# Toward Accurate Group Scheduling in Multi-core Environments

Kenta Ishiguro*, Kenichi Yasukata**, Toshio Hirotsu*
* Hosei University ** IIJ Research Laboratory

## Background

- **Bandwidth control of a CPU scheduler** is used to isolate CPU resources between VMs
- Linux Completely Fair Scheduler (CFS) implements it as part of the ***group scheduling*** feature
  - Virtual CPUs are (POSIX) threads and they are grouped by each VM

## Challenge

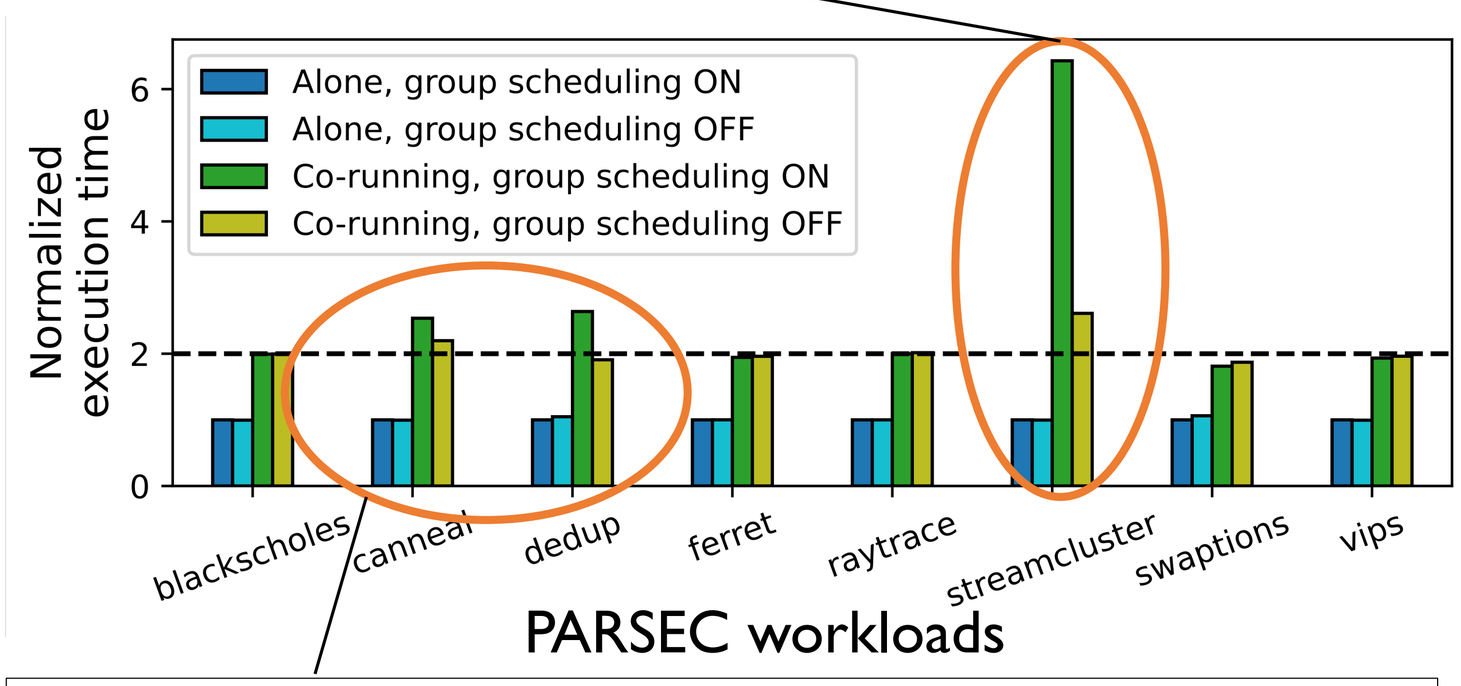
Hard to achieve both simultaneously
- Accurate CPU time assignment to each group across **multiple cores**
- Scheduling processes/threads in a **work-conserving** manner

## Experiment setup

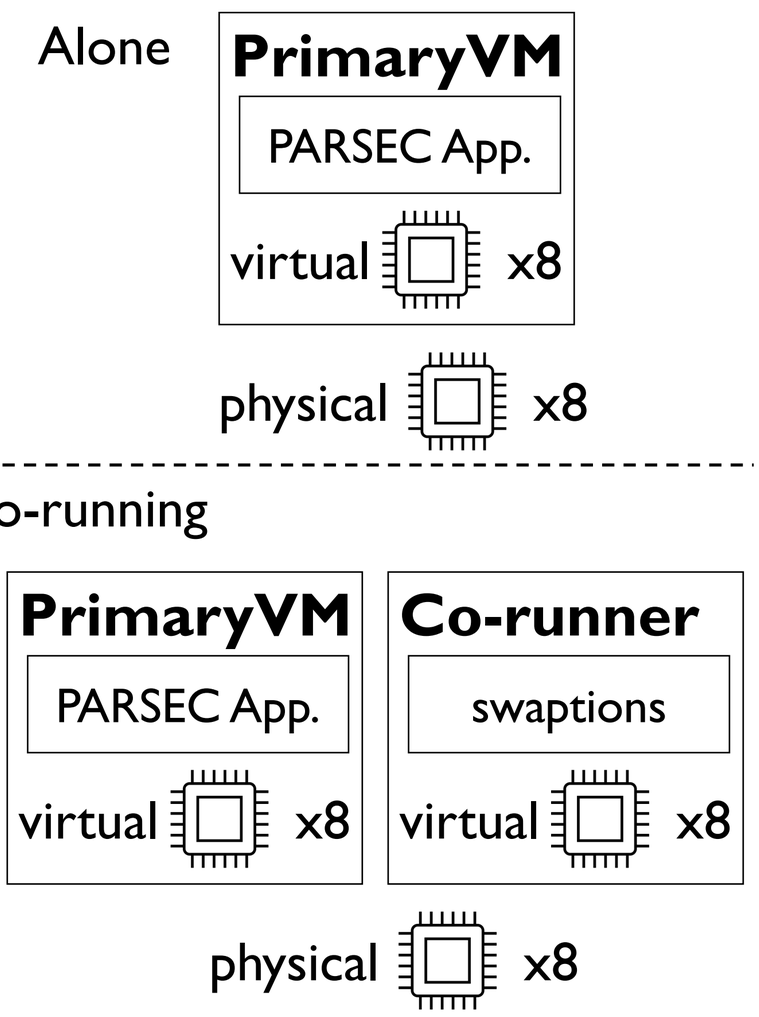
| | |
|---|---|
| CPU | Intel Xeon E-2378G<br>8 cores (2.80 GHz) |
| DRAM | 64 GB |
| Linux/KVM | v5.15.64 |

## Observation: significant slow down due to group scheduling

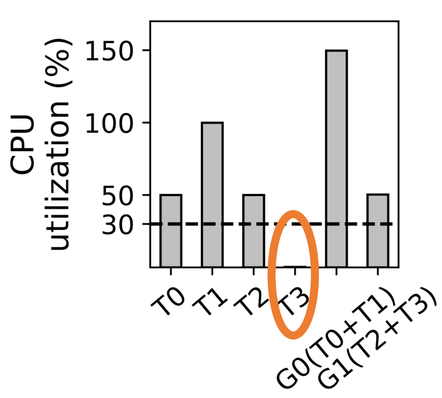
streamcluster w/ group scheduling slows down by **6x** (> 2x)
while w/o group scheduling slows down by **2.8x**




**slowdown (> 2x)** can be seen other than streamcluster



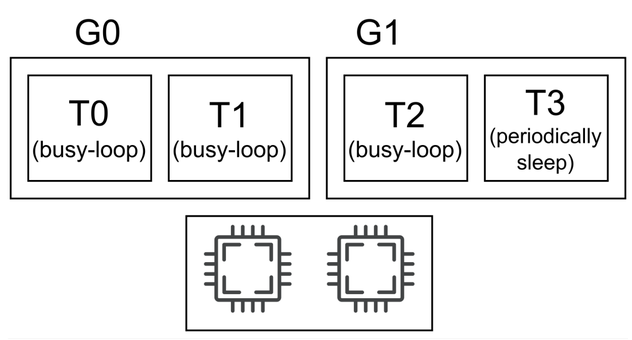

## What is causing the slowdown?




- **Degradation** of CPU utilization of **non-CPU-intensive** threads
- T3 (non-CPU-intensive) is almost never scheduled
- CPU time of T3 is consumed by T1

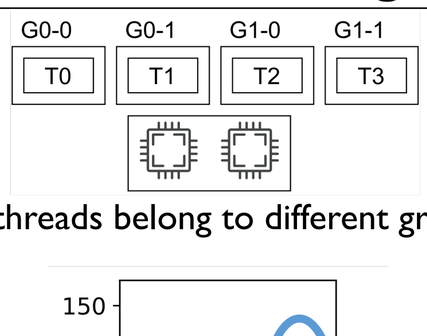
Synthetic workload
- 2 groups (G0, G1) (same weight)
- 4 threads (T0-3) (same weight)
- T0-2: busy-loop
- T3: periodically sleep (max CPU util.: 30%)

**Group (Naïve) configuration**

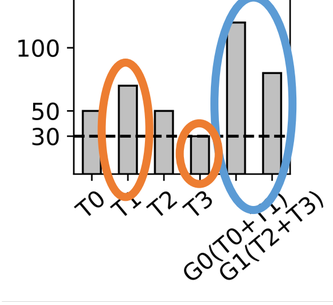



## No perfect configurations

**Individual config**

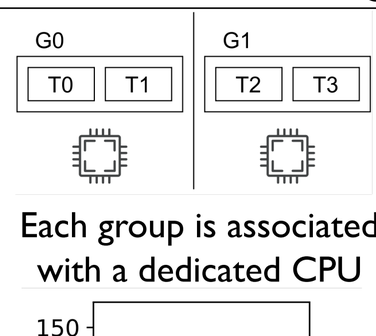



All threads belong to different groups

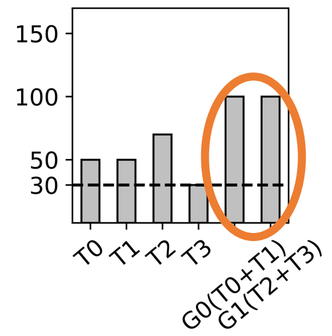



😄 T3 achieves 30%
😞 G0 (120%) vs G1(80%)

**Dedicated config**




Each group is associated with a dedicated CPU




😄 Ideal bandwidth control
😞 Load balancing is unavailable

## Conclusion & Future work

- Group scheduling causes significant performance degradation in some workloads
- Slowdown derives from degradation of CPU utilization of non-CPU-intensive workloads
- No config exists to achieve ideal group scheduling

| | Per-core | Aggregated | Load balance |
|---|---|---|---|
| Group | ✗ | ✗ | ✔ |
| Individual | ✔ | ✗ | ✔ |
| Dedicated | ✔ | ✔ | ✗ |
| Our goal | ✔ | ✔ | ✔ |

This work was supported in part by Japan Science and Technology Agency (JST CREST JPMJCR21M4).