Using Checkpointing to Recover from Poor Multi-site Parallel Job Scheduling Decisions

William M. Jones
Electrical Engineering Department
United States Naval Academy
Annapolis, Maryland, USA

http://www.parl.clemson.edu/beosim
Multi-site Co-allocation
Scheduling w/o co-allocation
Scheduling w/ co-allocation

makespan without co-allocation

makespan with co-allocation

improvement

Cluster 1

Cluster 2

running job

waiting job 1

running job

waiting job 2

waiting job 1

waiting job 2

Time
Co-allocation w/ slowdown

- Makespan without co-allocation
- Makespan with co-allocation and some NW congestion
- Makespan with co-allocation and no NW congestion

Cluster 1
- Running job
- Waiting job 1
- Waiting job 2

Cluster 2
- Running job

Time

Improvement

Execution slowdown

Waiting job 1

Waiting job 2
Interprocess Communication

All-to-all personalized

Cluster 1
1
2
3

Cluster 2
4
5
6

Cluster 3
7
8
9

2D nearest neighbor

Cluster 1
1
2
3

Cluster 2
4
5
6

Cluster 3
7
8
9
Previous Results

Improvement Over Migration Only
w/ No Estimate Inaccuracy

% improvement

Max (0% Error)

150 Mbps
300 Mbps
400 Mbps
What happens if there is significant inaccuracy in the user-provided bandwidth estimates?
Turnaround Time vs Info w/ PPBW = 400 Mbps

Job turnaround time

Information Availability (% of jobs)

@ +/- 0% Error
@ +/- 50% Error
@ +/- 80% Error
@ +/- 100% Error
Migration Only
Impact of +/- 100% Estimate Inaccuracy

% Improvement Over Migration Only

Max (0% Error)  
| w/o CPing |

<table>
<thead>
<tr>
<th>150 Mbps</th>
<th>300 Mbps</th>
<th>400 Mbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.5</td>
<td>15.5</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

130 Mbps

-2.5

-5
What potential gain might there be to employ checkpointing and run-time job migration to mitigate network over-subscription?
Turnaround Time vs CP Overhead vs Error w/ PPBW = 150 Mbps

Job Turnaround Time (mins)
Turnaround Time vs CP Overhead vs Error w/ PPBW = 300 Mbps

Job Turnaround Time (mins)

CP overhead (mins)

+/- % Error in User Estimate
Turnaround Time vs Error w/ PPBW = 300 Mbps

+/-% Error in User Estimate

Job Turnaround Time

Equivalent RMS % Error in User Estimate

w/o CPing
CPing w/ OH=00 min
CPing w/ OH=05 min
CPing w/ OH=10 min
CPing w/ OH=20 min
CPing w/ OH=30 min
Migration Only
Turnaround Time vs CP Overhead vs Error w/ PPBW = 400 Mbps

Job Turnaround Time
(mins)

CP overhead (mins)

+/- % Error in User Estimate

1350 1300 1250 1200 1150 1100 1050 1000 950

90 80 70 60 50 40 30 20 10 0

1250 1200 1150 1100 1050 100

80 60 40 20 0

90 80 70 60 50 40 30 20 10 0
Results Summary at 5 and 10 Minute CP Overhead at 100% Estimate Error

- **Max (0% Error)**
- **CPing (5 min)**
- **CPing (10 min)**
- **w/o CPing**

% Improvement Over Migration Only

- **150 Mbps**
- **300 Mbps**
- **400 Mbps**
Turnaround Time vs CP Overhead w/ PPBW=150 Mbps

- @ 80% Error w/o CPing
- @ 80% Error w/ CPing
- @ 100% Error w/o CPing
- @ 100% Error w/ CPing
- Migration Only

Job Turnaround Time (mins)

Checkpoint Overhead (mins)
Conclusions

- Bandwidth-aware Co-allocation
  - Improvement over migration-only
  - Estimate inaccuracies
    - Negligible --> severe

- Checkpoint/Migrate/Restart
  - Recovery from network contention
  - Cost
    - Effective even when CP overhead is high
Thank you!

Questions?

William M. Jones
http://www.parl.clemson.edu/beosim