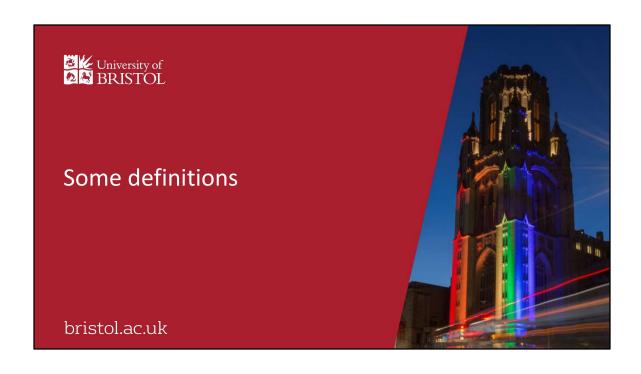


Computer Systems B COMS20012

Introduction to Operating Systems and Security



What is scheduling

- The act of selecting a process/thread to run
- Abstraction: threads (see last week)
- Mechanism: context switching (see last week)
- Policy: scheduling (this week)

Goal of scheduling

- Preserve the illusion that a process has sole access to the hardware
- Processes should be oblivious to scheduling
- Each process (in the absence of internal thread) act as it was a sequential process with full control of the hardware
- Two flavors of resource
 - Preemptable: can take the resource away
 - Need a scheduling policy: How long do you get the resource? In what order do you grant the resource?
 - ➤ e.g., memory
 - Non-preemptable: cannot take the resource away
 - > Need an allocation policy: Who get the resource?
 - ➤ e.g., printer

Metrics for a scheduling policy

- Throughput: efficiency of resource utilization
 - Keep the CPU, disk etc. busy
- Latency: minimize response time
- Fairness: distribute resource equitably
 - What does that mean?

Metrics for a scheduling policy

- Throughput: efficiency of resource utilization
 - Keep the CPU, disk etc. busy
- Latency: minimize response time
- Fairness: distribute resource equitably
 - What does that mean?
 - > As in real life no clear-cut answer

Metrics for a scheduling policy

- Throughput: efficiency of resource utilization
 - Keep the CPU, disk etc. busy
- Latency: minimize response time
- Fairness: distribute resource equitably
 - What does that mean?
 - > As in real life no clear-cut answer
 - > Faculty members get more resources?
 - > Final year students get more resources?
 - > Everyone get the same number of resources?
 - > You have more work to do so you get more resources?

