IEEE International Conference on
2003 CLUSTER COMPUTING
H O N G K O N G

Cluster Architecture with Lightweighted Redundant TCP Stacks

Hai Jin and Zhiyuan Shao

Availability of the services provided by the cluster system is greatly emphasized in today's system domain. We propose a new technique, named Redundant TCP Stacks (RTS), to enhance the connection-level reliability and availability of the services provided by cluster system.

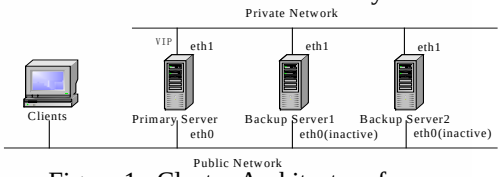


Figure 1. Cluster Architecture for Connection-level Availability

Communication Synchronization

Algorithm 1: *Send When the Minimum Updated* (SWMU)

When and only when the Primary Server gathers all the responses from the server nodes, the response is eventually sent to the client.

Pros. works very safe and stable.

Cons. harms the performance of the communication by increase the *Round Trip Time* (RTT) between the client and the servers.

Algorithm 2: *Send When the Fastest Response* (SWFR)

When the response of the fastest server in the server farm reaches the primary server, the response is sent to the client. The only prerequisite is that before sending the fastest response, all the communications of the server nodes should be synchronized.

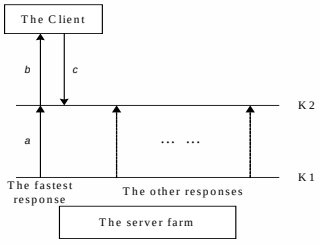


Figure 2. An Example of SWFR Algorithm

When "*a*" arrives at the primary server, the primary server checks the current status of the communications of the server nodes. If synchronized, response "*b*" is sent out. (In Fig.2 K1 and K2 represents the TCP sequence numbers) After processing by the client, packet *c* is sent back to the server farm. At that time, the TCP stacks of the server nodes are very likely to have another synchronized state again.

Related Technique:Backup Responses Compressing (BRC)

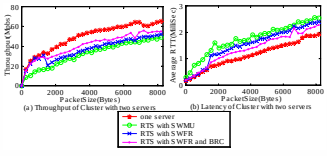
The responses of the backup servers are greatly compressed to alleviate the ACK implosion problem of the primary server.

Pros. enhanced network performance

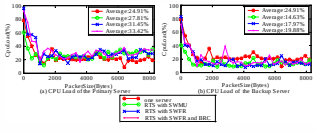
Cons. only tolerate silent fault

Performance Evaluation

I Performance of Two Servers

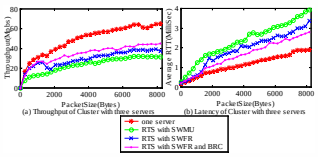


Performance of Communication (Benchmark: Netpipe 2.4)

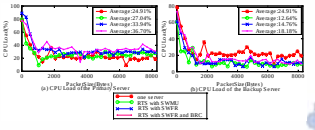


CPU Load of the Primary and Backup Servers

II Performance of Three Servers



Performance of Communication



CPU Load of the Primary and Backup Servers