

# Q-GSM: QoS Oriented Grid Service Management\*

Hanhua Chen, Hai Jin, Feng Mao, and Hao Wu

Cluster and Grid Computing Lab,  
Huazhong University of Science and Technology,  
Wuhan, 430074, China  
hjin@hust.edu.cn

**Abstract.** Effective and efficient *Quality of Service* (QoS) management is critical for a service grid to meet the requirements of both grid users and service providers. We introduce Q-GSM, a scalable framework for service management in grid environments, to address this problem. Main characteristics of our framework are 1) grid level resource reservation mechanism; 2) mapping application level service quality to sufficient quantitative reserved resources through negotiating with resources reservation interface; and 3) contract-like service level agreement management of grid services.

## 1 Introduction

Service oriented grid architecture presents a vision of an integrated approach to supporting both e-science and e-business [1]. In a highly competitive service grid environment, *Quality of Service* (QoS) is one of the substantial aspects for differentiating among similar service providers. QoS problem results from resource sharing among applications. Unless the resource has provisioning as a fundamental capability, predictable quality of service cannot be delivered to a grid consumer.

Research on QoS in grid services is still at its infancy. [2] introduces an efficient wide-area distributed service discovery strategy. It uses caching and propagation of discovery results with client QoS feedbacks in the discovery server hierarchy. QCWS project [3] is a web service architecture that guarantees QoS of multimedia web services by deploying a QoS Broker between web service clients and web service providers. However QCWS does not relate QoS of web services with the supporting resources. G-QoSM [4] provides an application level service management system.

Q-GSM aims at using existing reservation mechanisms to map service level capacity into sufficient quantitative reserved resources. Q-GSM enables service capacity predictability, which is required for proper service scheduling in a service grid. The QoS obligation terms described in a SLA are negotiated before use and monitored during the service lifetime. Q-GSM is also responsible for managing the lifecycle of the service level agreement for a grid service.

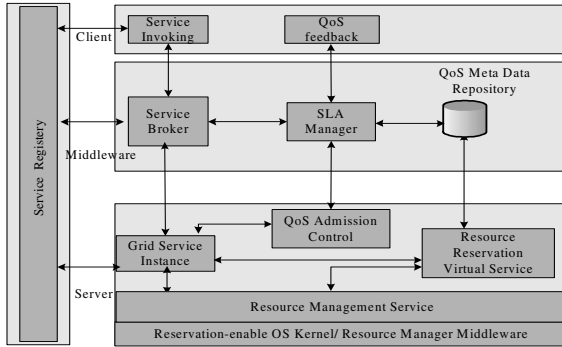
---

\* This paper is supported by National Science Foundation under grant 60273076 and 90412010, ChinaGrid project from Ministry of Education, and the National 973 Key Basic Research Program under grant No.2003CB317003.

## 2 Q-GSM Architecture

In our previous work [5], we proposed the model of *Service Virtualization* in grid environment. The main idea of *Service Virtualization* is encapsulating diverse implementations, which may have different QoS capacities, behind a common service interface that standardizes the business function. Based on this model, we design Q-GSM to provide a more available framework to guarantee QoS of grid services.

Fig.1 shows the architecture of the framework of G-QSM, with three components: client components, grid middleware components, and grid server components.



**Fig. 1.** Architecture of Q-GSM

Grid server is the host environment of grid services. It consists of an admission controller, reservation enabled resource management service, and a virtual resource reservation service. They work together to provide sufficient resources to guarantee the QoS of grid services.

Admission controller is responsible to create grid service instances. During the creation of the grid service instances, the service instance interacts with the resource reservation virtual service to bind sufficient resource to the service instance. The server admits a request only when it can reserve a sufficient amount of resources to achieve the desired service quality.

The resource management service can be any type of the existing reservation enabled resource management system, which has the ability to assign different amount of system resources to different service instances. The virtual resource reservation service serves as an abstract interface to different resource reservation system. It is responsible for function semantics abstraction and protocol transmission.

Different reservation systems are implemented at different levels. *Globus Architecture for Reservation and Allocation* (GARA) provides a flexible architecture that makes it possible for the application to make advance reservation for networks, CPUs, disks, and so on. Q-GSM uses existing reservation mechanisms to map service level capacity into sufficient quantitative reserved resources.

The main responsibility of the service middleware is service virtualization and service QoS management. Service middleware consists of a service broker, *Service Level Agreement* (SLA) manager and the QoS metadata repository.

We deploy a service broker of every common service on the grid as a virtual service and the services located on local resources are called physical services. A broker acts as an intermediary between a client and a set of physical services by providing a single point of submission for request. The requests with QoS requirements from the client to the service broker are scheduled to physical services according to QoS capacity of physical services in the QoS metadata repository.

The QoS metadata repository helps to predictable quality of service delivered to a grid consumer. It is responsible for querying the QoS capacity of the every physical service registered to the virtual service and identifying those services whose QoS capacities match those desired by the service consumer. The selection is driven by high-level application criteria, such as time to completion, reliability, or cost.

The distributed nature of the grid environment makes precise determination of the service state impossible. The QoS metadata repository does not imply any commitment. QoS metadata repository only gives a prediction of the QoS capacity of physical services. To actually allocate the request, the SLA manager will interact with every candidate server to make sure that the service provider will guarantee the QoS of the request and negotiate a contract like SLA. The QoS metadata is assumed to be specified and be updated by the service provider according to the resource status and the feedbacks of the client are also used to adjust the QoS metadata.

The client of the virtual service is for the end user to send their service request with QoS requirements to the service broker. When the client receives the response of the service, it sends the feedback information of the QoS to the SLA manager. Such feedback will help to enhance the precise of the predication. SLA manager determines whether the QoS requirement of the user has been met or not.

Fig 2 shows a sequence diagram among service client, service middleware component and server side component.

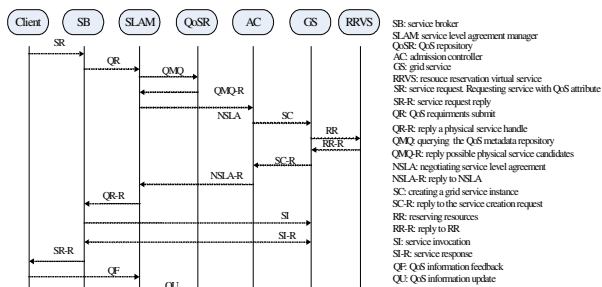
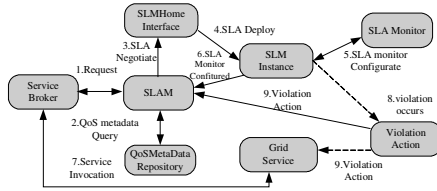


Fig. 2. Sequence Diagram among the Components of the Framework

### 3 SLA Management of Grid Services

Service management in Q-GSM is SLA-driven. The SLAM dynamically determines whether enough spare capacity is available to accommodate additional SLAs. Fig.3. shows the management framework for the SLA instances during the lifetime. When there is a request from the service broker, the SLAM queries the QoS metadata repository and selects candidate services with the proper QoS metadata.

SLA manager is responsible for negotiating SLA contracts between client and service provider. It is also responsible for SLA management during the service lifetime. QoS objective terms specified in SLAs are monitored during the service lifetime and actions are taken upon violation occurs.



**Fig. 3.** Service Level Agreement Management Framework

The SLAM is designed in Factory pattern. SLMHomeInterface creates a SLM instance to manage the deployed SLA instance. SLM instance has a lifecycle longer than SLA instance. It configures the SLA monitor service, which has been specified in SLA document. During the access to a grid service, SLA Monitor monitors the SLA Obligations. SLM Instance evaluates the obligations whether they are violated or not. Once violation occurs ViolationAction service will take necessary action. ViolationAction is a grid service specified in SLA. The monitor interfaces can be implemented by service provider or be treated as out-sourcing to third-party.

## 4 Conclusions

We propose Q-GSM, a QoS oriented grid service management framework. We design an abstract interface to existing resources reservation mechanisms and the management system of the SLA. Our framework makes it easier to build complex application with QoS management.

## References

- [1] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, "Grid Service for Distributed System Integration", *IEEE Computer*, Vol.35, No.6, June 2002, pp.37-46.
- [2] D. Y. Xu, K. Nahrstedt, and D. D. Wichadakul, "QoS-Aware Discovery of Wide-Area Distributed Services", *Proceedings of the First IEEE/ACM International Symposium on Cluster Computing and the Grid(CCGrid 2001)*, Brisbane, Australia, May 2001, pp.92-99.
- [3] H. G. Chen, T. Yu, and K. J. Lin, "QCWS: an implementation of QoS-capable multimedia Web services", *Proceedings of the Fifth International Symposium on Multimedia Software Engineering*, December 10-12, 2003, pp.38-45.
- [4] R. J. AI-Ali, O. F. Rana, D. W. Walker, S. Jha, and S. Sohail, "G-QoSM: Grid Service Discovery Using QoS properties", *Journal of Computing and Informatics*, 2003.
- [5] H. Jin, H. Chen, J. Chen, P. Kuang, L. Qi, and D. Zou, "Real-time Strategy and Practice in Service Grid", *Proceedings of The 28th Annual International Computer Software & Applications Conference (CompSac'04)*, September, 2004, Hong Kong, pp.161-166.