ALTER: First Step towards Dependable Grids

Xuanhua Shi, Hai Jin, Weizhong Qiang
Cluster and Grid Computing Lab,
Huazhong University of Science and Technology,
Wuhan, 430074, China
{xhshi, hjin, wzqiang}@hust.edu.cn

ABSTRACT
This paper presents ALTER, an adaptive failure detection service, which incorporates the technique of unreliable failure detection service and the idea of R-GMA. ALTER is organized in a hierarchical structure and it can be adaptive to the system conditions and user requirements with changing the system parameters and system organizations. With mathematical evaluation, ALTER shows good scalability and flexibility, which is suitable for grid environments.

Categories and Subject Descriptors
D.4.7 [Operating Systems]: Organization and Design - Distributed systems

General Terms
Algorithms, Performance, Design, Experimentation

Keywords
dependable grid, adaptive failure detection, scalability

1. INTRODUCTION
Due to the diverse failures and error conditions in grid environments, developing, deploying, and executing large-scale applications is a challenge, dependability is a key factor for grid computing. Failure detector is well known as a fundamental building block for dependable systems. Providing a generic scalable and high-efficiency failure detection service as the first step towards dependable Grids is of primary importance in Grid computing.

In this paper we present ALTER, an adaptive failure detection service with good scalability and flexibility, and this implementation addresses the unique requirements for failure detection for Grids, This implementation is a blend of unreliable failure detectors [2] for distributed systems and R-GMA (Relational Grid Monitoring Architecture) [4]. ALTER can tune system performance with different QoS requirements for applications and different requirements for different system components. Furthermore, ALTER can change its topology according to grid changes, such as some resources are added in, some key components crash or even some components of failure detectors crash.

2. IMPLEMENTATION OF ALTER

We now present implementation of ALTER, which includes its architecture and the adaptive protocols

2.1 System Architecture
ALTER is organized in a hierarchical structure, as shown in Fig. 1. The system architecture composes two levels: local groups and global groups, and in local group there is a unique group leader. Failure detectors in local groups monitor the objects in the local space, and the monitored objects in one local space may be in one LAN or cross LANs, but the network condition in one space should be good. Failure detectors in global groups monitor the global objects by means of monitoring the detectors in local groups. Thus, in this architecture there are two different types of failure detectors: local failure detector and group leader. The monitored objects send "I'm alive" messages to local failure detectors periodically, while the messages that a group leader send is a list containing the monitored objects and their status, and the group leaders share failure detection messages with epidemic method.

For management simplicity, there is an index service in the system implementation, and the index service works as a directory registry of the global group failure detectors and the local group failure detectors. Besides, the index service provides some decision-making capability for the organization of the group leaders. There are three components or roles in ALTER system: consumers, producers, and an index service. A consumer wants to detect some objects in grid, first queries the index service and gets the location of the producer, the failure detector is a producer which gives the status of the monitored objects.

2.2 Adaptive Protocols
Grid fault detection service addresses two things, one is how to satisfy the QoS between two processes, and the other is how to satisfy the Grid dynamic nature. The QoS between the monitored process and the detector can be donated in a tuple \((T_{UD}, T_{MR}, T_{UM})\), where \(T_{UD}\) is an upper bound of the
Then $n = \lceil 2 \log S_g N_{total} \rceil$; the number of groups needed, $S_g$ is the size of the local groups.

$$\text{if } \forall i,j, T_{ij} \leq \Phi[i,j] \in LAN \rightarrow \text{good} , \text{/network evaluation then MergGrp[i,j]; /smallest annexed to second smallest} \}$$

### 3. ARITHMETIC ANALYSIS

This section we present an arithmetic method to analyse the performance of ALTER.

**Definition 1** Completeness (adopted from [1][2]):
\[ \exists \gamma : \forall t, \Phi \in \text{correct}(t), \forall q \in \text{crashed}, q \notin \text{correct}(t) \]

**Evaluation:** As depicted in Algorithm 1, $ET_{n+1}$ is a constant,

so $\Delta_1 : ET_{n+1} \leq \Delta \Rightarrow \exists \Delta_2 : \tau_{n+1} \leq \Delta_2$

$$= \Delta ; \Delta \text{timeout} = \tau_{n+1} - \text{Total} \leq \Delta$$

so if the $p$ process does not get any message from $q$ after $\Delta$
time, the $q$ will be suspected by process $p$.

**Definition 2** Accuracy (adopted from [1][2]):
\[ \exists \gamma : \forall t_n, t_p, q \in \text{correct}(t), q \notin \text{correct}(t) \]

**Evaluation:** From Algorithm 1, we can conclude that:

$$\exists \Delta_1 : \tau_q < ET_{n+1} \leq \tau_p + \Delta_1$$

$t_q$ is the time when $q$ sends message $n_{a+1}$.

Thus it’s easy to conclude that:

$$\exists \Delta_2 : \forall t_q, \tau_q < \Delta_2$$

$$\Rightarrow \exists t_q : \forall t_q \in \text{correct}(t), q \notin \text{correct}(t) \}

**Definition 3** Scalability:
\[ \exists \gamma : \forall NT, \tau, \text{log} \eta \leq \tau \text{correct}, N_M \leq \gamma \text{log} \eta \text{NT} \text{Total} \]

**Evaluation:** obviously

$$\exists \Delta_1 : \text{monitor} (t), q \in \text{local}(t), \tau_q < \Delta_1$$

$$\Rightarrow \exists \gamma : (t \in \text{local}(t), j \in \text{groupleader}(t), T_2 \leq \tau_q \text{log} N_y$$

$

N_g$ is the number groups. The total message can be computed
as $N_M = N_{Total} + N_g (N_g - 1)$.

With algorithm 2, we get:

$$\exists \gamma : (t \in \text{local}(t), j \in \text{groupleader}(t), T_2 \leq \tau_q \text{log} N_y$$

**Definition 4** Flexibility:
\[ \exists \gamma : \forall \text{app}, \forall t_q \geq t_p, \forall q \in \text{correct}(q), \forall q \in \text{crashed}, q \notin \text{correct}(t) \]

**Evaluation:** The system is adaptive to the users’ requirements and system conditions, without any knowledge about the applications, so our method is independent of applications, and this model is flexible.

### 4. CONCLUSION AND FUTURE WORK

In this paper, we focus on the dependability of the grid. To increase the dependability, an adaptive failure detection service, ALTER is presented. ALTER is organized in a hierarchical structure, which incorporates the technique of unreliable failure detection service and the idea of R-GMA. ALTER is adaptive to the system conditions and the users’ requirements by changing the system parameters and the system topology. Currently, ALTER is deployed on the campus grid of Huazhong University of Science and Technology. Failure avoidance and failure recovery in grid is the future work based on ALTER.

### 5. REFERENCES


