum number of neighbors a peer has.

$$goodness(p) = \sum_{j=1..H} \frac{S_{p.j}}{F^{j-1}} \tag{4}$$

3.3 Example

The use of HRIs is demonstrated through an example based on the topology depicted in Figure 4. Suppose that all the peers (recall that each peer represents a whole administrative

domain) in Figure 4 have 2 machines, each one with a speed of 1 GHz and 1 GB of memory; $k_1 = k_2 = 50$ and the current number of processes is 40 for all of them. Link bandwidths appearing in the figure have been calculated as [8] suggests. Finally, the horizon is 3, and fanout is 3. In order to calculate the HRI of peer P_1, each I_p^l is calculated as shown in Table 3. These I_p^l produce the HRI depicted in Table 4 by adding the results presented in each cell of Table 3.

Table 3: Calculation of I_p^l.

Peer	P_2	P_3
1 hop	$I_{P_2}^{P_1} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.5 = 2.5$	$I_{P_3}^{P_1} = \left(\frac{100}{40} + \frac{100}{40}\right) * 1 = 5$
2 hops	$I_{P_4}^{P_2} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.2 = 1$	$I_{P_6}^{P_3} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.5 = 2.5$
	$I_{P_5}^{P_2} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.9 = 4.5$	$I_{P_7}^{P_3} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.5 = 2.5$
3 hops	$I_{P_8}^{P_4} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.2 = 1$	$I_{P_{12}}^{P_6} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.2 = 1$
	$I_{P_9}^{P_4} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.9 = 4.5$	$I_{P_{13}}^{P_6} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.9 = 4.5$
	$I_{P_{10}}^{P_5} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.2 = 1$	$I_{P_{14}}^{P_7} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.2 = 1$
	$I_{P_{11}}^{P_5} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.9 = 4.5$	$I_{P_{15}}^{P_7} = \left(\frac{100}{40} + \frac{100}{40}\right) * 0.9 = 4.5$

Table 4: Example HRI for peer P_1.

Peer	P_2	P_3
1 hop	2.5	5
2 hops	5.5	5
3 hops	11	11

If a computing resource is required at peer P_1, then the following goodness function is applied:

$$goodness(P_2) = \frac{2.5}{3^0} + \frac{5.5}{3^1} + \frac{11}{3^2} = 5.5$$

$$goodness(P_3) = \frac{5}{3^0} + \frac{5}{3^1} + \frac{11}{3^2} = 7.8$$

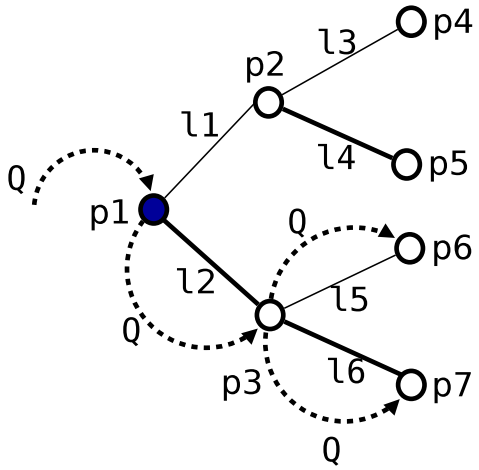


Figure 5: A query (Q) is forwarded from $p1$ to the best neighbors ($p3$, $p6$, and $p7$).

The goodness function produces a higher value for P_3 compared to P_2. This occurs because the network connection to P_3 makes it more suitable to execute jobs than P_2. Thus, the job would be forwarded to P_3.

3.4 Search technique

Several techniques are used for searching in P2P networks, including flooding (e.g. Gnutella) or centralized index servers (e.g. Napster). More effective searches are performed by systems based on distributed indices. In these configurations, each node holds a part of the index. The index optimizes the probability of finding quickly the requested information, by keeping track of the availability of data at each neighbor.

Algorithm 1 shows the way that our architecture performs the scheduling of jobs to computing resources. In our system, when a user wants to run a job, he submits a query to the GNB of the local domain. This query is stored (line 7) as it arrives for the first time to a GNB. The GNB looks for a computing resource in the local domain matching the

Algorithm 1 Search algorithm.

```
1:  Let q = new incoming query
2:  Let LocalResource = a resource in the local domain
3:  Let NextBestNeighbor = a neighbor domain select by the goodness function
4:  Let ToTry = the next neighbor domain to forward the query to
5:  for all q do
6:     LocalResource := null
7:     if (QueryStatus(q) = not present) then
8:        {the first time the query arrives at this domain, store the query}
9:        QueryStatus(q) := 1
10:       {look for a computing resource in the local domain}
11:       LocalResource := MatchQueryLocalResource(q)
12:    end if
13:    if (LocalResource == null) then
14:       {no computing resource in the local domain, so forward the query to a neighbor
          domain}
15:       ToTry := QueryStatus(q)
16:       NextBestNeighbor := HRI(q, ToTry)
17:       if (NextBestNeighbor == null) then
18:          {the query must be bounced back}
19:          Recipient := Sender (q)
20:       else
21:          Recipient := NextBestNeighbor
22:          QueryStatus(q) += 1
23:       end if
24:       ForwardQueryToRecipient(q, Recipient)
25:    else
26:       {tell the requester a computing resource has been found}
27:       SendResponseToRequester(q)
28:    end if
29: end for
```

requirements of the query (line 11). If the GNB finds a computing resource in the local domain that matches the requirements, then it tells the user to use that resource to run the job (line 27). Otherwise, the GNB will forward the query to the GNB of one of the neighbor domains. This neighbor domain will be chosen based on the *Hop-Count Routing Index, HRI*, explained before (line 16). The parameter $ToTry$ is used to decide which neighbor should be contacted next, as shown in Figure 5 – where $p3$ will contact $p6$); if the query is bounced back, then the 2^{nd} best neighbor will be contacted ($p3$ will contact peer $p7$), and so

on. Hence, a neighbor domain is only contacted if no suitable local computing resources are available.

4 Evaluation

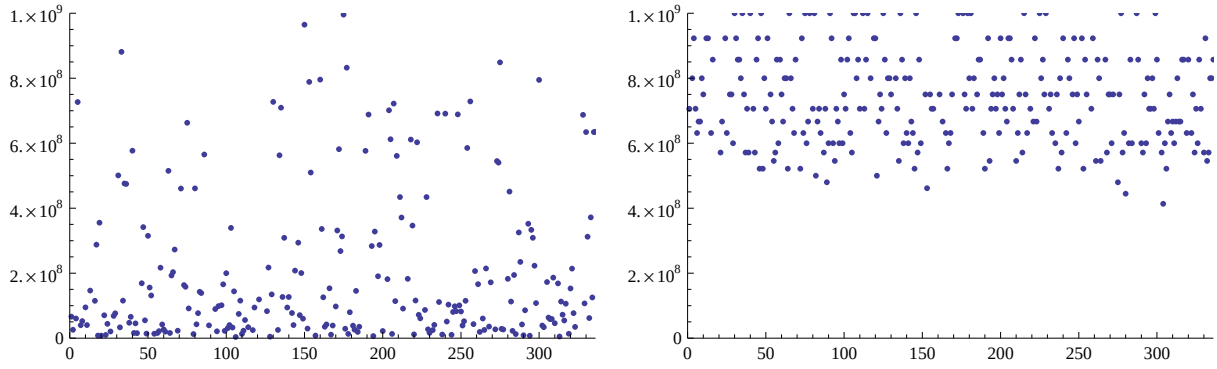
Two types of evaluations have been undertaken to validate our approach. First, we focus on how HRIs evolve when varying system parameters – an evaluation from the point of view of the *system*. The second evaluation is carried out to evaluate the approach from the point of view of the *users*.

4.1 System point of view

For the first evaluation, the topology presented in Figure 5 is used; all the data presented here refer to peer $p1$. In the simplest case, all link bandwidths are assumed to be 1 Gbps, and all the peers have 1 resource made of 1 machine, with 4 Gb of memory and CPU speed of 1 GHz.

For Equation 1, the values of $current\_num\_processes_i$ have been approximated as a uniform distribution between 10 and 100, and the $max\_num\_processes_i$ as 100. Regarding the $eff\_bw(l, p)$, a Poisson distribution has been considered for those links that are heavily loaded, and a Weibull distribution for those links which are not, as [9] suggests. In Figure 5, the even links will be heavily used, and are depicted with a thicker line.

To calculate the mean μ for the Poisson distribution, and scale β and shape α for the Weibull distribution, it has been considered that the level of use of heavily used links is



(a) Use of a not heavily loaded link (Weibull distribution).

(b) Use of a heavily loaded link (Poisson distribution).

Figure 6: Variation of link usage.

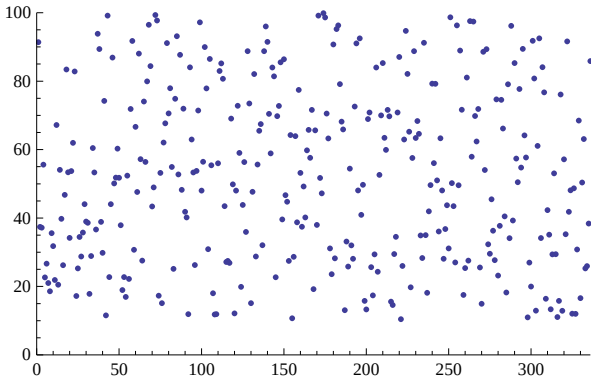
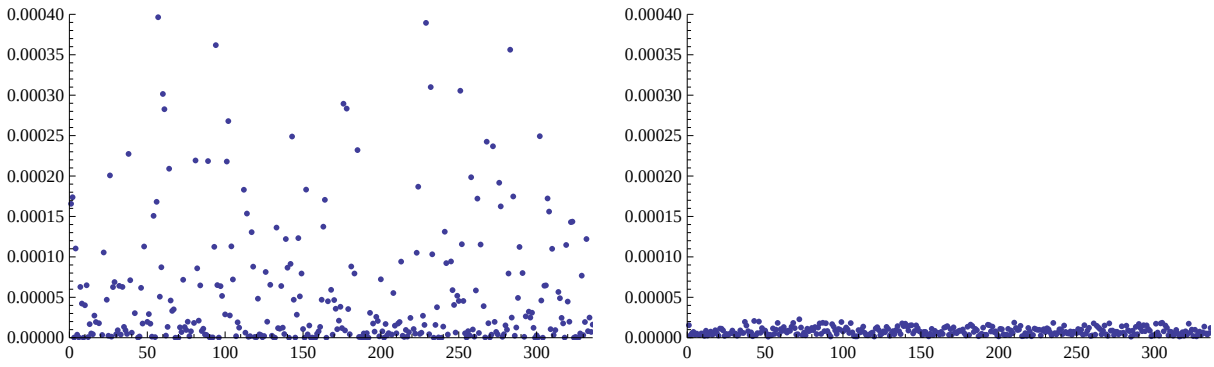


Figure 7: Variation of the number of processes (Uniform distribution).

80 %, whilst less heavily used links exhibit a 10 % usage. This way, if a heavily used link transmits 800 Mb in 1 second, and the maximum transfer unit of the links is 1500 bytes, the inter-arrival time for packets is 0.000015 *seconds* – corresponding to μ of the Poisson distribution. In the same way, the value for the β parameter of the Weibull distribution is calculated to be 0.00012 *seconds*.

A measurement period of 7 days has been simulated, with measurements collected every 30 minutes. Figures 6 and 7 present the variation in the use of links and the number of processes, following the mathematical distributions explained before. Figure 6 represents the level of use of links compared to the actual bandwidth (1 Gbps), per measurement. Heavily



(a) $S_{2.1} = I_{p2}^{p1}$

(link $p1 - p2$ is not heavily loaded).

(b) $S_{3.1} = I_{p3}^{p1}$

(link $p1 - p3$ is heavily loaded).

Figure 8: Variation of $S_{x.y}$ for loaded/ unloaded links.

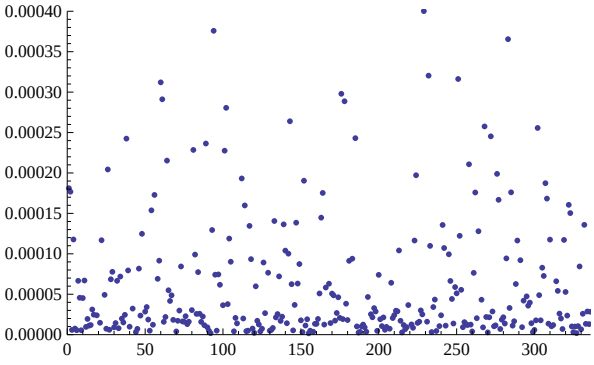
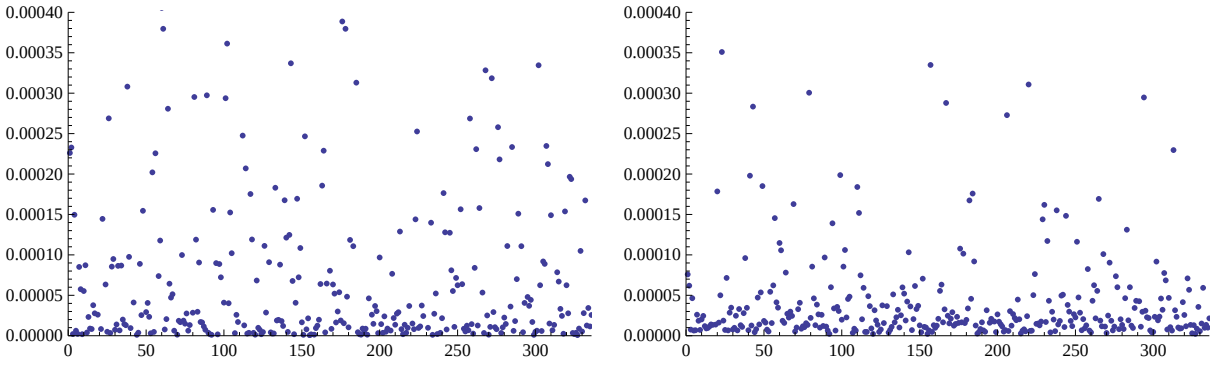


Figure 9: $S_{2.2}$ ($S_{3.2}$ would also look like this).

used links get a higher used bandwidth than other links. This data is used to determine along which link a query may be forwarded.

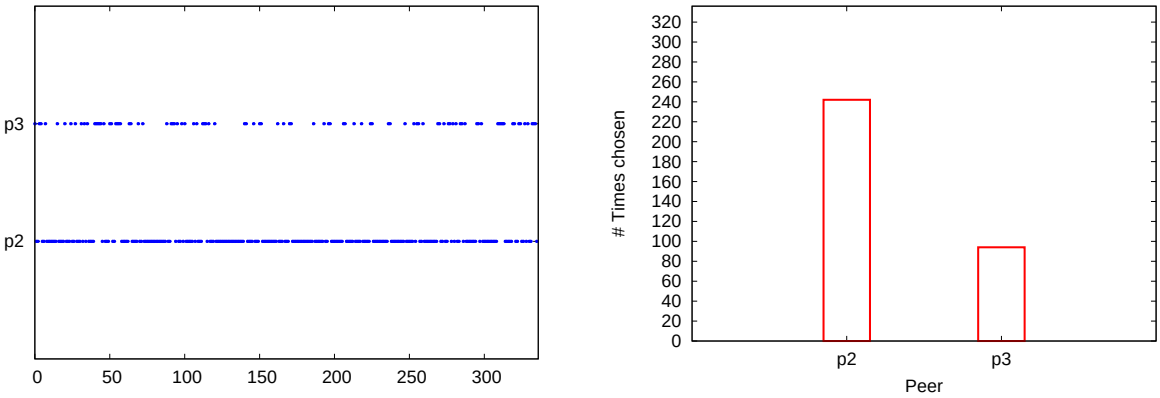
Figure 8 and 9 present the variation of the $S_{x.y}$ for both heavily/ less heavily loaded links. These figures have been calculated by means of the formulas explained in Section 3.1, and applied to the mathematical distributions mentioned above. From Tables 1 and 2, $S_{2.1} = I_{p2}^{p1}$, and $S_{3.1} = I_{p3}^{p1}$. It can be seen that the network performance affects the HRI, as was expected. A higher HRI is better, as it means that the peer is powerful and well connected. Also, when the link is not heavily loaded, S takes higher values and has a greater spread. Conversely, when the link is heavily loaded, more values are grouped together at the bottom



(a) Goodness function for peer $p2$ (link $p1 - p2$ unloaded).

(b) Goodness function for peer $p3$ (link $p1 - p3$ loaded).

Figure 10: Variation of the goodness function.



(a) Peer chosen.

(b) Aggregate count for peer selection.

Figure 11: Peer chosen (link $p1 - p2$ unloaded, link $p1 - p3$ loaded).

of the figure. Also, for Figure 9, $S_{2.2} = I_{P_4}^{P_2} + I_{P_5}^{P_2}$, and $S_{3.2} = I_{P_6}^{P_3} + I_{P_7}^{P_3}$, which means that to calculate $S_{2.2}$ and $S_{3.2}$, both heavily and less heavily used links are used.

Figure 10 shows the variation of the goodness function for the 2 neighbors of peer $p1$. Recall that the link between $p1$ and $p2$ is unloaded, and the link between $p1$ and $p3$ is loaded. It can be seen that the goodness function for $p2$ has higher values, and for $p3$ it has more values grouped at the bottom of the figure. Thus, peer $p2$ will be chosen more often than $p3$. This is depicted in Figure 11. Figure 11 (a) shows the peer that is chosen each time, and Figure 11 (b) shows an aggregate count of how often a particular peer was chosen. Peer

$p2$ is chosen 242 times out of 336 (around 72%), and peer $p3$ is chosen 94 times (around 28%).

4.2 Users' point of view

A complementary evaluation from a users' point of view is now presented. This approach is compared with other approaches from literature, namely *GriDM* and *flooding*, which were explained in Section 2. The aim of such a comparison is to emphasise that the network is an important resource that influences the performance received by users in a Grid. Thus, approaches that do not consider the network will not perform as efficiently as possible. Besides, when a query must be forwarded, the process of finding a suitable destination must be performed in a scaleable manner, so that it can efficiently fit into such a dynamically changing environment. Furthermore, considering only the direct neighbors of an administrative domain (instead of the whole Grid system) considerably reduces the scope of the search, making the approach more scaleable.

4.2.1 Experiments and results

A network scenario based on the EU DataGRID Testbed has been created, as shown in Figure 12 [21]. The original topology has been modified (3 links have been removed) to avoid loops when constructing Routing Indices (the issue of keeping HRI working and avoiding loops has already been treated in [11], but this is not related to our Grid meta-scheduling proposal). The topology shows eleven computing resources spanning several locations in Europe. Each location is an administrative domain, with the structure shown in Figure 3. For the sake of clarity, only routers and computing resources are depicted.

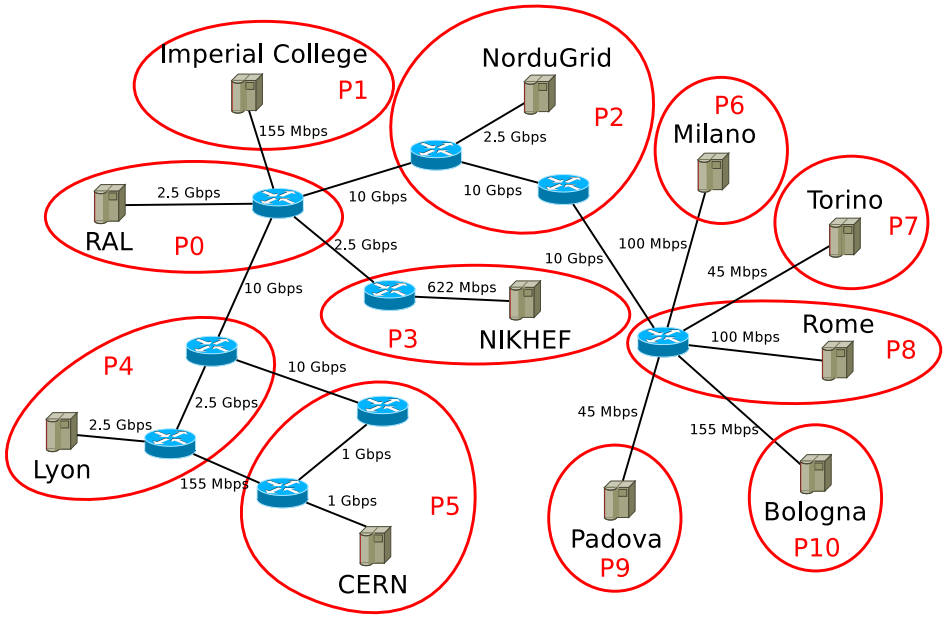


Figure 12: EU DataGRID Testbed 1.

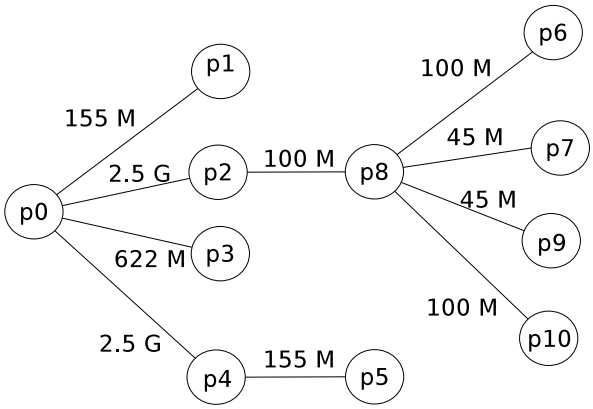


Figure 13: Peer-to-peer topology.

Boundaries between administrative domains are shown in circles in Figure 12, and the bandwidth of the link connecting the GNB is the same as that of the computing resource in that domain. Hence, the connectivity structure leads to the P2P topology depicted in Figure 13, where links between peers are the bottleneck of the network paths between GNBs. From now on, link bandwidths mentioned in this section are those appearing in Figure 13. The three proposals (ID-GNB, GriDM and flooding) have been implemented in GridSim. The following decisions have been made:

- Meta-scheduling is performed in *meta-scheduling rounds*, with an interval of 20 seconds.

- The monitoring of neighbors (ID-GNB and GriDM) is undertaken every 10 seconds. This has been chosen to allow 2 monitoring rounds to complete for every meta-scheduling round, so that more accurate information on the status of the neighbors is compiled.

- Peers accept a job to be executed in their local resource when the resource has idle CPUs at the moment the query reaches the peer. If a query reaches the peer more than once, this is done every time the query reaches the broker.

- Job queries in both GriDM and flooding experiments have a TTL, which has been chosen to allow queries to reach all the peers in the topology. For GriDM, it is equal to 11; for flooding, the TTL is 5.

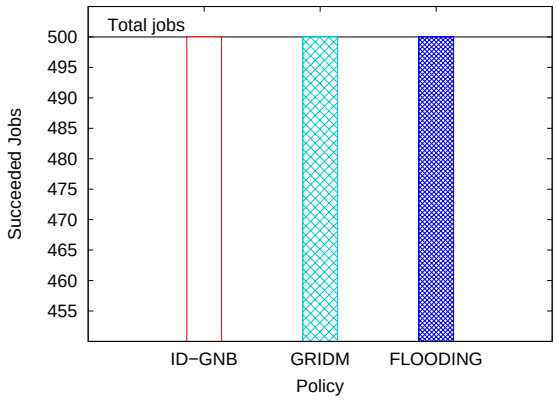
- For GriDM, the load of the computing resource provided by GridSim is used to decide which neighbor a query must be forwarded to. The least loaded computing resource is chosen each time.

- Several computing resources have full local (non-Grid) computing load, in the same way as in the intra-domain scenario. These computing resources are `Res_0`, `Res_1`, `Res_2`, `Res_3`, `Res_4`, and `Res_5`. Their local load covers around 95% of the computing power of the resources. That is, only around 5% of the computing power of each CPU at those resources is available for Grid users. For the other resources, the local load is nearly 0%. This has been decided in order to simulate a real Grid scenario, in which resources may have local load, that may differ between resources.

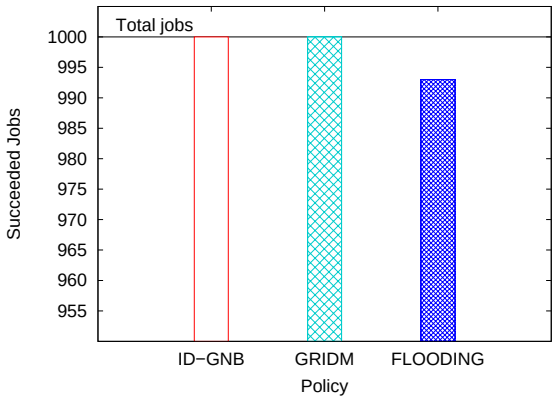
Table 5 summarizes the characteristics of simulated resources, which were obtained from a real LCG testbed [22]. The CPU rating is defined in MIPS (*Millions of Instructions Per Second*) as per SPEC (*Standard Performance Evaluation Corporation*) benchmark. The number of nodes for each resource have been scaled down by 10, due to memory limitations

Table 5: Resource specifications.

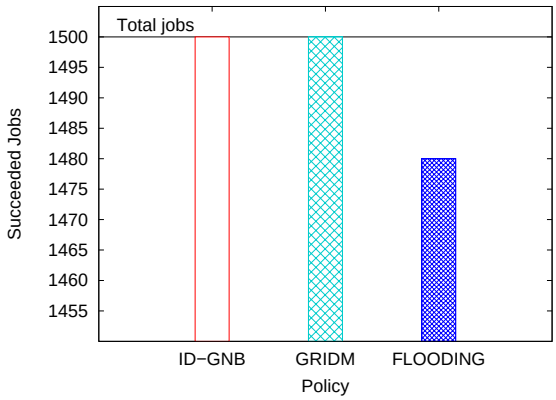
Peer ID	Res. (Location)	# Nodes	CPU Rating	# Users
0	RAL (UK)	41	49,000	12
1	Imp. College (UK)	52	62,000	16
2	NorduGrid (Norway)	17	20,000	4
3	NIKHEF (Netherlands)	18	21,000	8
4	Lyon (France)	12	14,000	12
5	CERN (Switzerland)	59	70,000	24
6	Milano (Italy)	5	70,000	4
7	Torino (Italy)	2	3,000	2
8	Rome (Italy)	5	6,000	4
9	Padova (Italy)	1	1,000	2
10	Bologna (Italy)	67	80,000	12



(a) 5 jobs per user



(b) 10 jobs per user



(c) 15 jobs per user

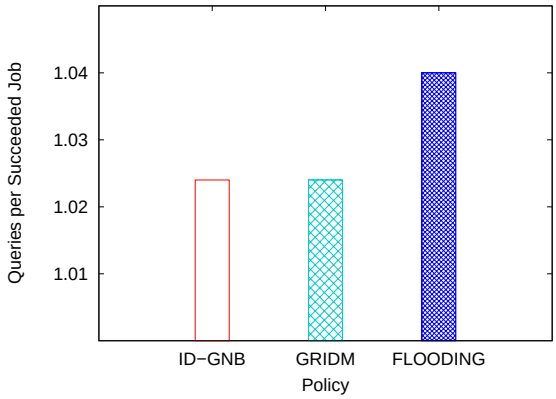
Figure 14: Number of succeeded jobs.

– otherwise the full experiment would require more than 2 GB of memory, and would take several weeks of processing. Finally, each resource node has four CPUs. For this experiment, 100 users were created and distributed among the locations, as shown in Table 5. Each user has multiple jobs, with the processing power of each job being $1,400,000$ *Million Instructions* (*MI*), which means that each job takes about 2 seconds if it is run on the CERN resource. Also, I/O file sizes are 24 MB. All jobs have the same parameters that are taken from ATLAS online monitoring and calibration system [4].

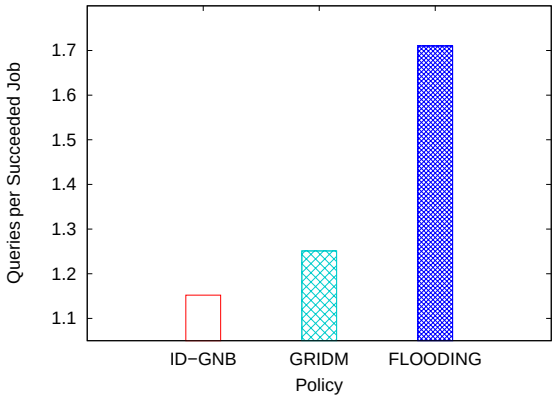
Our experiment is aimed at determining the behavior of the inter-domain meta-scheduling algorithm. Hence, the aim of the experiment is seeing how different algorithms affect the performance received by users in terms of number of queries forwarded, rate of queries per job, and the overall job execution time. Statistics related to the amount of data transferred between peers to keep HRIs up-to-date are also presented.

Figure 14 presents results regarding number of jobs that were successfully completed for each inter-domain meta-scheduling policy, as the number of jobs each user wants to run varies. It can be seen that there is no difference between the use of GriDM and ID-GNB approaches, since both of them can find a computing resource for all the jobs in all the experiments. On the other hand, as the number of jobs per user increases, there is an increase in the number of jobs in the flooding approach that cannot be allocated to any computing resource. Hence, those jobs remain unexecuted.

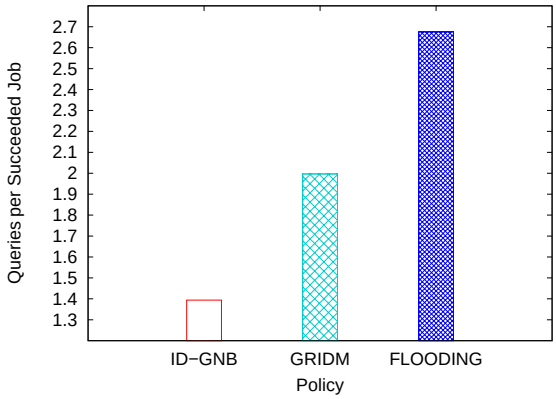
Now consider Figure 15, which depicts the number of queries forwarded per successfully completed job. This statistic has been calculated by dividing the actual number of queries forwarded by the number of successfully completed jobs. Hence, this statistic includes queries forwarded for those jobs which could not be executed. As expected, flooding requires more queries per job, since each peer forwards incoming queries it cannot fulfill to all its neighbors. With regard to ID-GNB and GriDM, ID-GNB shows the smallest values for this statistic, and the difference gets bigger as the number of jobs per user increases. For



(a) 5 jobs per user



(b) 10 jobs per user



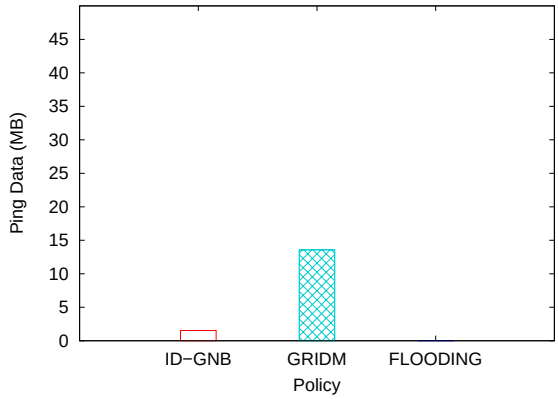
(c) 15 jobs per user

Figure 15: Number of queries per succeeded job.

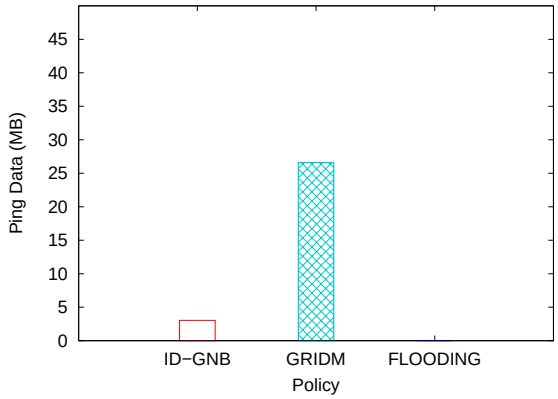
the case of 15 jobs per user, ID-GNB requires 30% less queries than GriDM, for the same amount of successful jobs.

Figure 16 shows the amount of data forwarded through the network, and includes ping requests made from peer to peer to support meta-scheduling. However, flooding does not require such information. As expected, ID-GNB requires less bytes to be forwarded, since it only requires information from the neighbors. Conversely, GriDM requires information from all the peers, thus increasing the amount of information forwarded through the network.

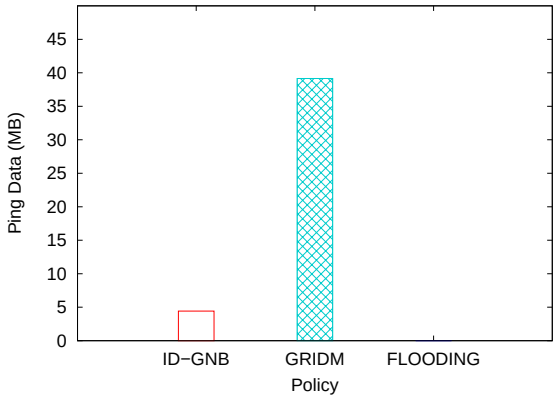
Figure 17 illustrates the number of bytes transferred through the network in queries. This is calculated as the sum of the size of each query that is propagated through the system. Each query has a number of parameters, including an identification for the user, another



(a) 5 jobs per user



(b) 10 jobs per user

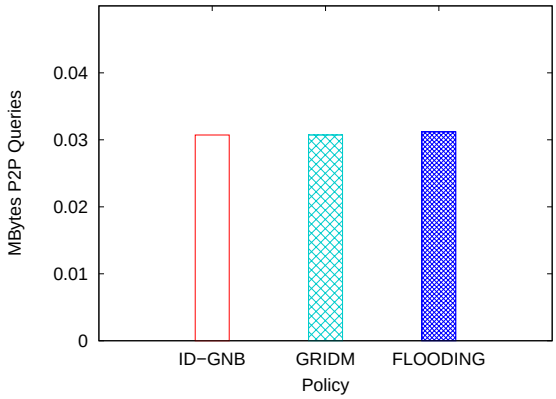


(c) 15 jobs per user

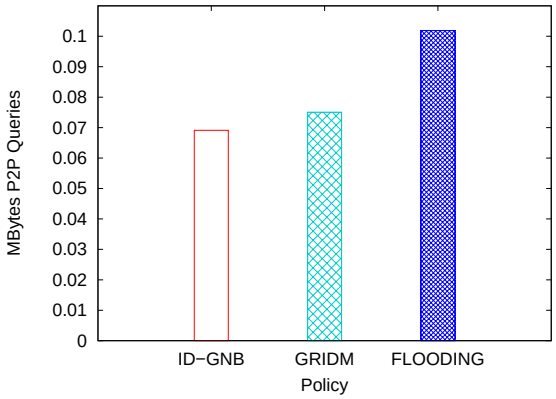
Figure 16: Data forwarded through the network when getting information.

identification for the computing resource chosen to run the job, TTL, identification of the GNB that forwarded the query to the current GNB, the size of the job, and sizes of input and output files. All of them make a job request object size of 60 bytes. This figure shows that when the number of jobs per user is small, there are negligible differences between strategies, but as the number of jobs per user increases, differences increase as well. As was expected, the flooding approach has the highest value for this statistic, followed by GriDM, and ID-GNB respectively.

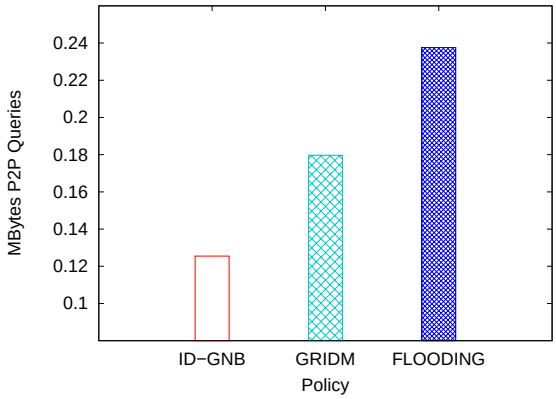
The number of jobs executed in each computing resource is depicted in Figures 18, 19 and 20. When there is a small number of jobs per user (see Figure 18), there are negligible



(a) 5 jobs per user



(b) 10 jobs per user



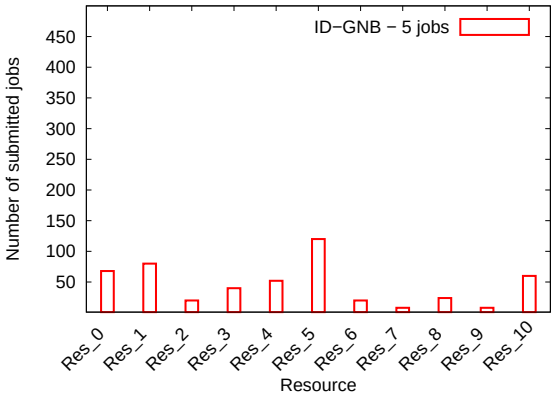
(c) 15 jobs per user

Figure 17: Data forwarded through the network in queries.

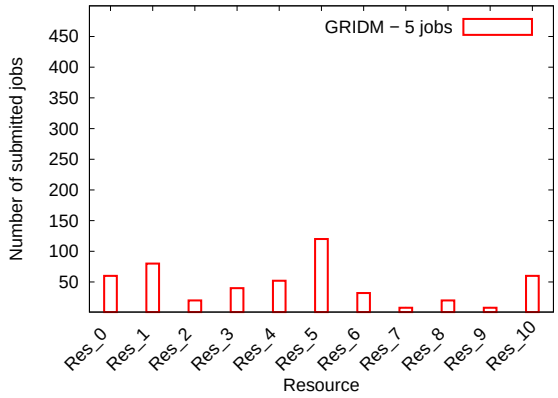
differences between strategies since all the resources run the same number of jobs in each case.

When each user has 10 jobs (see Figure 19), differences are still negligible. But when each user has 15 jobs (see Figure 20) differences clearly arise. In this last case, it can be seen that there are some computing resources that execute a high number of jobs for all the strategies (namely, `Res_0` (RAL), `Res_1` (Imperial College), `Res_5` (CERN) and `Res_10` (Bologna)). These are the most powerful computing resources, since they have more nodes than others.

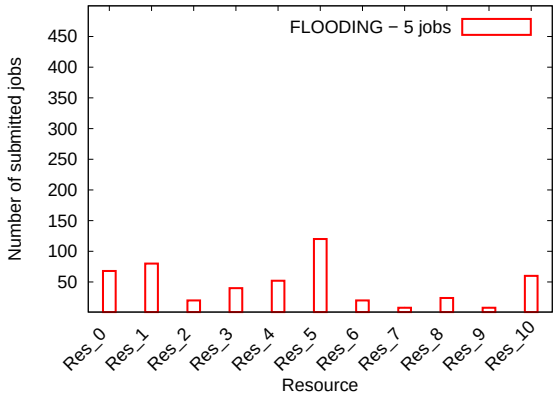
Apart from this observation, when ID-GNB is being used (see Figure 20 (a)), it can be seen that resource `Res_4` (Lyon) runs around 175 jobs, a considerably higher number of jobs



(a) ID-GNB



(b) GriDM

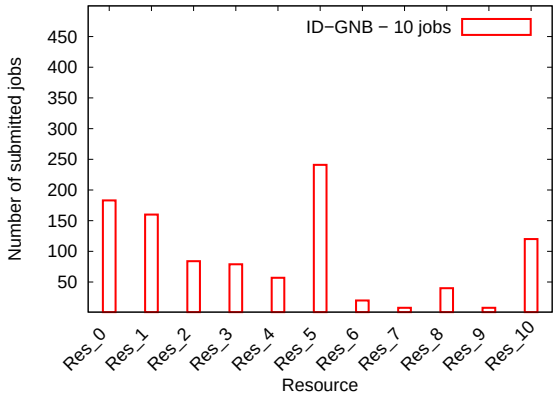


(c) Flooding

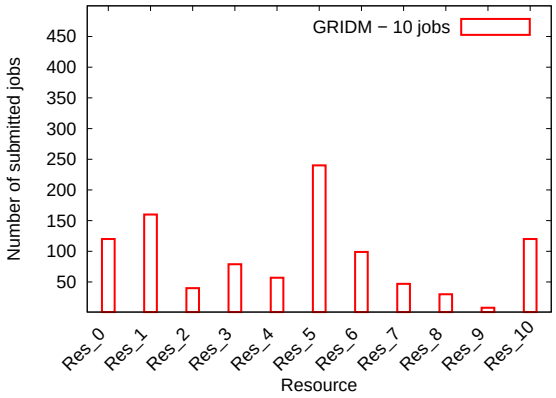
Figure 18: Number of jobs submitted to each computing resource, for 5 jobs per user.

than when GriDM is used (Figure 20 (b)). This is because this resource has a high bandwidth link of 2.5 GB (see Figure 13) which does not get overload. As Routing Indices are heavily influenced by the effective bandwidth of a link, this makes `Res_4` a good candidate to execute jobs.

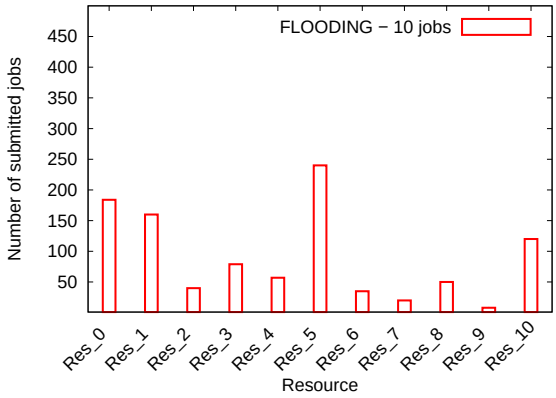
When ID-GNB is being used, resources `Res_6` (Milano) and `Res_7` (Torino) execute hardly any jobs, as opposed to the case when GriDM is running. This is explained by the fact that their links do not have good bandwidth, thus ID-GNB does not consider them as good candidates to run jobs. But they have low local load, thus GridM considers them as good candidates to run jobs.



(a) ID-GNB



(b) GriDM

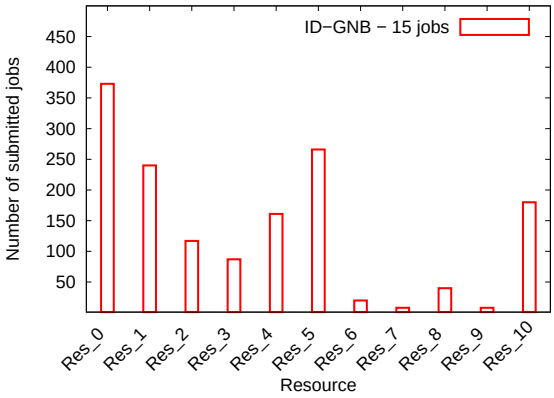


(c) Flooding

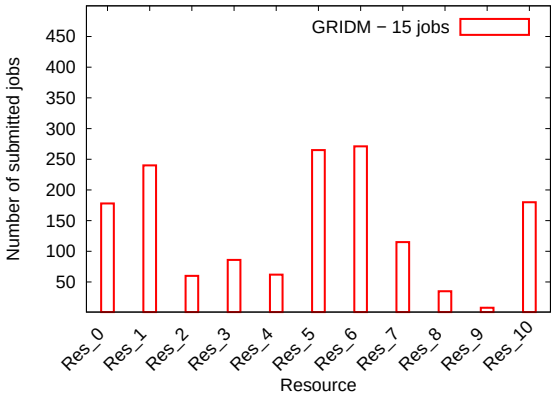
Figure 19: Number of jobs submitted to each computing resource, for 10 jobs per user.

For flooding (depicted in Figure 20 (c)), it can be seen that although computing resource `Res_8` (Rome) is less powerful, it executes more jobs than resource `Res_6` (Milano). This is because `Res_8` has 5 neighbors (as can be seen in Figure 13). So, it gets flooded with queries from them, and whenever its computing resource gets idle another request arrives and is accepted for execution in that resource.

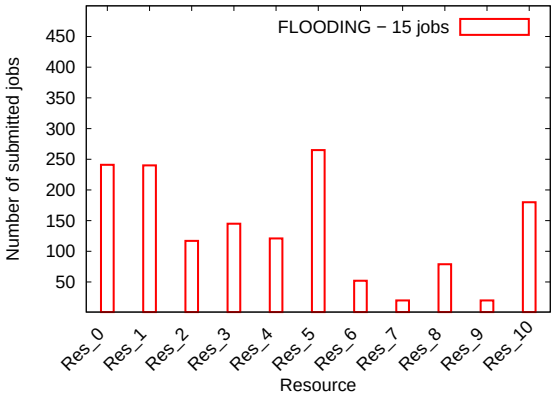
Figure 21 depicts the average network latencies of jobs. This statistic is calculated for each job, for example, the average network latency is calculated for job 0 for all the users. This is undertaken to demonstrate latencies of different jobs which were submitted in the same order (jobs with the same number for all the users). As before, when each user has 5 jobs (represented in Figure 21 (a)), differences between approaches are negligible (being



(a) ID-GNB



(b) GriDM

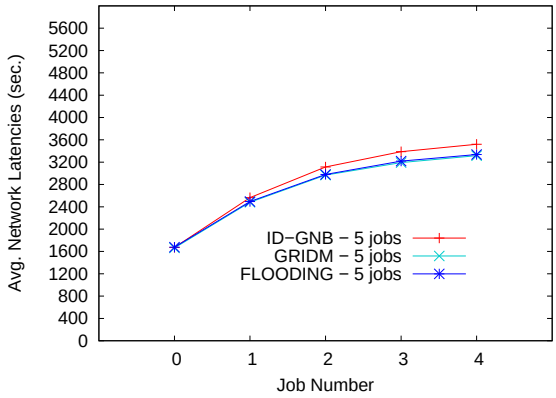


(c) Flooding

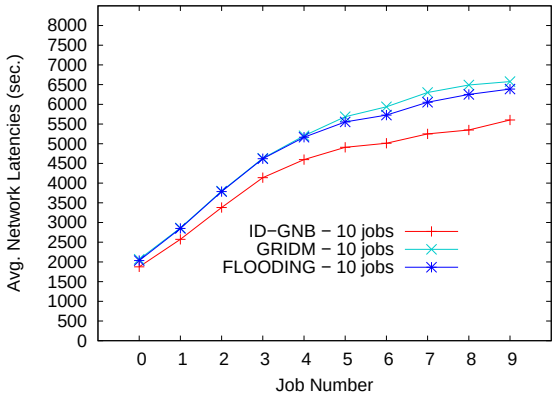
Figure 20: Number of jobs submitted to each computing resource, for 15 jobs per user.

ID-GNB slightly worse). When each user has 10 jobs (represented in Figure 21 (b)), GriDM and flooding approaches show similar results, and ID-GNB performs better. The reason is that the network is more loaded than before, thus the network performance becomes more important than in the previous case, and the resource workload becomes less important. This fact is supported by the number of queries per succeeded job (presented in Figure 15 (b)), which shows that GriDM needs more queries to find a suitable resource for each job than ID-GNB.

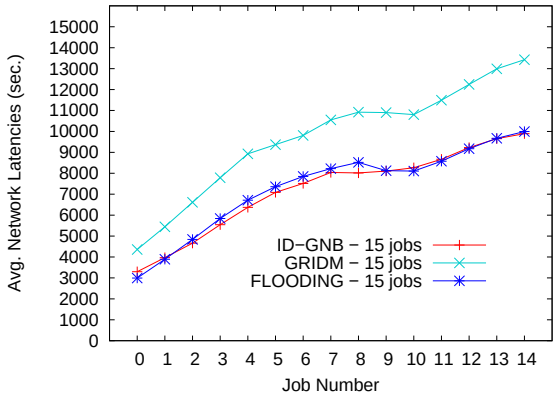
When users have 15 jobs (shown in Figure 21 (c)), the average network latency is higher for GriDM than for the other approaches. Since GriDM does not consider the network load, the resource chosen is not the most suitable. This is confirmed by the number of queries



(a) 5 jobs per user



(b) 10 jobs per user

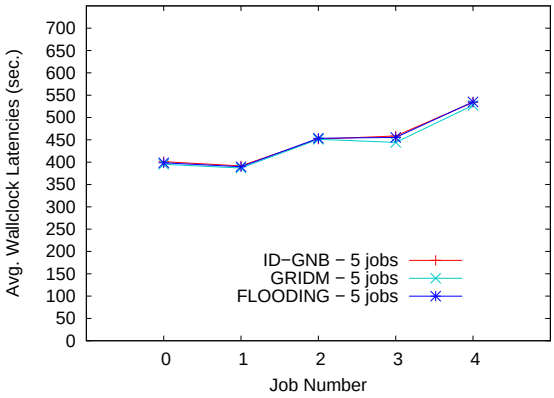


(c) 15 jobs per user

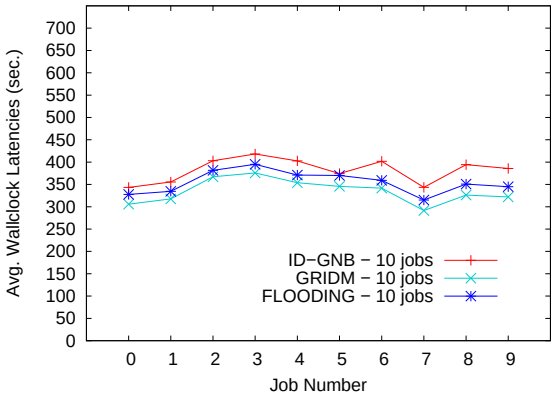
Figure 21: Average network latencies.

per successfully completed job (presented in Figure 15 (c)). Also, flooding presents similar latencies than ID-GNB, because of the nature of flooding. Recall that with flooding, every peer forwards each query to all its neighbors, thus it reaches a suitable computing resource, at the expense of a really high number of queries per succeeded job (presented in Figure 15 (c)) and a greater amount of interchanged information in queries (presented in Figure 17 (c)).

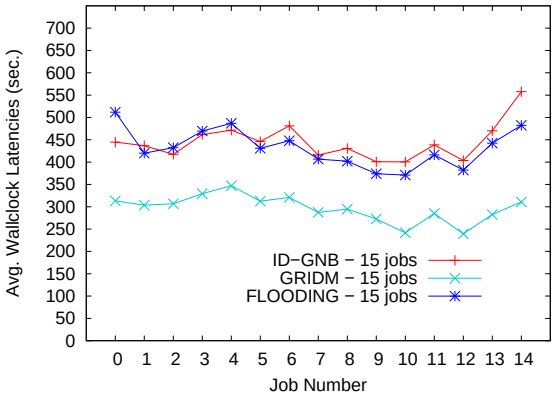
Figure 22 depicts the average wallclock latencies of jobs, and represents the total time jobs spend in computing resources, including waiting time (when no CPU is idle when the job arrives at the resource), and execution time. As before, when each user has 5 jobs (Figure 22 (a)) differences between strategies are negligible. When users have 10 jobs (Figure 22 (b)) differences start to arise, and ID-GNB performs slightly worse than the other



(a) 5 jobs per user



(b) 10 jobs per user

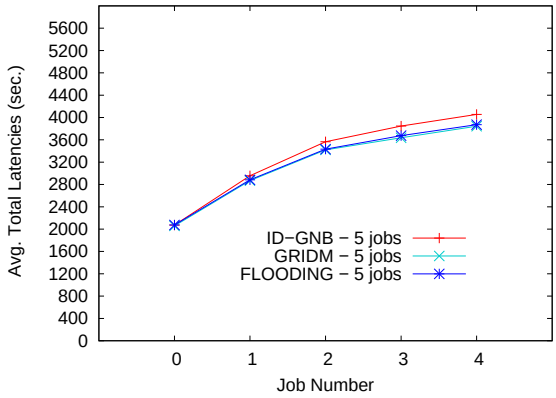


(c) 15 jobs per user

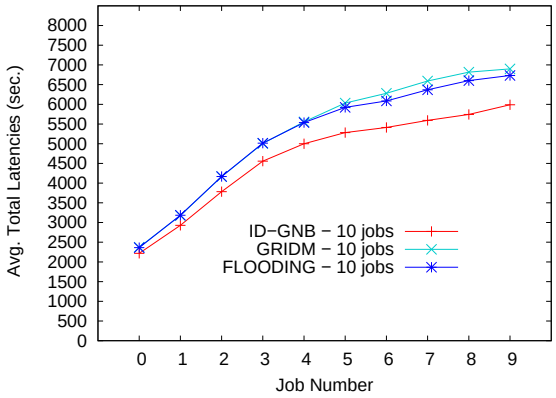
Figure 22: Average wallclock latencies.

strategies. When users have 15 jobs (Figure 22 (c)), differences are clearer, and GriDM shows the best results. This is explained by the fact that GriDM always chooses the least loaded computing resource. But differences are almost negligible (a few tens of seconds), compared with the differences in network latencies (shown in Figure 21).

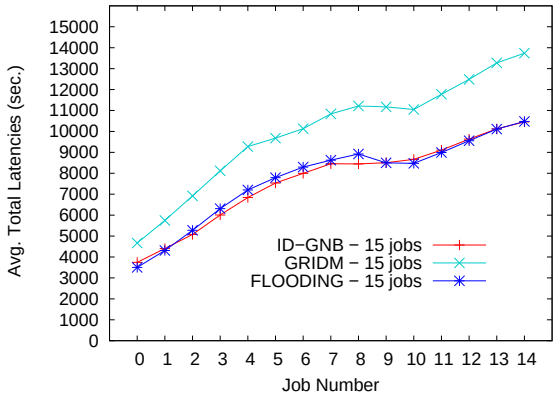
The last statistic presented is the average total latency for jobs, which is shown in Figure 23. This statistic includes the elapsed time since users submitted the job to the computing resource, until the output of the job reaches the user. It includes the transmission time, queueing time at the resource (if no CPU is idle at the moment), and the execution time of the job. Thus, it is the result of adding the statistics presented in Figures 21 and 22. They present similar tendencies as for the network latencies (presented in Figure 21).



(a) 5 jobs per user



(b) 10 jobs per user



(c) 15 jobs per user

Figure 23: Average total latencies.

The results of this evaluation show that ID-GNB outperforms both GriDM and flooding in terms of number of queries required for each job, and network and total latency times. ID-GNB achieves a better rate of successfully completed jobs and lower latencies, with less queries per job. Besides, less information must be sent through the network to keep the architecture working. This therefore demonstrates that ID-GNB is scalable, hence it is a more appropriate technique for realistic Grid environments.

5 Conclusions and future work

Grids are made of different administrative domains connected with each other. Thus, relations between domains are key in Grid computing and must be considered when performing meta-scheduling. An extension to an existing meta-scheduling framework has been proposed to allow network-aware multi-domain meta-scheduling based on peer-to-peer techniques.

More precisely, the proposal is based on *Routing Indices* (RI). This way we allow nodes to forward queries to neighbors that are more likely to have answers. If a node cannot find a suitable computing resource for a user's job within its domain, it forwards the query to a subset of its neighbors, based on its local HRI, rather than by selecting neighbors at random or by flooding the network by forwarding the query to all neighbors.

Results presented here demonstrate the better performance and the scalability of *Inter-Domain GNB, ID-GNB*. The results of the evaluation depict that our meta-scheduler outperforms existing proposals in terms of number of queries required for each job, network time and total latency. ID-GNB achieves better rate of succeeded jobs and better latencies, with less queries per job. Also, less information is transmitted through the network to keep the infrastructure working (which makes it scalable).

In this paper, a combination of data on the status of the network and the status of the computing resources are used to perform inter-domain meta-scheduling, but this may not be enough. For example, a job may have the following requirements: < OS = Linux, SW = Java 5, MatLab 7, HW = 200 GB available hard disk >. In this case, the most powerful unloaded computing resource, whose network is also powerful and unloaded cannot execute this job unless this computing resource fulfills the requirements of the job. Tryin to identify how job properties precisely match resource properties has already been the subject of

considerable research in matchmaking (e.g. Condor classads). Without the use of such approaches, utilizing the inter-domain scenario presented here still has limitations, as GNBs must decide which information to provide to their neighbor GNBs, and this must be done in an efficient and scalable manner. The use of a summarization process may be a useful approach to summarize capabilities of multiple resource properties or job classads, before spreading it to the neighbors. Thus, further research can be conducted following this direction.

Acknowledgement

This work has been jointly supported by the Spanish MEC and European Commission FEDER funds under grants “Consolider Ingenio-2010 CSD2006-00046” and “TIN2006-15516-C04-02”; jointly by JCCM and Fondo Social Europeo under grant “FSE 2007-2013”; and by JCCM under grants “PBI08-0055-2800” and “PII1C09-0101-9476”.

References

[1] Adami, D., Giordano, S., Repeti, M., Coppola, M., Laforenza, D., and Tonellotto, N. Design and implementation of a Grid network-aware resource broker. In *Proc. of the Intl. Conference on Parallel and Distributed Computing and Networks (PDCN)* (Innsbruck, Austria, 2006).

[2] Al-Ali, R., Sohail, S., Rana, O., Hafid, A., von Laszewski, G., Amin, K., Jha, S., and Walker, D. Network QoS provision for distributed Grid applications. *Intl. Journal of Simulations Systems, Science and Technology, Special Issue on Grid Performance and Dependability 5*, 5 (December 2004), 13–28.

[3] Anjum, A., McClatchey, R., Stockinger, H., Ali, A., Willers, I., Thomas, M., Sagheer, M., Hasham, K., and Alvi, O. DIANA scheduling hierarchies for optimizing bulk job scheduling. In *Proc. of the Second Intl. Conference on e-Science and Grid Computing* (Amsterdam, Netherlands, 2006).

[4] ATLAS online monitoring and calibration system. Web page at `http://dissemination.interactive-grid.eu/applications/HEP`, 2009.

[5] Baker, R., Gommans, L., McNab, A., Lorch, M., Ramakrishnan, L., Sarkar, K., and Thompson, M. R. Conceptual Grid authorization framework and classification. In *Global Grid Forum Working Group on Authorization Frameworks and Mechanisms* (2003).

[6] Barrère, F., Benzekri, A., Grasset, F., Laborde, R., and Nasser, B. Automated inter-domain security policy generation. In *Proc. of the 11th Workshop of the HP OpenView University Association* (Paris, 2004).

[7] Bhatti, S., Sørensen, S., Clark, P., and Crowcroft, J. Network QoS for grid systems. *The Intl. Journal of High Performance Computing Applications 17*, 3 (August 2003), 219 – 236.

[8] Caminero, A., Rana, O., Caminero, B., and Carrión, C. Performance evaluation of an autonomic network-aware metascheduler for Grids. *Concurrency and Computation: Practice and Experience (in press, accepted on Oct. 17, 2008).*

[9] Cao, J., Cleveland, W., Lin, D., and Sun, D. *Nonlinear Estimation and Classification*. Springer Verlag, New York, USA, 2002, ch. Internet traffic tends toward Poisson and independent as the load increases.

[10] Carpenter, B. E., and Janson, P. A. Abstract interdomain security assertions: A basis for extra-Grid virtual organizations. *IBM Systems Journal 43*, 4 (2004), 689–701.

[11] Crespo, A., and Garcia-Molina, H. Routing Indices For Peer-to-Peer Systems. In *Proc. of the Intl. Conference on Distributed Computing Systems (ICDCS)* (Vienna, Austria, 2002).

[12] de Assunção, M. D., Buyya, R., and Venugopal, S. InterGrid: A case for internetworking islands of Grids. *Concurrency and Computation: Practice and Experience* (2007).

[13] di Vimercati, S. D. C., and Samarati, P. Access control in federated systems. In *Proc. of the 1996 Workshop on New Security Paradigms* (Lake Arrowhead, USA, 1996).

[14] Fitzgerald, S., Foster, I., Kesselman, C., von Laszewski, G., Smith, W., and Tuecke, S. A directory service for configuring high-performance distributed computations. In *Proc. 6th Symposium on High Performance Distributed Computing (HPDC)* (Portland, USA, 1997).

[15] Foster, I. T. The anatomy of the Grid: Enabling scalable virtual organizations. In *Proc. of the 1st Intl. Symposium on Cluster Computing and the Grid (CCGrid)* (Brisbane, Australia, 2001).

[16] G-lambda Project. Web page at `http://www.g-lambda.net`, 2009.

[17] Gnutella. Web page at `http://rfc-gnutella.sourceforge.net/`, 2009.

[18] Greenwald, S. A new security policy for distributed resource management and access control. In *Proc. of the 1996 Workshop on New Security Paradigms* (Lake Arrowhead, USA, 1996).

[19] GridPP, real time monitor. Web page at `http://gridportal.hep.ph.ic.ac.uk/rtm/`, 2009.

[20] Gu, X., and Nahrstedt, K. A Scalable QoS-Aware Service Aggregation Model for Peer-to-Peer Computing Grids. In *Proc. 11th Intl. Symposium on High Performance Distributed Computing (HPDC)* (Edinburgh, UK, 2002).

[21] Hoschek, W., Janez, F. J., Samar, A., Stockinger, H., and Stockinger, K. Data management in an international Data Grid project. In *Proc. of the 1st Intl. Workshop on Grid Computing* (Bangalore, India, 2000).

[22] LCG Computing Fabric Area. Web page at `http://lcg-computing-fabric.web.cern.ch`, 2009.

[23] LCG (LHC Computing Grid) Project. Web page at `http://lcg.web.cern.ch/LCG`, 2009.

[24] Magaña, E., and Serrat, J. Distributed and heuristic policy-based resource management system for large-scale Grids. In *Proc. of the First Intl. Conference on Autonomous Infrastructure, Management and Security, (AIMS)* (Oslo, Norway, 2007).

[25] Massie, M. L., Chun, B. N., and Culler, D. E. The Ganglia distributed monitoring system: design, implementation, and experience. *Parallel Computing 30*, 5-6 (2004), 817–840.

[26] McCloghrie, K., and Rose, M. T. Management information base for network management of TCP/IP-based internets:MIB-II. RFC 1213, March 1991.

[27] McGuffin, L. J., Smith, R. T., Bryson, K., Sorensen, S. A., and Jones, D. T. High throughput profile-profile based fold recognition for the entire human proteome. *BMC Bioinformatics 7* (June 2006), 288+.

[28] O'Brien, A., Newhouse, S., and Darlington, J. Mapping of Scientific Workflow within the E-Protein Project to Distributed Resources. In *UK e-Science All-hands Meeting* (Nottingham, UK, 2004).

[29] Phosphorus Project. Web page at http://www.ist-phosphorus.eu/, 2009.

[30] Puppin, D., Moncelli, S., Baraglia, R., Tonellotto, N., and Silvestri, F. A Grid Information Service Based on Peer-to-Peer. In *Proc. of the 11th Intl. Euro-Par Conference* (Lisbon, Portugal, 2005).

[31] Rekhter, Y., Li, T., and Hares, S. A Border Gateway Protocol 4 (BGP-4). Internet proposed standard RFC 4271, January 2006.

[32] Roy, A. *End-to-End Quality of Service for High-End Applications*. PhD thesis, Dept. of Computer Science, University of Chicago, 2001.

[33] Sohail, S., Pham, K. B., Nguyen, R., and Jha, S. Bandwidth broker implementation: Circa-complete and integrable. Tech. rep., School of Computer Science and Engineering, The University of New South Wales, 2003.

[34] Talia, D., and Trunfio, P. A P2P Grid services-based protocol: Design and evaluation. In *Proc. of the 10th Intl. Euro-Par Conference* (Pisa, Italy, 2004).

[35] VIOLA : Vertically Integrated Optical Testbed for Large Application in DFN. Web page at http://www.viola-testbed.de/, 2009.

[36] Waldrich, O., Wieder, P., and Ziegler, W. A meta-scheduling service for co-allocating arbitrary types of resources. In *Proc. of the 6th Intl. Conference on Parallel Processing and Applied Mathematics (PPAM)* (Poznan, Poland, 2005).

[37] Wang, Y., and Hu, J. Adaptive trust management framework in P2P Grid computing. *Wuhan University Journal of Natural Sciences 1*, 13 (2008), 33–36.

[38] Xiong, Z., Yang, Y., Zhang, X., Chen, F., and Liu, L. Integrating Genetic and Ant algorithm into P2P Grid resource discovery. In *Proc. of the Third Intl. Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP)* (Washington, DC, USA, 2007).

[39] Xiong, Z., Yang, Y., Zhang, X., and Zeng, M. Grid resource aggregation integrated P2P mode. In *Proc. of the 4th Intl. Conference on Intelligent Computing (ICIC)* (Shanghai, China, 2008).

[40] Xu, D., Nahrstedt, K., and Wichadakul, D. QoS-aware discovery of wide-area distributed services. In *Proc. of the First Intl. Symposium on Cluster Computing and the Grid (CCGrid)* (Brisbane, Australia, 2001).

[41] Zhao, Y., An, Y., Wang, C., and Gao, Y. A QoS-satisfied interdomain overlay multicast algorithm for live media service Grid. In *Proc. of the 4th Intl. Conference on Grid and Cooperative Computing (GCC)* (Beijing, China, 2005).

[42] Zheng, W., Hu, M., Yang, G., Wu, Y., Chen, M., Ma, R., Liu, S., and Zhang, B. An efficient web services based approach to computing Grid. In *Proc. of Intl. Conference on E-Commerce Technology for Dynamic E-Business, (CEC-East)* (Washington, DC, USA, 2004).