Load Sharing

Reference: Pradeep K Sinha,

"Distributed Operating Systems: Concepts and Design", Prentice Hall of India, 2007

Load sharing approach

- For the proper utilization of the resources of a distributed system it is not required to balance the load on all the nodes.

- It is necessary and sufficient to prevent the nodes from being idle while some other node have more than two processes.
- This rectification is often called dynamic load sharing instead of dynamic load balancing.

Issues in designing load-sharing algorithms

- Load sharing algorithms do not attempt to balance the average workload on all the nodes of the system, rather they only attempt to ensure that no node is idle when a node is heavily loaded.

- The priority assignment policies and migration limiting policies are same as that for the load-balancing algorithms. Other policies are described here.

Load Estimation Policy

- It is sufficient to know whether a node is busy or idle.

- Methods for estimating load:
  - Count the total number of processes on a node.
  - Measure CPU utilization.

Process transfer policies

- All-or-nothing strategy:
  - Uses the single threshold policy with the threshold value of all the node fixed at 1 and some uses 2.

  - Drawback : Loss of available processing power in the system.
  - Solution : use a threshold value of 2 instead of 1.

Location policies

1. Sender-initiated policy
2. Receiver-initiated policy

Sender initiated location policy

- Heavily loaded nodes search for lightly loaded node to which work may be transferred.

- When load becomes more than the threshold value, it either broadcasts a message or randomly probes the other nodes one by one to find a lightly loaded node.

Sender initiated location policy

- In the broadcasting method, the presence or absence of a suitable receiver node is known as soon as the sender node receives reply messages from the other nodes.

Sender initiated location policy

- In the random probing method, the probing continues until either a suitable node is found or the no. of probes reaches a static probe limit, Lp.

- Fixed limit has better scalability than broadcast method.

Receiver-initiated location policy

- Lightly loaded node search for heavily loaded nodes from which work may be transferred.

- When a node's load falls below the threshold value either it broadcasts a message indicating its willingness to receive processes or randomly probes the other nodes one by one to find a heavily loaded node.

Receiver-initiated location policy

- In the broadcast method, a suitable node is found as soon as the receiver node receives reply messages from the other nodes.

- In the random probing method, the probing continues until either a suitable node is found or the no. of probes reaches a static probe limit, Lp.

Location Policies

- Both sender-initiated and receiver-initiated policies offer substantial performance advantages over the situation in which no load sharing is attempted.

- Sender-initiated policies are preferable at light to moderate system loads, while receiver-initiated policies are preferable at high system loads.

Location Policies

- If the cost of process transfer under receiver-initiated policies is significantly greater than under the sender-initiated policies due to the preemptive transfer of processes

- Sender-initiated policies provide uniformly better performance.

State information Exchange Policy

1. Broadcast when state changes.
2. Poll when state changes.

Thank You