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

- 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

Colorado State University

Estimated Time to Compute (ETC) Matrix

- ETC obtained from real world systems and benchmark data

		machines	
		m1	m2
tasks	t1	3	5
	t2	7	4
	t3	20	18

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

	m1	m2
t1	3	5
t2	7	4
t3	20	18

Machine Heterogeneity vs. Task Heterogeneity

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

	m1	m2
t1	3	5
t2	7	4
t3	20	18

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 μ be the mean of a set of values
  - and σ the standard deviation of a set of values
  - the COV, *V*, is given by

$$V = \sigma / \mu$$

Statistical Measures - Average Task COV

- $V_j^{(m)}$ is the tasks COV for machine *j*
  - $\sigma_j^{(m)}$ task standard deviation for machine *j*
  - $\mu_j^{(m)}$ task mean for machine *j*

$$V_j^{(m)} = \frac{\sigma_j^{(m)}}{\mu_j^{(m)}}$$

Statistical Measures – Average Task COV

- $V_j^{(m)}$ is the tasks COV for machine *j*
  - $\sigma_j^{(m)}$ task standard deviation for machine *j*
  - $\mu_j^{(m)}$ task mean for machine *j*

$$V_j^{(m)} = \frac{\sigma_j^{(m)}}{\mu_j^{(m)}}$$

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

$$ATC = [\sum_{j=1}^{M} V_j^{(m)}]/M$$

Statistical Measures – Average Machine COV

- $V_i^{(t)}$ is the machine COV for task *i*
  - $\sigma_i^{(t)}$ machine standard deviation for task *i*
  - $\mu_i^{(t)}$ machine mean for task *i*

$$V_i^{(t)} = \frac{\sigma_i^{(t)}}{\mu_i^{(t)}}$$

Statistical Measures – Average Machine COV

- $V_i^{(t)}$ is the machine COV for task *i*
  - $\sigma_i^{(t)}$ machine standard deviation for task *i*
  - $\mu_i^{(t)}$ machine mean for task *i*

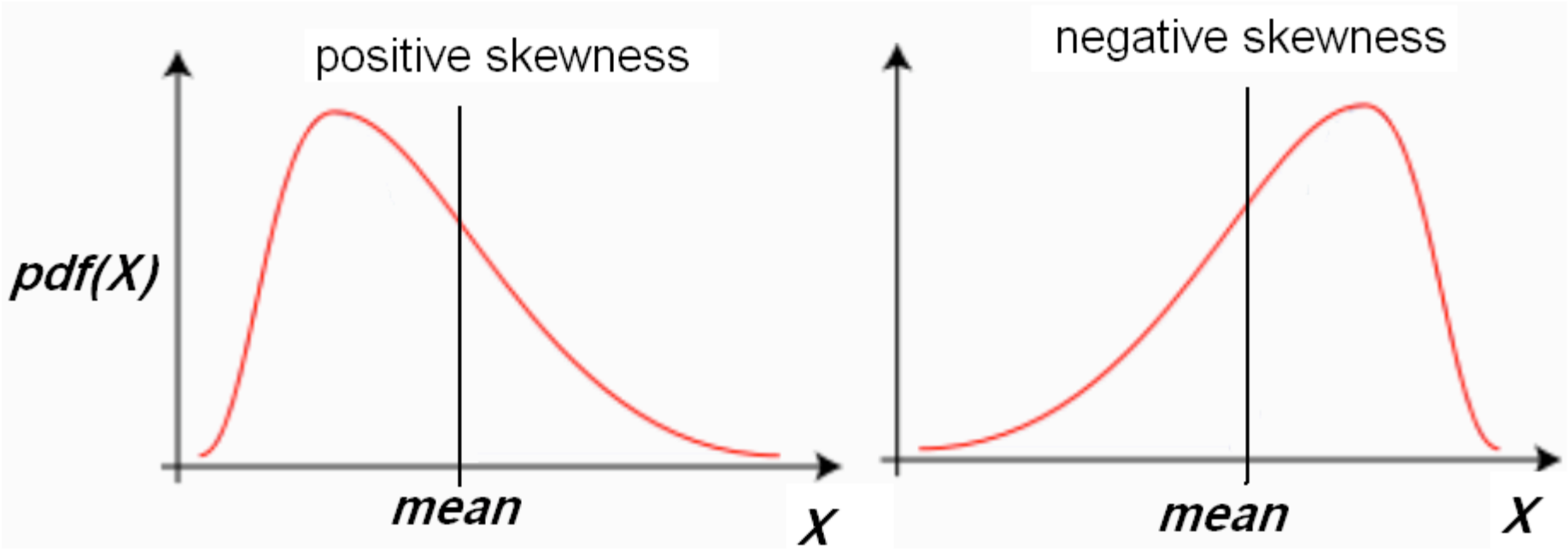
$$V_i^{(t)} = \frac{\sigma_i^{(t)}}{\mu_i^{(t)}}$$

- task heterogeneity as measured by the Average Machine COV (*AMC*)
  - T: number of tasks

$$AMC = [\sum_{i=1}^{T} V_i^{(t)}] / T$$

Statistical Measures – Skewness (Third Central Moment)

- positive skewness
  - more values less than mean
- negative skewness
  - more values greater than mean



Statistical Measures – Average Task Skewness

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

$$S_j^{(m)} = [\frac{1}{T}\sum_{i=1}^{T}(\mathrm{ETC}(i,j) - \mu_j^{(m)})^3]/(\sigma_j^{(m)})^3$$

Statistical Measures – Average Task Skewness

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

$$S_j^{(m)} = [\frac{1}{T}\sum_{i=1}^{T}(\mathrm{ETC}(i,j) - \grave{\imath}_j^{(m)})^3]/(\acute{o}_j^{(m)})^3$$

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

$$ATS = [\sum_{j=1}^{M} S_j^{(m)}]/M$$

Statistical Measures – Average Machine Skewness

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

$$S_i^{(t)} = [\frac{1}{M}\sum_{j=1}^{M}(\mathrm{ETC}(i,j) - \grave{\imath}_i^{(t)})^3]/(\acute{o}_i^{(t)})^3$$

Statistical Measures – Average Machine Skewness

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

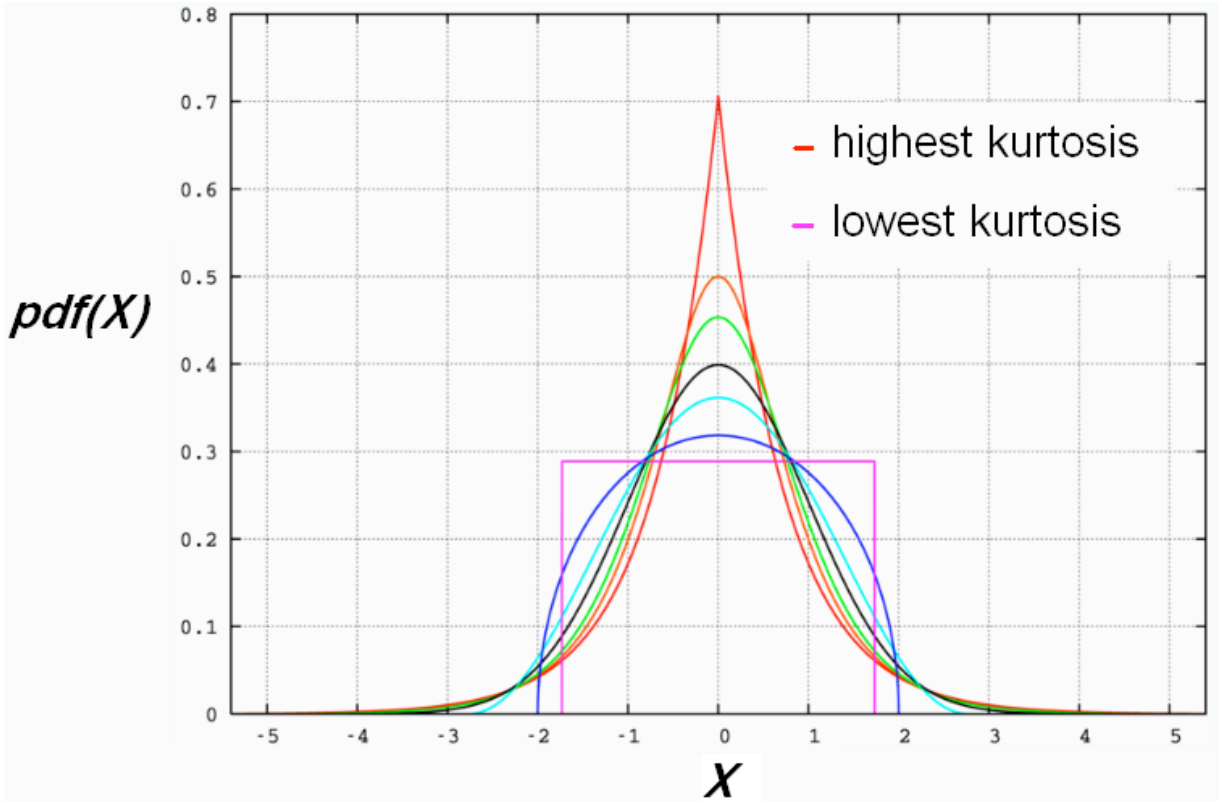
$$S_i^{(t)} = [\frac{1}{M}\sum_{j=1}^{M}(\mathrm{ETC}(i,j) - \grave{\imath}_i^{(t)})^3]/(\acute{o}_i^{(t)})^3$$

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

$$AMS = [\sum_{i=1}^{T} S_i^{(t)}]/T$$

Statistical Measures – Kurtosis (Fourth Central Moment)

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



Source: http://en.wikipedia.org/wiki/Kurtosis

Colorado State University

Statistical Measures – Average Task Kurtosis

- $K_j^{(m)}$ tasks kurtosis for machine *j*

$$K_j^{(m)} = [(\frac{1}{T}\sum_{i=1}^{T}(\mathrm{ETC}(i,j) - \mu_j^{(m)})^4)/(\sigma_j^{(m)})^4] - 3$$

Statistical Measures – Average Task Kurtosis

- $K_j^{(m)}$ tasks kurtosis for machine *j*

$$K_j^{(m)} = [(\frac{1}{T}\sum_{i=1}^{T}(\mathrm{ETC}(i,j) - \grave{\imath}_j^{(m)})^4)/(\acute{o}_j^{(m)})^4] - 3$$

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

$$ATK = [\frac{1}{M}\sum_{j=1}^{M} K_j^{(m)}]$$

Statistical Measures – Average Machine Kurtosis

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

$$K_i^{(t)} = [(\frac{1}{M}\sum_{j=1}^{M}(\mathrm{ETC}(i,j) - \grave{\imath}_i^{(t)})^4)/(\acute{o}_i^{(t)})^4] - 3$$

Colorado State University

Statistical Measures – Average Machine Kurtosis

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

$$K_i^{(t)} = [(\frac{1}{M}\sum_{j=1}^{M}(\mathrm{ETC}(i,j) - \grave{\imath}_i^{(t)})^4)/(\acute{o}_i^{(t)})^4] - 3$$

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

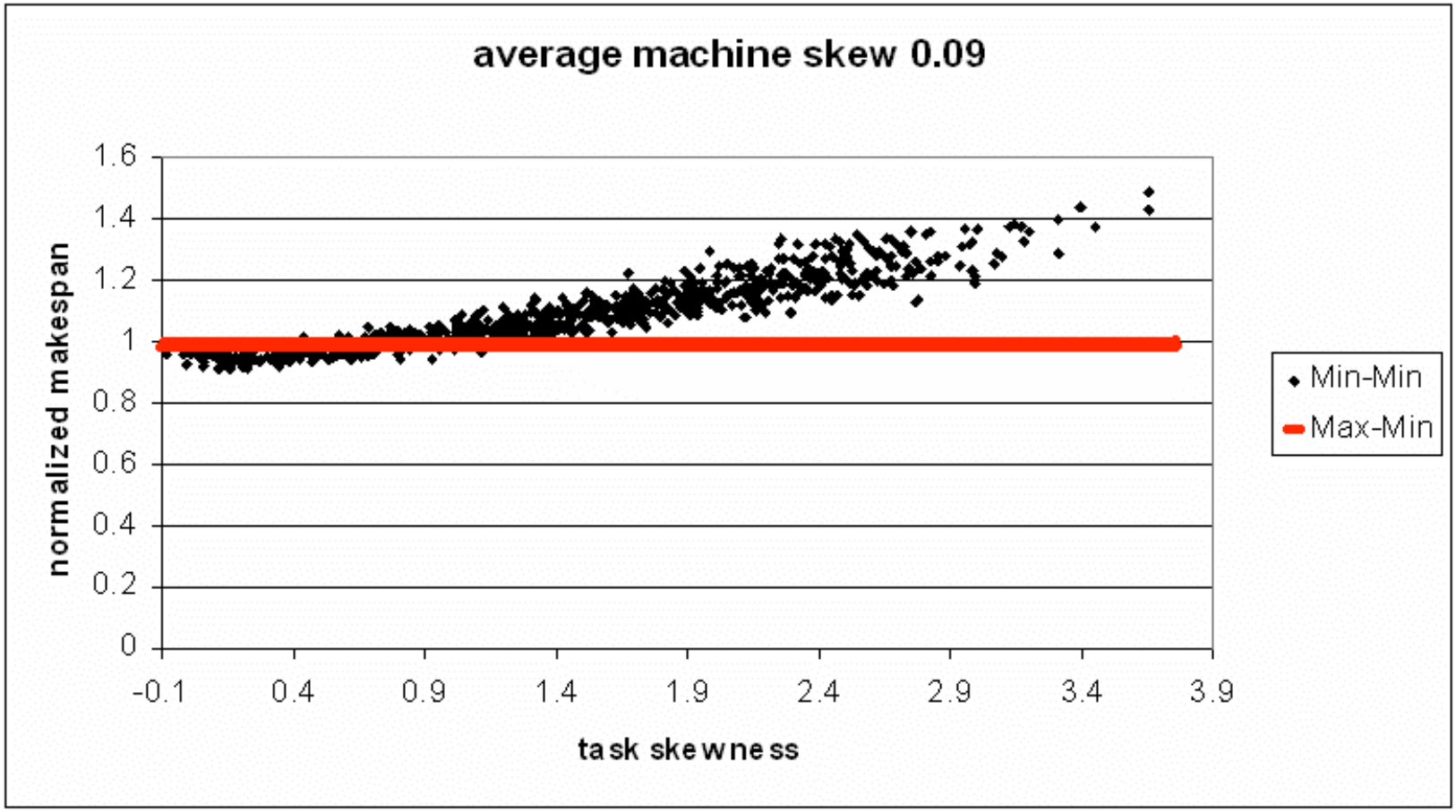
$$AMK = [\frac{1}{\mathrm{T}}\sum_{\mathrm{i}=1}^{\mathrm{T}} K_i^{(t)}]$$

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

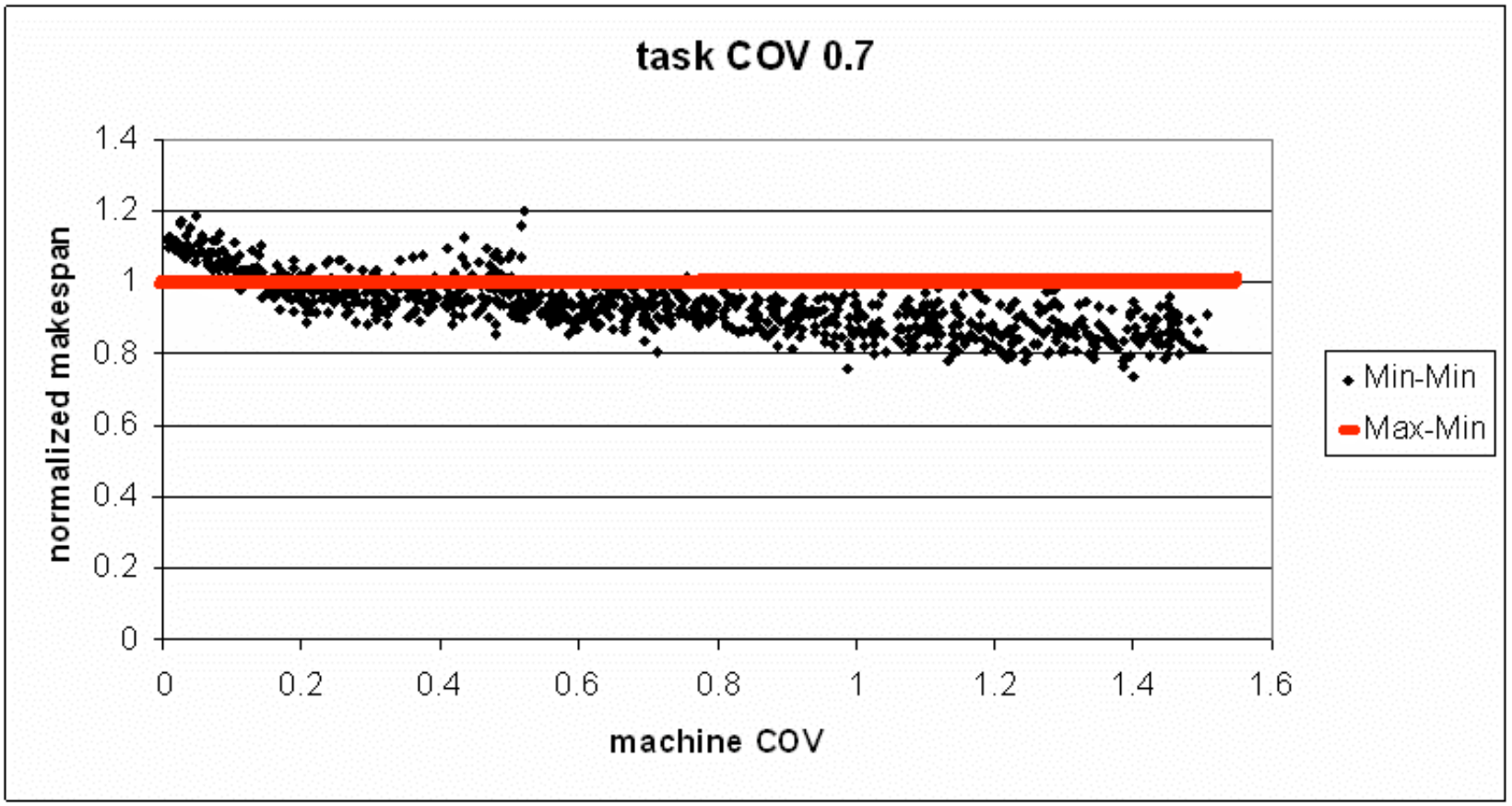
Simulation Results – Effect of ATS

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