

A Novel Architecture for Grid Information Systems

Zoltán Balaton, Gábor Gombás, Zsolt Németh

MTA SZTAKI Computer and Automation Research Institute

P.O. Box 63., H-1518 Hungary

E-mail: {balaton, gombasg, zsnemeth}@sztaki.hu

1. Introduction

Grids facilitate large-scale distributed resource sharing. In such an environment a priori knowledge is not available due to the diversity of resources and their dynamics. Information in the grid ranges from static to highly dynamic. We focus on information required for resource brokering, i.e. how appropriate resources for applications can be found and what requirements it poses for an information system.

One of the representative grid information systems is the Globus Metacomputing Directory Service (MDS) currently in its second incarnation called MDS-2 [1]. It performs better than its predecessor with respect to scalability and performance, yet it still may fail to satisfy every needs of a resource broker. In MDS-2 Grid Information Index Service (GIIS) servers (specialised indexes of the available resources) are used to provide a central repository for clients to facilitate searches. They collect, cache and manage information from resources belonging to a so called virtual organisation. It works well in small virtual organisations but it is not scalable for thousands of resources.

One of the reasons is that MDS-2 aims to provide a uniform information and monitoring system, i.e. handles static and dynamic data in the same way. Since frequently changing data becomes stale quickly, the tree of the GIIS servers forming a cache chain cannot be tall. Also as a consequence of the pull data delivery model used in MDS-2 the amount of data moved is proportional to the number of queries. This limits scalability and efficiency.

Furthermore, MDS-2 uses the LDAPv3 [4] protocol where information is organised as a hierarchical tree called Directory Information Tree (DIT) defined by the distinguished names (DN) of the entries. Although the LDAP query language permits searches based on any properties of the entries, a query that does not match the hierarchy of the distributed LDAP database can be very inefficient. This is because the DIT is distributed along the hierarchy defined by the DNs and one can only restrict the set of servers to be searched in terms of this hierarchy (with the base and scope parameters in the query). Thus, most queries which

do not match the hierarchy predefined by the DNs of entries result in querying every server storing the distributed LDAP database. Therefore, the distributed LDAP database can only answer certain searches efficiently that match the predefined hierarchy.

The solution proposed in this paper tries to tackle with these problems. Key points are introduced in the following.

2. Proposed information system architecture

Figure 1 shows the basic components of the system (drawn with solid lines). The parts drawn with dashed lines show how it is integrated with the rest of the grid.

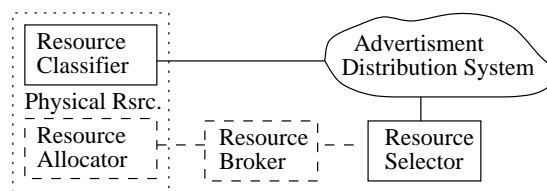


Figure 1. Basic components of the system

The **Resource Classifier** (RC) periodically produces an advertisement about parameters of the resource. A RC belongs to a site and may handle more than one physical resource that are under the same administrative control.

The **Advertisement Distribution System** (ADS) is responsible for distributing the advertisements from the Resource Classifiers to Resource Selectors. It must be distributed, robust, fault tolerant and should minimise the network traffic required to dispatch advertisements.

The **Resource Selector** (RS) is the consumer of the advertisements distributed by the ADS. It consists of four major components depicted in Figure 2.

The *Filter* discards advertisements that do not have a proper syntax and an acceptable digital signature, or those that are known to never match local needs. This step reduces the size of the database and increases its efficiency.

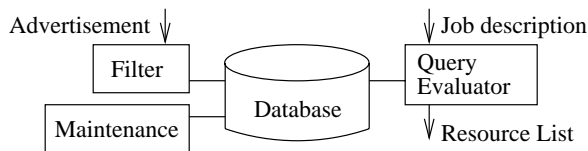


Figure 2. Structure of the Resource Selector

The *Database* stores all filtered information and provides an efficient way to perform complex search operations locally. It is not directly visible from outside so it can be chosen to fit the possible queries the best. The most likely candidate is an SQL-based RDBMS, also suggested by [2].

The *Maintenance* subsystem handles database maintenance tasks such as removing expired advertisements.

The *Query Evaluator* receives a job description from the Resource Broker (RB) and extracts the list of matching resources from the database. It also sorts the list based on some heuristics that can depend on several factors including the probability of the resource having enough free capacity to run the job. The sorted list is then passed to the RB for further processing and to select the resource to run the job.

Advertisements carry information used by the Resource Selector to decide if a resource matches a request. They may contain:

- Administrative information (creation and expiration date, digital signature)
- Resource description (architecture, operating system)
- Availability information (maximum number of processors offered, forecast of resource allocation, etc.)
- Policy information (e.g. the accounting method)

Resource brokers require precise and timely information about resources. Advertisements can include features of the resource but they should not contain actual state information that expires quickly. Additionally, state information might not be interpreted without the context of the resource and also raises security concerns. Therefore, highly dynamic data are estimated based on a stochastic model and only characteristic parameters of the model are included in advertisements. The stochastic model realises the necessary context where data can be interpreted and in such a way the RS can estimate actual state information.

If we look at USENET News it turns out that it can implement the functionality of ADS described above:

- Resource Classifiers act as news user agents and post the resource advertisements to the nearest news server.
- The News network distributes the posted articles to all news servers.
- The Filter of a RS acts as a simple news client.

The News system is completely distributed and the network traffic is optimal since information spreads along a spanning tree only [3]. It is also very robust and scalable.

3. Conclusion

Our approach aims at supporting resource brokering in large-scale grid applications. It has four fundamental points. First, highly dynamic information is not stored in the information system, rather replaced by some less frequently changing parameters, i.e. made pseudo-static. Based on these parameters and a stochastic model, the actual data can be estimated. Second, information is propagated in advertisements from resources to consumers using a push data delivery model. At the client side it is stored in relational databases that can be tailored to local needs and queried efficiently locally. Third, the network traffic is proportional to the number of resources and not to the number of queries. The proposed system ensures better information locality and utilises local resources only. Fourth, neither the RC nor the RS needs to know anything about each other's locations but they can rely on the ADS. Security can be addressed by RCs signing advertisements. The RSs can then decide which RCs to trust and can configure their Filters to drop advertisements coming from untrusted sources. When implemented using USENET News, all these features can be realised.

Future work includes the further refinements in the specification of advertisements and query language. It is anticipated and currently being checked by simulation that a trade-off can be found between the frequency of advertisements and the precision of the data estimation.

4. Acknowledgement

We would like to thank Péter Kacsuk, Norbert Podhorszki, Ferenc Szalai and Ferenc Vajda for their valuable comments, discussions and help in forming these ideas.

This work was partially supported by the European Commission under contract number IST-2000-25182, the Hungarian Scientific Research Fund (OTKA) under grant number T032226 and the Research and Development Division of the Hungarian Ministry of Education under contract number IKTA-00111/2000.

References

- [1] K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman: Grid Information Services for Distributed Resource Sharing. Proc. 10th IEEE International Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press, 2001.
- [2] S. Fisher: Relational Model for Information and Monitoring. Technical Report GWD-Perf-7-1, GGF, 2001.
- [3] B. Kantor, P. Lapsley: Network News Transfer Protocol, IETF RFC 977, February 1986.
- [4] M. Wahl, T. Howes, S. Kille: Lightweight Directory Access Protocol (v3), IETF RFC 2251, December 1997.