Resource Discovery Techniques in Distributed Desktop Grid Environments

Presented by:
Arun Balasubramanian

Approach

- Centralized Systems – Condor
- Distributed Systems – BOINC, Seti@home
- Approach used in Paper
  - A combination of both centralized and distributed.

How can this be Accomplished?

- Match–making strategy
  - Match jobs to processing nodes that satisfies resource constraints.
- 3 types of Strategies are discussed.
  - Rendezvous Node Tree
  - CAN (Content Addressable Network)
  - Centralized matchmaker

Goals of MatchMaking Algorithms

- Low overhead
  - Routing cost should be much less compared to job execution cost.
- Completeness
  - A valid assignment should be found.
- Precision
  - A node shouldn’t be over–provisioned to a job.
- Load Balance
  - Jobs should be distributed across the nodes.
Basic Terminologies

- There’s an underlying DHT Infrastructure.
- Jobs have an associated profile.
  - CPU speed, memory requirement, disk space.
- Clients submit jobs to any system node.
- Nodes receiving jobs from client assign GUIDs and forward the jobs to owners.
- Heart beats exchanged between compute nodes and owners.

Rendezvous Node Tree

- Implemented on “Chord”.
- Once the RN–Tree is constructed, the parent–child relationship is determined.
- Each node periodically sends local subtree resource information to its parents.
- Owner initiates search for the node to run the job.
- Search proceeds through sub-tree rooted at owner. Then, ancestors.
- K– best candidates are found.

CAN
(Content Addressable Network)

- Each resource type is a distinct dimension.
- Volume divided into zones managed by individual nodes.
- Job is inserted using its resource constraints
- Owner of the job’s zone becomes owner of the job.
- Candidate nodes are drawn from the owners of the neighboring zones. (each candidate should more capable in at least one constraint than owner)

Virtual dimension

- Nodes may have similar resource capacities. Zone division becomes a challenge.
- Jobs may have similar resource constraints.
- A randomly assigned value as a virtual dimension is introduced.
Centralized Match maker

- Maintains global information about current capabilities and load information for all nodes in the system.

- Extreme form of both RN_tree and CAN.

- Finds all candidate nodes that meets job constraints and picks the one with shortest job queue.

Performance Results

- Clustered Workloads
  - RN Tree has lower match-making costs.
  - CAN has lower wait times and queue lengths.

- Mixed Workloads
  - Match-making costs favor CAN for heavy workloads.
  - But CAN performs poorly for light workloads.