Lecture 38

23 October 2018 10:35

scheduling policies

- pre-emptive vs. non-preemptive
  work conserving vs non-work conserving
  real time vs. interactive vs compute centric
  } categorization of schedulers.

- metrics
  - response time (user-level metric)
  - distr. of cpu util — cpu. util.
  - wait time / queuing delays
  - throughput / size of run queues
  - ~ scheduler overheads.

  

  || fairness . proportionate behaviours.

sample policies.

- FIFO / FCFS ~ work-conserving. non-deterministic worst-case large | wait times.
- RR — round robin
  - time quantum / slot — pick processes for every slot in a round robin manner.

    ??

    t — very fair / low WT, efficiency?

    t >> 0 ~ start looking like FIFO

- WRR
  DRR
  weighted RR
  deficit RR
  || proportionate scheduling. time quantum is associated with a priority / weight.

  deficit RR — collect time units in a slot if not ~~util~~ utilized (deficit ≤ max)

- Lottery scheduling.

  RT
  interactive

- lottery scheduling.
- priority scheduling.
  - interactive
  - foreground
  - background.

  } RT

  - multi-queue priorities
    - schedule processes in higher priority before others.
    - starvation.

① yesteryear Linux
- used multi-queue priorities
- assigned time intervals per process
- (shifted) processes to higher priorities based on wait times.

Linux processes
- priority
- nice

② CFS — completely fair scheduler
- priority, nice (inputs).
- 'N' processes, assign $\frac{1}{N}$ of cpu to each.

- SCHED_NORMAL
- SCHED_FIFO
- SCHED_RR

} RT

scheduling class
own scheduling policy

① vruntime ~ normalized runtime of each process on cpu.
est.
of ~~each~~ scheduled process

② schedule process w/ lowest vruntime