

# Kronos: Towards Bus Contention-Aware Job Scheduling in Warehouse Scale Computers

Shuai Xue, Shang Zhao, Quan Chen, Zhuo Song, Shanpei Chen, Tao Ma, Yong Yang, Wenli Zheng, Minyi Guo

Frontiers of Computer Science, DOI: [10.1007/s11704-021-0418-5](https://doi.org/10.1007/s11704-021-0418-5)

# Problems & Ideas

- Problems of mitigating memory bus contention due to split lock
  - Co-located VMs share resources of host node, e.g. memory bus
  - Split locks are pervasive in cloud and result in server memory bus contention
  - Co-located VMs suffer from **9X performance degradation** in worst case
- Ideas: Bus contention-aware Job scheduling
  - An online bus contention tolerance meter
  - A bus contention-aware job scheduler

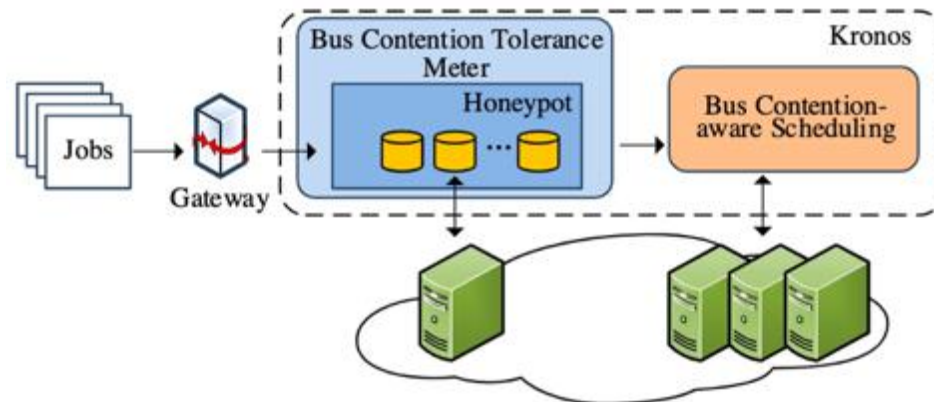


Fig.1: The design architecture of Kronos

# Main Contributions

- **Policy 1 : reduces 42.1% of physical node while ensuring the SLA**

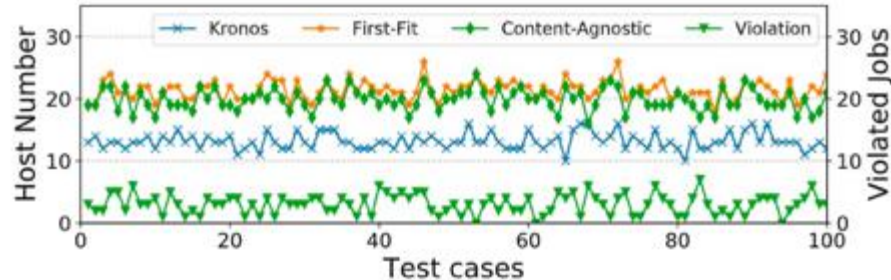


Fig.2: The number of required physical nodes to fulfill the SLA

- **Policy 2 : reduces 72.8% SLA violations**

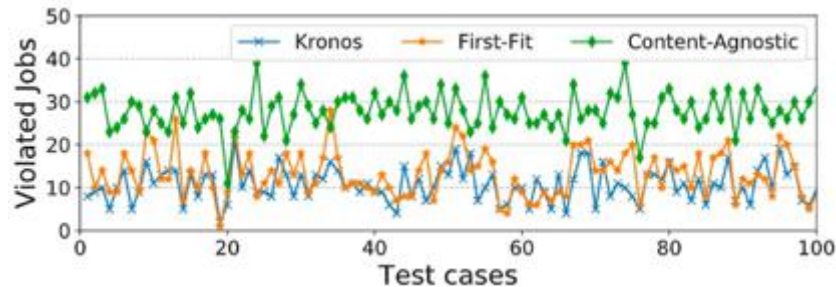


Fig.3: The number of jobs that suffer from SLA violation

- **Policy 3 : improves the overall performance by 35.2%**

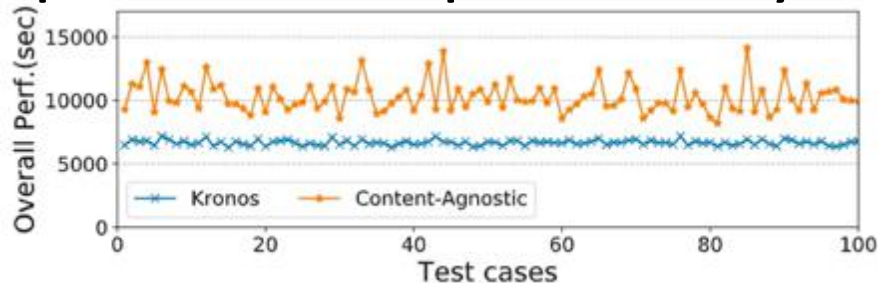


Fig.4: The overall running time performance