NSF/IUCRC CAC PROJECT

INTEGRATED VISUALIZING, MONITORING, AND MANAGING HPC SYSTEMS

Jie Li

Doctoral Student, TTU

09/25/2020

Advisors:

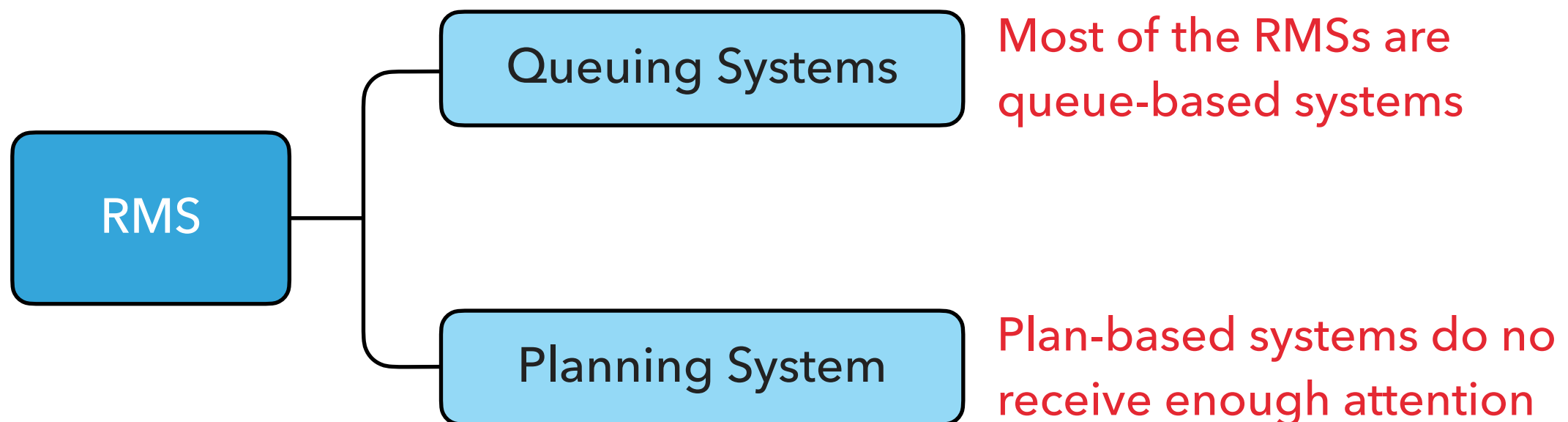Mr. Jon Hass, SW Architect, Dell Inc.

Dr. Alan Sill, Managing Director, HPCC, TTU

Dr. Yong Chen, Associate Professor, CS Dept, TTU
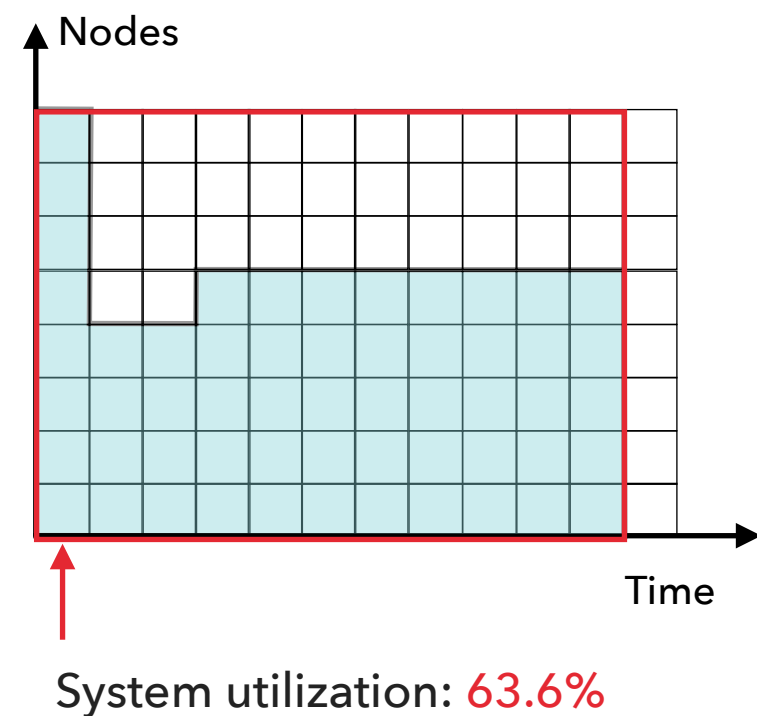
Dr. Tommy Dang, Assistant Professor, CS Dept, TTU

Revisiting Scheduling in HPC Resource Management Systems(RMS)

- RMS: schedule and allocate computing resources to jobs

  - Scheduling: the process of computing a schedule

  - Jobs: consist of information about processing environment, startup parameters, resource request etc.

    - Resource request: the number of requested resources and a duration of how long the resources are requested

```
                    ┌─────────────────────┐      Most of the RMSs are
                    │  Queuing Systems    │      queue-based systems
                    └─────────────────────┘
┌───────────┐
│    RMS    │
└───────────┘
                    ┌─────────────────────┐      Plan-based systems do no
                    │  Planning System    │      receive enough attention
                    └─────────────────────┘
```

Hovestadt, Matthias, Odej Kao, Axel Keller, and Achim Streit. "Scheduling in HPC resource management systems: Queuing vs. planning." In *Workshop on Job Scheduling Strategies for Parallel Processing*, pp. 1-20. Springer, Berlin, Heidelberg, 2003.
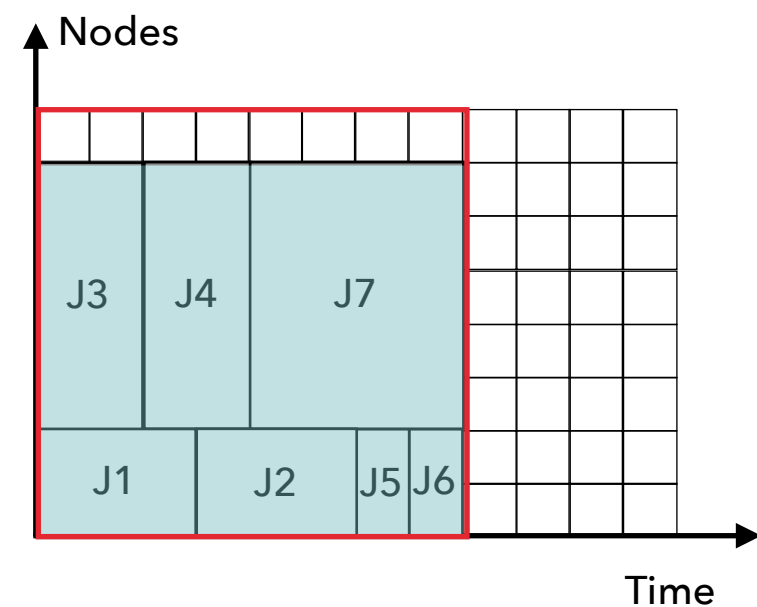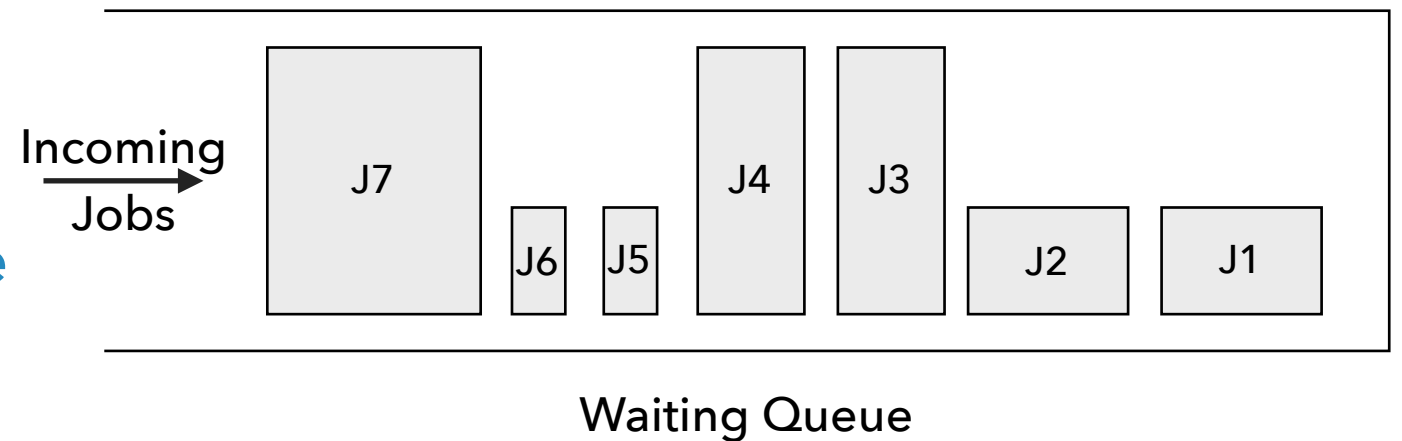
QUEUING SYSTEMS

- The highest prioritized request is always the queue head

- Queuing strategies: FCFS, shortest/ longest job first, fair-sharing etc.

- If not enough resources are available, the system waits until enough resources become available

- Idle resources may be utilized with less prioritized requests by backfilling mechanisms

  - Conservative backfilling: schedule as many jobs as possible

  - EASY backfilling: schedule the first less prioritized job for simplicity

Incoming Jobs →

| J7 | J6 | J5 | J4 | J3 | J2 | J1 |

Waiting Queue

Nodes

Time

System utilization: 63.6%

- Scheduling is fast but compromising system performance even with backfilling

Hovestadt, Matthias, Odej Kao, Axel Keller, and Achim Streit. "Scheduling in HPC resource management systems: Queuing vs. planning." In *Workshop on Job Scheduling Strategies for Parallel Processing*, pp. 1-20. Springer, Berlin, Heidelberg, 2003.
Zheng, Xingwu, Zhou Zhou, Xu Yang, Zhiling Lan, and Jia Wang. "Exploring plan-based scheduling for large-scale computing systems." In *2016 IEEE International Conference on Cluster Computing (CLUSTER)*, pp. 259-268. IEEE, 2016.

- Assign start times and location for each job in the waiting queue

- Consider not only currently available resource but also predictions on job execution time

- A new schedule is computed when a new request is submitted or a running request ends before its estimated finish time

- Achieve a better overall system utilization

- The cost of scheduling is higher than in queueing systems

- Inaccurate estimation of runtime degrades system utilization

Waiting Queue

System utilization: 87.5%

Hovestadt, Matthias, Odej Kao, Axel Keller, and Achim Streit. "Scheduling in HPC resource management systems: Queuing vs. planning." In *Workshop on Job Scheduling Strategies for Parallel Processing*, pp. 1-20. Springer, Berlin, Heidelberg, 2003.
Zheng, Xingwu, Zhou Zhou, Xu Yang, Zhiling Lan, and Jia Wang. "Exploring plan-based scheduling for large-scale computing systems." In *2016 IEEE International Conference on Cluster Computing (CLUSTER)*, pp. 259-268. IEEE, 2016.

QUESTIONS?/COMMENTS?