Self-organization in Grid Computing

Weifeng Tao

Department of system&information engineering

University of Tsukuba

What is grid?

The term “the Grid” arose in the mid1990s in the supercomputing community with the goal of making underutilized computing resources easily available for complex computations across geographically distributed sites. The *Grid problem* is defined as flexible, secure, coordinated resource sharing, among dynamic collections of sharing computing power, storage capacity, instruments, data, and applications. They involve many thousands or millions of distributed-computing devices, linked together in local or regional networks, which in turn are interconnected on a global level.

The P2P and grid

For another network trends as we can see are the internet computing and peer-to-peer computing, the introduction of the expanded addressing schemes, such as IPv6, will lead to an explosion of the number and types of devices integrated into the internet, and the increased functionality of wireless links will lead to integration of mobile nodes, increasing the spatial-temporal complexity of the network dynamics. Moreover, the persistent demand for free communication drives will continue to push for ubiquitous and universal automatic access, based more and more on peer-to-peer (P2P) interactions, which avoid dependence on centralized master nodes or servers, balanced with considerations of ease of use, security, and reliability. Those technologies are more general (“beyond client-server”) sharing modalities and computational structures which have much in common with Grid technologies. As these applications become more sophisticated and the need for interoperability becomes clearer we will see a strong convergence of interests between peer-to-peer, Internet, and Grid computing. It means that although Grid concepts were first developed in the scientific research fields. It will also become important in personal、commercial and enterprise distributed computing applications over the Internet. Just as the World Wide Web began as a technology for scientific collaboration and was adopted for e-business, we expect a similar trajectory for Grid technologies.

Why using self-organization for grid?

The future grid system can be thought of as a "complex system" comprised of many interacting elements connected by various network media whose behavior is governed by control programs implementing network protocols. One of the most important features of complex system is self-organization. The characteristics of self-organizing systems can be concludes as followings:

1. Global order from local interactions: it means that the interactions of the elements on a local level are those responsible for the organization.
2. Distributed control: the control of the organized system is also distributed among the elements.
3. The distributed control of the system contribute to it robustness and resilience.
4. Non-linearity and feedback: small changes can have large effects through positive and conversely large changes can have small affect through negative feedback.

From those points of views self-organization is ideal for achieving efficient use of the infrastructure under dynamic conditions such as the grid system, it will result in simplified network management, efficient resource discovery and collection, robust and flexibility.

How can we do

Ian Foster and Carl Kesselman had presented in his article “The Anatomy of the Grid” that the grid architecture is comprised of five protocol layers (Figure 1). We will explore each layer about how the self-organization technology can be used in grid system.

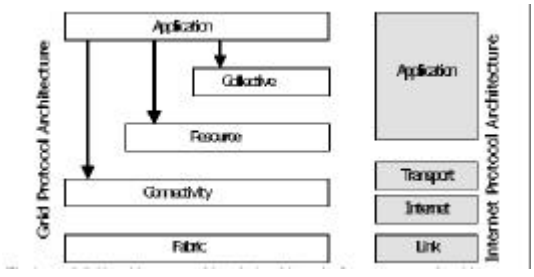

Figure 1

First we survey from base part—physical and connect layer. We concentrate on two main fields routing and security in network layer. Current routing systems are not adequate to tackle the increasing complexity of the internet. Centralized algorithms have scalability problems; static algorithms have trouble keeping up-to-date with network changes; and other distributed and dynamic algorithms have oscillations and stability. Now scientists and engineers had developed the new routing system based on self-organizing algorithm swarm intelligence and ants to provide scalability, fault tolerance, adaptation, speed, modularity, autonomy and parallelism on increasingly diverse and heterogeneous communication networks.

For security problems as we known it uses the spread-spectrum or WEP protocols in physical layer and use symmetric key and encryption technologies in data link layer. But now we can get some idea which presented as AODV protocol to improve the security of network layer with the self-organization routing. In this design, it takes a self-organizing approach by exploiting fully localized network without assuming any prior trust or secret association between network nodes. each node has a token in order to participate in the network operations, and its local neighbors collaboratively monitor it to detect any misbehavior in routing or packet forwarding services. Upon expiration of the token, each node renews its token via its multiple neighbors. The period of the validity of a node's token is dependent on how long it has stayed and behaved well in the network. A well-behaving node accumulates its credit and renews its token less and less frequently as time evolves. Also Hubaux and others proposed the self-organized public-key infrastructure for the ad hoc networks, which was similar to the PGP. In this public-key infrastructure, the certificate of each node is issue by other nodes, and the certificate chain is used to verify a certificate.

Then we exploit the resource and collective layers. The current efforts of the Grid community are focusing on local cluster resource management. Resource management mechanisms are very well suited for single Grid clusters incorporating optimized scheduling algorithms and load sharing mechanisms, but they are not work well between multiple Grids. The next stage is the inter-working evolution between the different communities of Grids and underlying mechanisms that can provide hard quality of service. A self-organized approach to resource management will provide the infrastructure for a reliable and consistent Grid. The three main components are resource discovery, scheduling, and dynamic resource information. The self-organized function of a resource management system consists of self-configuring, self-optimizing and self-healing properties. Self-configuring arises from being able to discover resources, and schedule jobs and work load autonomously. In order to provide high throughput, the system should be self-optimizing. This can include allocation of workloads to specific resources based on policies that ensure hard quality of service guarantees and preferential attachment to resources. In the event that a resource is no longer available, the system should self-heal. In such a scenario alternative resources that are capable of completing the jobs can be used, the system will adapt and reschedule accordingly. For example some research group is investigating the use of small world techniques for resource management.

There also have some research works been done in the grid information services which use the CAN-based DHT-system into an indexing infrastructure, the advantages of such an approach are the self-organizing characteristics, fault-tolerance and scalability inherent to P2P system.

At last for application layer, there are many application use the self-organization technology. Those issues have beyond the scope of this article. People who are interest in this field can read some other books and articles.

Work of the future

As more and more works focuses on the study of self-organization system, we will know more about the science of self-organization, and we will applied it more effectively and accurately. Beside that the suitable architecture is the key to apply self-organization science in the grid system.