Evaluating Workload and Machine Heterogeneity in Distributed Computing Systems

A. M. Al-Qawasmeh, A. A. Maciejewski, H. J. Siegel, J. Smith, and J. Potter
Colorado State University
Fort Collins, Colorado, USA
Abdulla.Al-Qawasmeh@ColoState.edu

Outline

● Distributed Heterogeneous Computing Systems
● Statistical Measures for Evaluating Heterogeneity
● Simulation Results
● Conclusions
Distributed Heterogeneous Computing System

- A cluster of different types of machines
  - Varied computational capabilities among the system machines

- Workload
  - Tasks with different computational requirements

- Each task may perform differently on each machine
  - Furthermore: machine A can be better than machine B for task 1 but not for task 2

- Resource allocation: assign tasks to machines to optimize some performance measure
  - NP-complete (cannot find optimal in reasonable time)
  - Use heuristics to find near-optimal allocation
Estimated Time to Compute (ETC) Matrix

- ETC obtained from real world systems and benchmark data

<table>
<thead>
<tr>
<th></th>
<th>m1</th>
<th>m2</th>
</tr>
</thead>
<tbody>
<tr>
<td>tasks</td>
<td>t1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>t2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>t3</td>
<td>20</td>
</tr>
</tbody>
</table>

*estimated time to compute task 1 on machine 1*

\[ETC(1, 1) = 3\]
Machine Heterogeneity vs. Task Heterogeneity

- task vs. machine heterogeneity
  - task heterogeneity: variation along a column

<table>
<thead>
<tr>
<th></th>
<th>m1</th>
<th>m2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>t2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>t3</td>
<td>20</td>
<td>18</td>
</tr>
</tbody>
</table>
Machine Heterogeneity vs. Task Heterogeneity

- task vs. machine heterogeneity
  - task heterogeneity: variation along a column
  - machine heterogeneity: variation along a row

<table>
<thead>
<tr>
<th></th>
<th>m1</th>
<th>m2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>t2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>t3</td>
<td>20</td>
<td>18</td>
</tr>
</tbody>
</table>
Outline

- Heterogeneous Computing Systems
- Statistical Measures for Quantifying Heterogeneity
- Simulation Results
- Conclusions
Measures of Heterogeneity

- goal: select heuristic that is best for an environment with a given heterogeneity
  - existing heterogeneity measures not enough
  - more measures correlated with heuristic performance

- studied measures of task and machine heterogeneities
  1. coefficient of variation
  2. skewness (third central moment)
  3. kurtosis (fourth central moment)
Statistical Measures - COV

- coefficient of variation (COV):
  - let $\mu$ be the mean of a set of values
  - and $\sigma$ the standard deviation of a set of values
  - the COV, $V$, is given by

$$V = \frac{\sigma}{\mu}$$
Statistical Measures - Average Task COV

\[ V_j^{(m)} \] is the tasks COV for machine \( j \)

\[ \dot{o}_j^{(m)} \] task standard deviation for machine \( j \)

\[ \dot{i}_j^{(m)} \] task mean for machine \( j \)

\[ V_j^{(m)} = \frac{\dot{o}_j^{(m)}}{\dot{i}_j^{(m)}} \]
Statistical Measures – Average Task COV

- \( V_j^{(m)} \) is the tasks COV for machine \( j \)
- \( \bar{\dot{\mu}}_j^{(m)} \) task standard deviation for machine \( j \)
- \( \bar{\dot{\mu}}_j^{(m)} \) task mean for machine \( j \)

\[
V_j^{(m)} = \frac{\dot{\mu}_j^{(m)}}{\bar{\dot{\mu}}_j^{(m)}}
\]

- Task heterogeneity as measured by the Average Task COV (ATC)
  - \( M \): number of machines

\[
ATC = \frac{\sum_{j=1}^{M} V_j^{(m)}}{M}
\]
Statistical Measures – Average Machine COV

\( V_i^{(t)} \) is the machine COV for task \( i \)

\( \dot{V}_i^{(t)} \) is the machine mean for task \( i \)

\( \ddot{V}_i^{(t)} \) is the machine standard deviation for task \( i \)

\[ V_i^{(t)} = \frac{\dot{V}_i^{(t)}}{\ddot{V}_i^{(t)}} \]
**Statistical Measures – Average Machine COV**

- $V_{i}^{(t)}$ is the machine COV for task $i$
- $\dot{\omega}_{i}^{(t)}$ machine standard deviation for task $i$
- $\dot{i}_{i}^{(t)}$ machine mean for task $i$

$$V_{i}^{(t)} = \frac{\dot{\omega}_{i}^{(t)}}{\dot{i}_{i}^{(t)}}$$

- task heterogeneity as measured by the Average Machine COV (AMC)

$$AMC = \frac{1}{T} \sum_{t=1}^{T} V_{i}^{(t)}$$
Statistical Measures – Skewness (Third Central Moment)

- **positive skewness**
  - 5 more values less than mean
- **negative skewness**
  - 5 more values greater than mean
Statistical Measures – Average Task Skewness

- $S_{j}^{(m)}$ tasks skewness for machine $j$

\[
S_{j}^{(m)} = \left[ \frac{1}{T} \sum_{i=1}^{T} (ETC(i,j) - i^{(m)}_j)^3 \right] / (\delta^{(m)}_j)^3
\]
Statistical Measures – Average Task Skewness

- $S_j^{(m)}$ tasks skewness for machine $j$

$$S_j^{(m)} = \frac{1}{T} \sum_{i=1}^{T} \left( \text{ETC}(i, j) - \bar{i}_j^{(m)} \right)^3 / (\bar{o}_j^{(m)})^3$$

- task heterogeneity as measured by the Average Task Skewness (ATS)

$$ATS = \left[ \sum_{j=1}^{M} S_j^{(m)} \right] / M$$
Statistical Measures – Average Machine Skewness

\[ S_{i(t)}^{(t)} \text{ machines skewness for task } i \]

\[ S_{i(t)}^{(t)} = \left[ \frac{1}{M} \sum_{j=1}^{M} (ETC(i, j) - \hat{i}_{i(t)})^3 \right] / (\hat{\sigma}_{i(t)})^3 \]
Statistical Measures – Average Machine Skewness

- $S_i^{(t)}$ machines skewness for task $i$

$$S_i^{(t)} = \frac{1}{M} \sum_{j=1}^{M} (ETC(i, j) - \bar{i}_i^{(t)})^3 / (\bar{o}_i^{(t)})^3$$

- machine heterogeneity as measured by the Average Machine Skewness (AMS)

$$AMS = \frac{\sum_{i=1}^{T} S_i^{(t)}}{T}$$
Statistical Measures – Kurtosis (Fourth Central Moment)

- **high kurtosis**
  - 5 small number of values having extreme deviations
- **low kurtosis**
  - 5 large number of values having modestly-sized deviations

Source: http://en.wikipedia.org/wiki/Kurtosis
Statistical Measures – Average Task Kurtosis

\[ K_j^{(m)} = \left( \frac{1}{T} \sum_{i=1}^{T} (ETC(i, j) - \bar{i}_j^{(m)})^4 \right) \left( \bar{\sigma}_j^{(m)} \right)^4 - 3 \]
Statistical Measures – Average Task Kurtosis

- $K^{(m)}_j$ tasks kurtosis for machine $j$

\[
K^{(m)}_j = \left[ \frac{1}{T} \sum_{i=1}^{T} \left( \text{ETC}(i, j) - i^{(m)}_j \right)^4 \right] / \left( \sigma^{(m)}_j \right)^4 - 3
\]

- Task heterogeneity as measured by the Average Task Kurtosis ($ATK$)

\[
ATK = \left[ \frac{1}{M} \sum_{j=1}^{M} K^{(m)}_j \right]
\]
Statistical Measures – Average Machine Kurtosis

$K_i^{(t)}$  machines skewness for task $i$

$$K_i^{(t)} = \left[ \frac{1}{M} \sum_{j=1}^{M} (\text{ETC}(i, j) - i_i^{(t)})^4 / (\dot{i}_i^{(t)})^4 \right] - 3$$
Statistical Measures – Average Machine Kurtosis

- $K_i^{(t)}$ machines skewness for task $i$

\[
K_i^{(t)} = \left[ \left( \frac{1}{M} \sum_{j=1}^{M} (ETC(i, j) - i_{i}^{(t)})^4 \right)/\left( \hat{o}_{i}^{(t)} \right)^4 \right] - 3
\]

- machine heterogeneity as measured by the Average Machine Kurtosis (AMK)

\[
AMK = \left[ \frac{1}{T} \sum_{i=1}^{T} K_i^{(t)} \right]
\]
Simulation Results

- **ETC size**
  - 8 machines
  - 128 tasks

- **Static mapping**
  - All tasks available for execution
  - Mapping done offline

- **Studied heuristics**
  - Max-Min and Min-Min

- **Makespan**
  - Time when all tasks finish
  - Lower is better

- **Normalized makespan**
  - Makespan of Min-Min divided by Max-Min
Simulation Results – Effect of ATS

Max-Min outperforms Min-Min for ATS of 1.4 and higher
Simulation Results – Effect of AMC

Min-Min outperforms Max-Min for AMC of 0.5 and higher
Conclusions

- currently used statistical measures (COV) not sufficient
- need to use additional measures
  - skewness (third central moment)
  - kurtosis (fourth central moment)
- skewness and kurtosis impact the performance of heuristics
- ignoring skewness and kurtosis may lead to wrong use of a heuristic