Resource Optimization and Security in Distributed Computing

With the number of e-Business applications dramatically increasing, service level agreements (SLA) will play an important part in distributed service computing. An SLA is a combination of several quality of service (QoS) metrics, such as security, performance, and availability, agreed between a customer and a service provider. Due to the complexity of these metrics, most existing research typically addresses only one of these QoS metrics. In the case of the response time as a performance metric, the average time to process and complete a job is typically used in the literature. However, this may not be of real interest to a customer. A statistically bounded metric, that is, a percentile response time, is more realistic than the average response time. Moreover, in enterprise service computing, customer requests are typically distinguished by different request characteristics and service requirements.

This dissertation includes a study of trustworthiness, percentile response time, service availability, and authentication among service stations or sites that may be owned by different service providers. The first part of this dissertation contains an analysis of percentile response time, which is one of the most important SLA metrics. Effective and accurate numerical solutions for the calculation of the percentile response time in single-class and multi-class queueing networks are obtained. Then, the numerical solution is incorporated in a resource allocation problem. Specifically, we present an approach for the resource optimization that minimizes the total cost of computer resources required while preserving a given percentile of the response time.

In the second part of this dissertation, we extend the approach to consider trustworthiness, service availability, and the percentile of response time in Web services. We clearly define these QoS metrics and provide their quantitative analysis. Then, we take into account these QoS metrics in a trust-based resource allocation problem in which a set of computer resources is used by a service provider to host a typical Web services application for single-class customer services and multiple-class customer services respectively. We formulate the trust-based resource allocation problem as an optimization problem under SLA constraints in which we calculate the number of servers in each service site that minimize a cost function that reflects operational costs for single-class customer services and multiple-class customer services respectively. We solve this problem using an efficient numerical procedure. Experimental results show the applicability of the procedure and validate its accuracy.

Finally, in the third part of this dissertation we first provide a systematic way to analyze several well-known authentication protocols. We give in-depth analysis of these protocols. Then, we propose new authentication protocols, and present a thorough performance evaluation of the authentication protocols compared to the others. Our analysis shows that the proposed protocols perform better than these well-known ones.