



COMPUTER ENGINEERING



UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

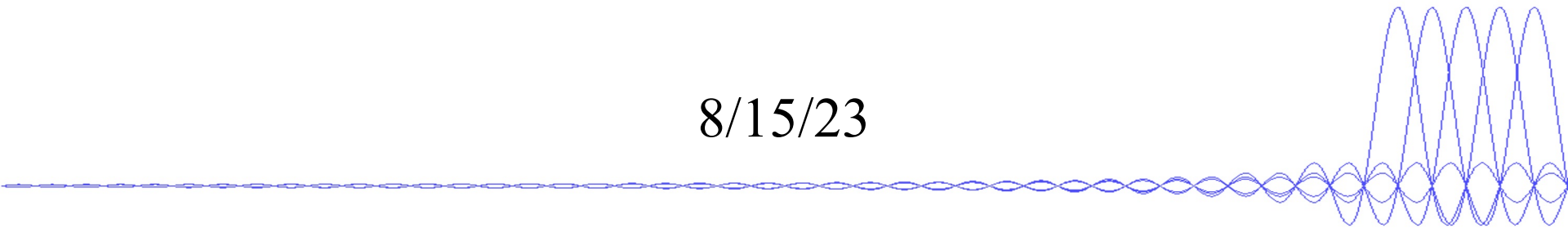
HỆ ĐIỀU HÀNH

Chương 4

ĐỊNH THỜI CPU

(Phần một)

8/15/23





Mục tiêu Chương 4

- Biết được các khái niệm cơ bản về định thời
- Biết được các tiêu chuẩn định thời CPU
- Hiểu được các giải thuật định thời
- Vận dụng các giải thuật định thời để làm bài tập và mô phỏng



Nội dung Chương 4

- Các khái niệm cơ bản về định thời
- Các loại định thời
- Các tiêu chuẩn định thời CPU
- Các giải thuật định thời
 - First-Come, First-Served (FCFS)
 - Shortest Job First (SJF)
 - Shortest Remaining Time First (SRTF)
 - Priority Scheduling

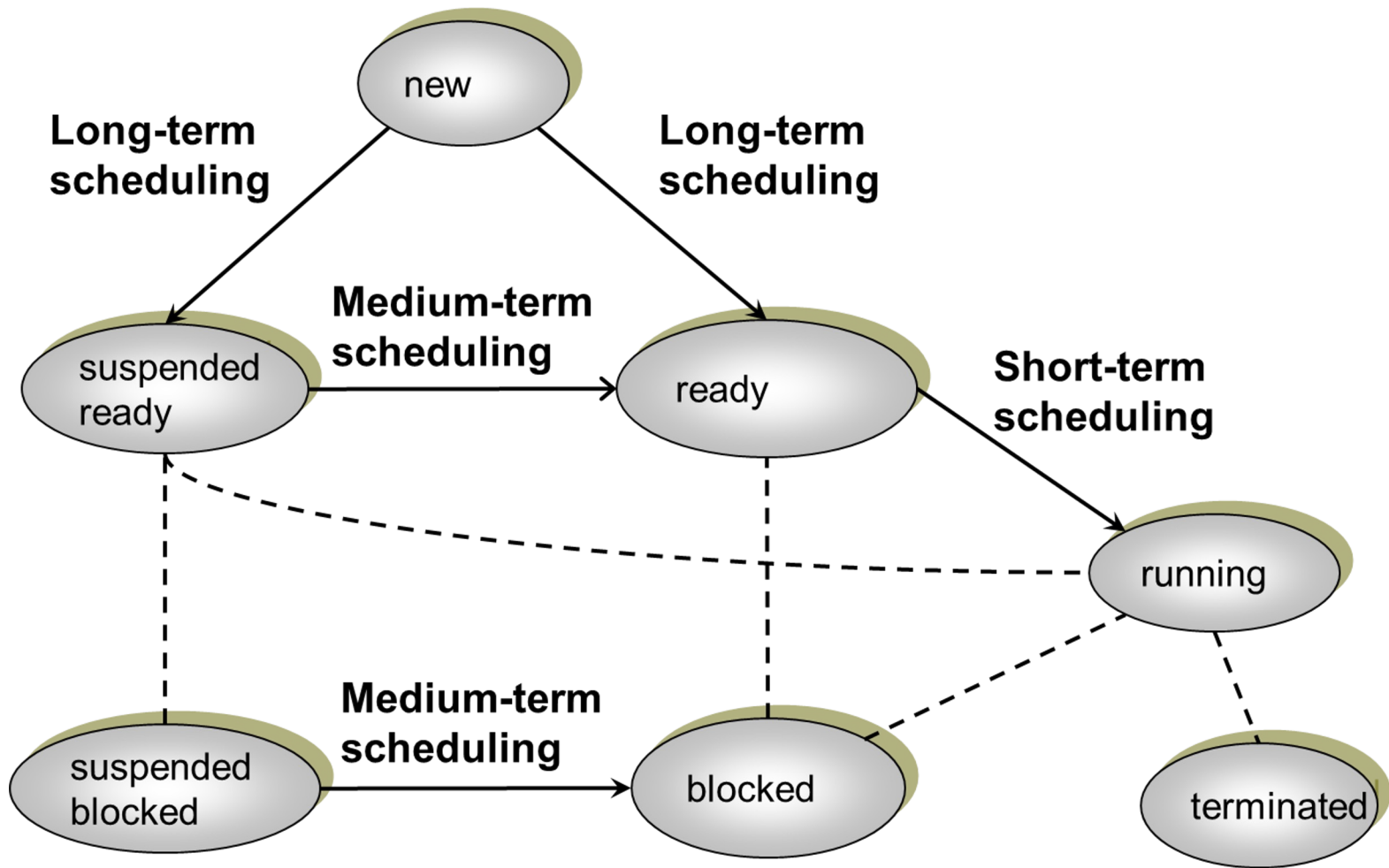


Khái niệm cơ bản

- Trong các hệ thống đa nhiệm (multitasking), đơn vi xử lý
 - ❑ Cho phép thực thi đồng thời nhiều chương trình để làm tăng hiệu suất hệ thống
(Cho phép nhiều chương trình được nạp vào bộ nhớ)
 - ❑ Tại mỗi thời điểm, chỉ có một tiến trình được thực thi
- Cần phải giải quyết vấn đề phân chia, lựa chọn tiến trình thực thi để đạt được hiệu quả cao nhất.
- Cần có những phương pháp chọn lựa phù hợp
- **Định thời** là chiến lược lựa chọn tiến trình phù hợp để được thực thi sao cho đạt được hiệu quả cao nhất



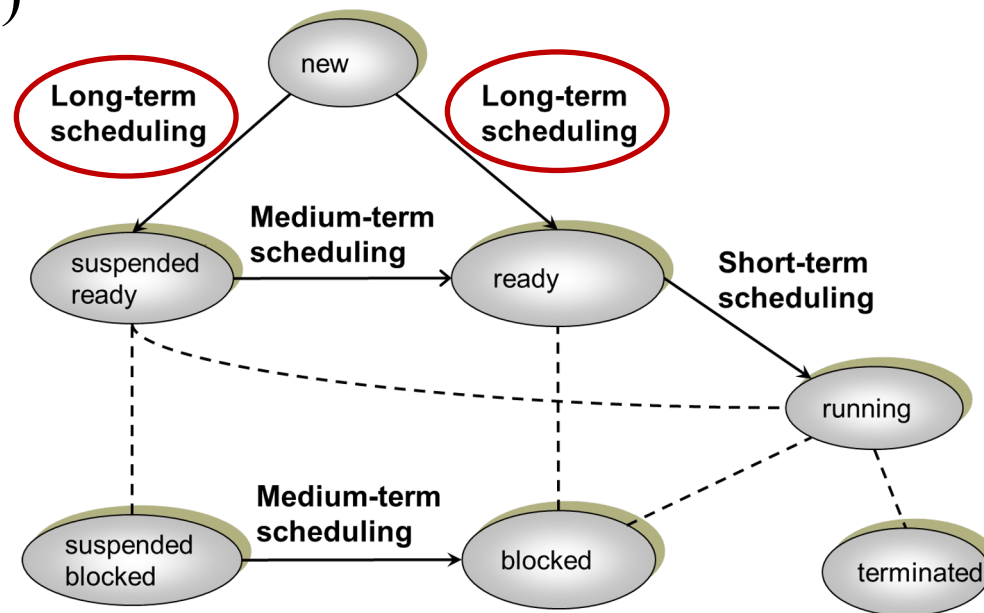
Các loại định thời





Định thời dài (Long-term Scheduling)

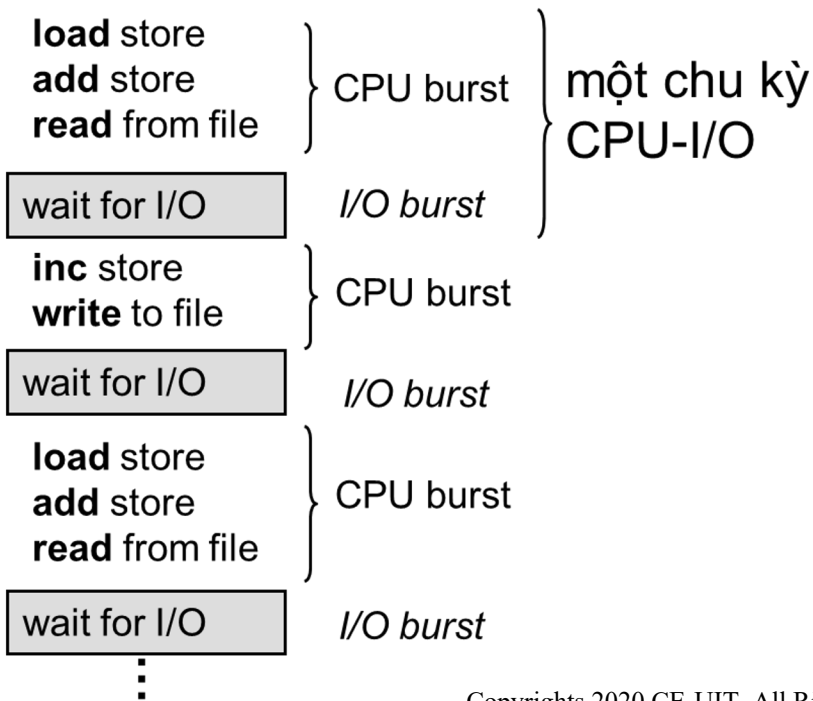
- Xác định chương trình nào được chấp nhận nạp vào hệ thống để thực thi
- Điều khiển mức độ đa chương của hệ thống
- Định thời dài thường cố gắng duy trì xen lẫn giữa tiến trình hướng CPU (CPU-bound process) và tiến trình hướng I/O (I/O-bound process)





Chu kỳ CPU-I/O

- **Service time** là thời gian một tiến trình cần CPU trong một chu kỳ CPU-I/O (hay còn gọi là *burst time*)
- Tiến trình có *service time* lớn được gọi là các tiến trình hướng CPU (CPU-bound process).



Process	Arrival Time	Service Time
1	0	3
2	2	6
3	4	4
4	6	5
5	8	2

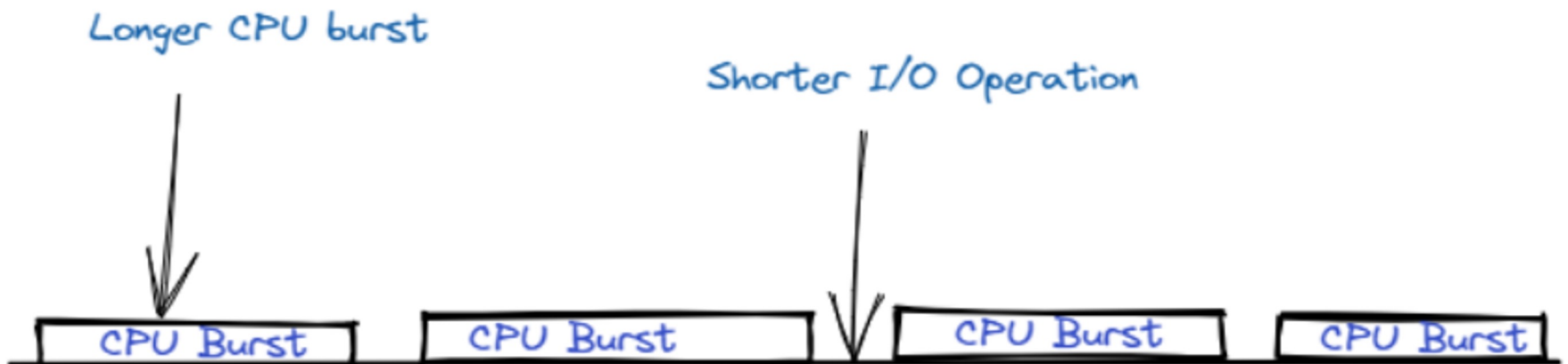


Tiến trình hướng CPU (CPU-bound)

- Tiến trình yêu cầu thời gian thực thi trên CPU nhiều
- Thời gian hoàn thành chương trình phụ thuộc vào tốc độ thực thi của CPU

```
#include <stdio.h>
```

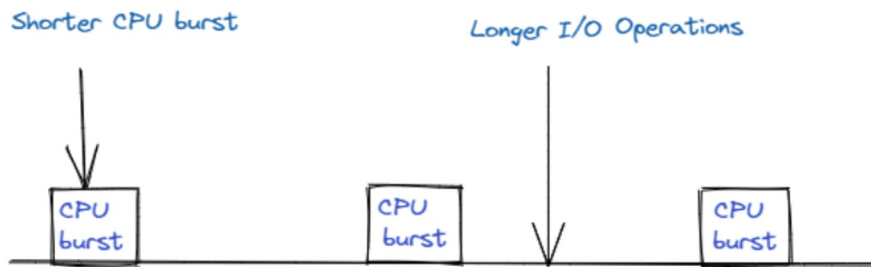
```
int main() {  
    long long start = 1, end = 1000000, total = 0;  
    for (long long i = start; i <= end; i++) {  
        total += i;  
    }  
    printf("Sum of numbers from %lld to %lld is  
%lld\n", start, end, total);  
    return 0;  
}
```





Tiến trình hướng I/O (I/O-bound)

- Tiến trình yêu cầu thời gian thực thi trên ngoại vi nhiều hơn
- Thời gian hoàn thành chương trình phụ thuộc chu kỳ đợi cho các thao tác nhập/xuất

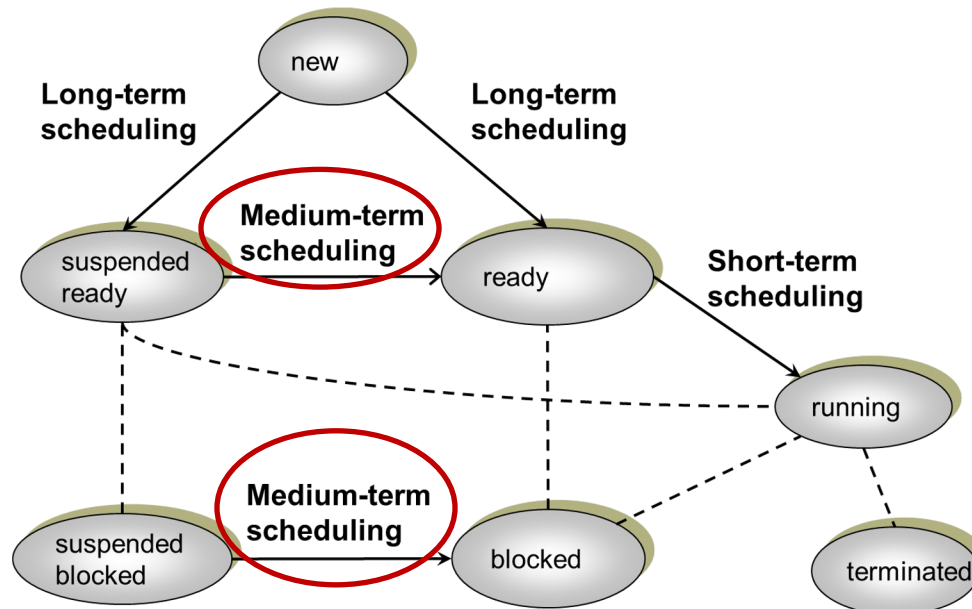


```
#include <stdio.h>
int main() {
    FILE *fp;
    char filename[] = "example.txt";
    int total = 0, ch;
    fp = fopen(filename, "r");
    if (fp == NULL) {
        printf("Failed to open file %s\n", filename);
        return 1;
    }
    while ((ch = fgetc(fp)) != EOF) {
        total++;
    }
    fclose(fp);
    printf("Total number of characters in file %s is %d\n", filename, total);
    return 0;
}
```



Định thời vừa (Medium-term scheduling)

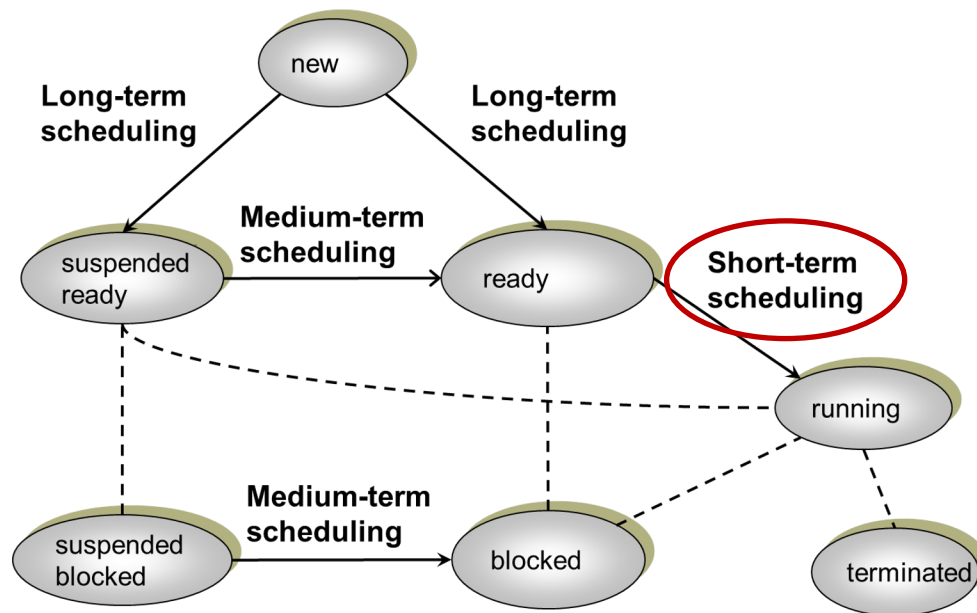
- Định thời trung hạn quyết định tiến trình nào được đưa vào (swap in) và đưa ra khỏi (swap out) bộ nhớ chính trong quá trình thực thi của hệ thống
- Được thực hiện bởi trình quản lý bộ nhớ (và sẽ được thảo luận ở chương về quản lý bộ nhớ).





Định thời ngắn (Short-term scheduling)

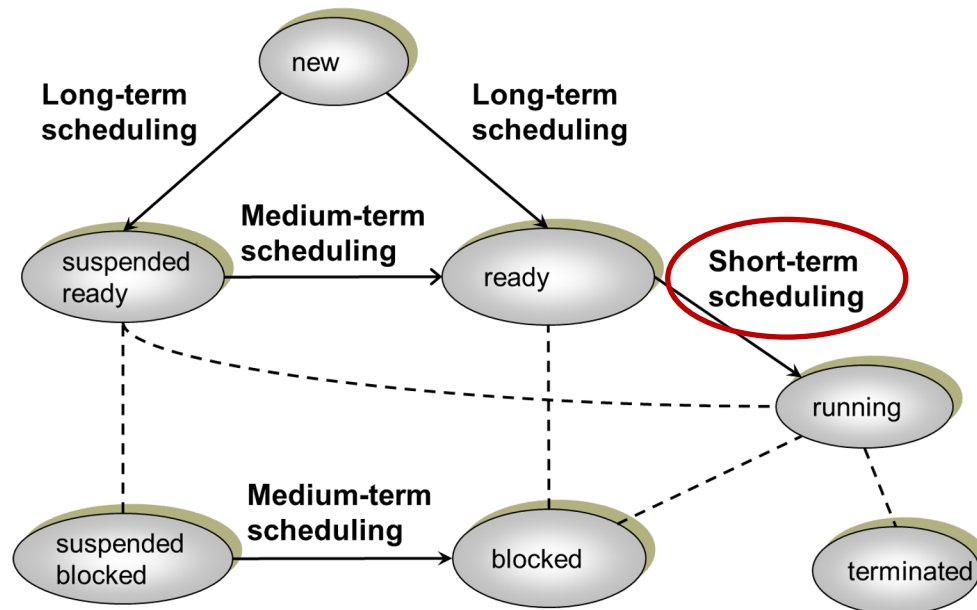
- Còn được gọi là định thời CPU (CPU scheduling).
- Xác định tiến trình nào trong *hạng đợi sẵn sàng* (*ready queue*) sẽ được chiếm CPU để thực thi kế tiếp.
- Đối với hệ thống hỗ trợ nhân đa luồng (multithreaded kernel), việc định thời CPU là do HĐH chọn *kernel thread* được chiếm CPU





Định thời ngắn hạn (tt)

- Bộ định thời ngắn được gọi khi có một trong các sự kiện/interrupt sau xảy ra:
 - Ngắt thời gian (clock interrupt)
 - Ngắt ngoại vi (I/O interrupt)
 - Lời gọi hệ thống (operating system call)
 - Tín hiệu đồng bộ hóa (Sẽ trao đổi sau ở Chương 5)





Bộ định thời ngắn hạn (CPU Scheduler)

- Bộ định thời sẽ chuyển quyền điều khiển CPU về cho tiến trình được chọn.
- Quá trình chuyển đổi bao gồm:
 - Chuyển ngữ cảnh (sử dụng thông tin ngữ cảnh trong PCB)
 - Chuyển chế độ người dùng
 - Nhảy đến vị trí thích hợp trong chương trình ứng dụng để khởi động lại chương trình (sử dụng thông tin địa chỉ tại *program counter* trong PCB)
- Công việc này gây ra phí tổn
 - ***Dispatch latency***: thời gian mà bộ định thời dừng một tiến trình và khởi động một tiến trình khác



Các tiêu chuẩn định thời CPU

- Hướng người dùng (User-oriented)
 - **Thời gian đáp ứng** (Response time): khoảng thời gian từ lúc tiến trình gửi yêu cầu thực thi đến khi yêu cầu được đáp ứng lần đầu tiên (trong các hệ thống time-sharing, interactive system) → cực tiểu
 - **Thời gian hoàn thành** (Turnaround time): khoảng thời gian từ lúc một tiến trình được nạp vào hệ thống đến khi tiến trình đó kết thúc → cực tiểu
 - **Thời gian đợi** (Waiting time): tổng thời gian một tiến trình đợi trong ready queue → cực tiểu



The diagram shows a horizontal axis labeled t with points 0 , r , t_0 , and f . Three orange rectangular pulses are shown, each labeled P . The first pulse starts at t_0 and ends at $t_0 + E_1$. The second pulse starts at $t_0 + E_2$ and ends at $t_0 + E_2 + E_2$. The third pulse starts at $t_0 + E_2 + E_2 + E_3$ and ends at f . A horizontal double-headed arrow labeled R spans from r to t_0 . A horizontal double-headed arrow labeled F spans from r to f . Above the axis, a blue wavy line represents a signal, and a horizontal line represents the baseline.

- Quá trình thực thi một tiến trình **P** gồm nhiều phần
- r là thời điểm xuất hiện của P trong hệ thống (Arrival time/Release time)
- t_0 là thời điểm P được thực thi lần đầu tiên
- f là thời điểm tiến trình P hoàn thành việc thực thi (Finishing time)

$$\begin{aligned} R &= t_0 - r \\ F &= f - r \\ W &= f - r - E \end{aligned}$$
$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$$



Các tiêu chuẩn định thời CPU (tt)

- **Hướng hệ thống (System-oriented)**
 - **Hiệu năng sử dụng CPU** (processor utilization): định thời sao cho CPU càng bận càng tốt → cực đại
 - **Tính công bằng** (fairness): tất cả tiến trình phải được đối xử như nhau
 - **Thông lượng** (throughput): số tiến trình hoàn tất công việc trong một đơn vị thời gian → cực đại



Giải thuật định thời

Một giải thuật định thời thông thường bao gồm hai yếu tố:

- **Hàm chọn lựa** (selection function): mô tả cách thức (căn cứ) để chọn tiến trình nào trong *ready queue* được thực thi (Các hàm chọn lựa thường được xây dựng dựa trên độ ưu tiên, yêu cầu về tài nguyên, đặc điểm thực thi của tiến trình,...)
- **Chế độ quyết định** (decision mode): quyết định thời điểm thực hiện hàm chọn lựa để định thời



Các chế độ quyết định

Có hai chế độ quyết định thường được áp dụng:

- Không trung dụng (Non-preemptive)
 - Khi ở trạng thái running, tiến trình sẽ thực thi cho đến khi kết thúc hoặc bị ngắt (blocked) do yêu cầu I/O.
- Trung dụng (Preemptive)
 - Tiến trình đang thực thi (ở trạng thái running) có thể bị ngắt giữa chừng và chuyển về trạng thái ready.
 - Chi phí cao hơn chế độ không trung dụng nhưng đánh đổi lại bằng thời gian đáp ứng tốt hơn vì không có trường hợp một tiến trình độc chiếm CPU quá lâu.



Thời điểm thực thi hàm chọn lựa

- Hàm chọn lựa được thực thi vào các thời điểm sau:
 - (1) Có tiến trình chuyển từ trạng thái running sang waiting
 - (2) Có tiến trình chuyển từ trạng thái running sang ready
 - (3) Có tiến trình chuyển từ trạng thái waiting, new sang ready
 - (4) Kết thúc thực thi của một tiến trình
- (1) và (4) không cần lựa chọn loại định thời, (2) và (3) cần
- Việc thực thi hàm chọn lựa trong trường hợp (1) và (4) không phụ thuộc vào loại giải thuật định thời và thường áp dụng chế độ không trung dụng.
- Ngược lại, trường hợp (2) và (3) phụ thuộc vào loại giải thuật định thời và thường áp dụng chế độ trung dụng.
- **Thực hiện theo cơ chế nào khó hơn? Tại sao?**



Các giải thuật định thời

- First-Come, First-Served (FCFS)
- Shortest Job First (SJF)
- Shortest Remaining Time First (SRTF)
- Round-Robin (RR)
- Priority Scheduling
- Highest Response Ratio Next (HRRN)
- Multilevel Queue
- Multilevel Feedback Queue



First-Come, First-Served (FCFS)

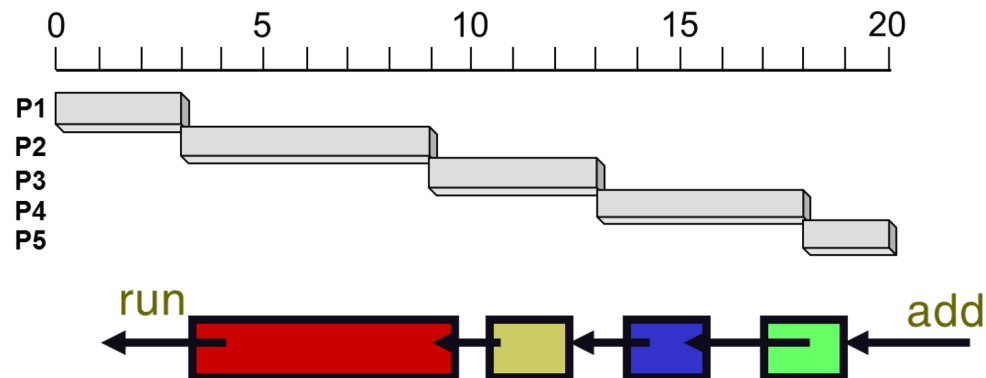
■ Hàm lựa chọn:

- Tiến trình nào yêu cầu CPU trước sẽ được cấp phát CPU trước
- Tiến trình sẽ thực thi đến khi kết thúc hoặc bị blocked do I/O

■ Chế độ quyết định: không trung dụng (non-preemptive)

■ Hiện thực: sử dụng hàng đợi FIFO (FIFO queues)

- ☐ Tiến trình mới xuất hiện được thêm vào cuối hàng đợi
- ☐ Tiến trình được lựa chọn để xử lý được lấy từ đầu của hàng đợi

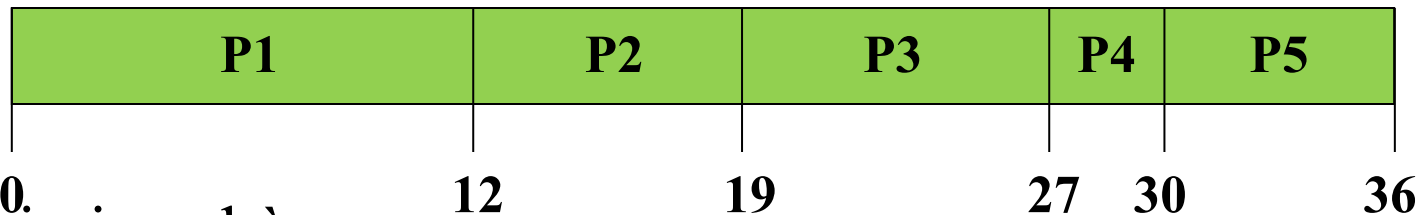




First-Come, First-Served (FCFS)

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian chờ:

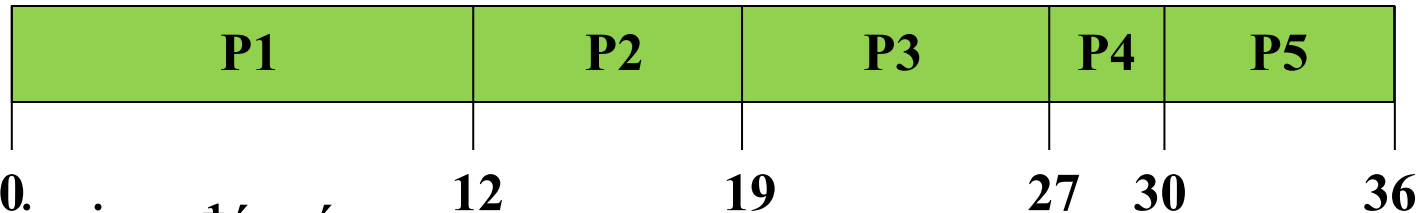
- $P1 = 0, P2 = 10, P3 = 14, P4 = 18, P5 = 18$
- Thời gian chờ trung bình: $(0 + 10 + 14 + 18 + 18)/5 = 12$



First-Come, First-Served (FCFS)

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian đáp ứng:

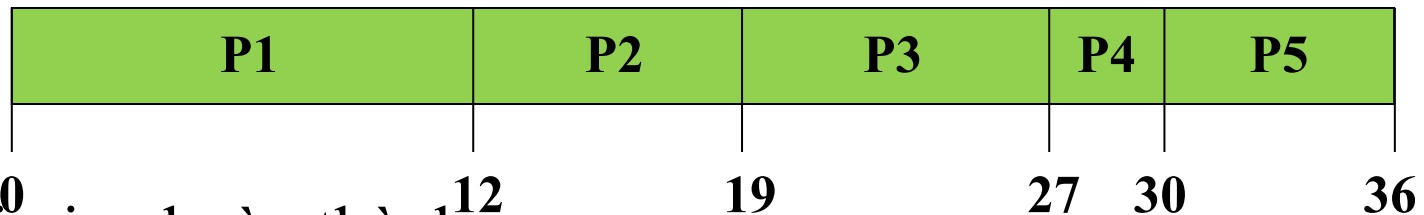
- $P1 = 0, P2 = 10, P3 = 14, P4 = 18, P5 = 18$
- Thời gian đáp ứng trung bình: $(0 + 10 + 14 + 18 + 18)/5 = 12$



First-Come, First-Served (FCFS)

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian hoàn thành:

- $P1 = 12, P2 = 17, P3 = 22, P4 = 21, P5 = 24$
- Thời gian hoàn thành trung bình: $(12 + 17 + 22 + 21 + 24)/5 = 19.2$



Shortest-Job-First (SJF)

- Hàm chọn lựa: tiến trình có thời gian yêu cầu thực thi (CPU burst) ngắn nhất sẽ được chọn
 - Khi CPU trống, HĐH sẽ chọn tiến trình có *CPU burst* ngắn nhất để được thực thi tiếp theo
 - Giải thuật này sử dụng chiều dài thời gian thực thi tiến trình làm căn cứ để chọn lựa.

- SJF có thể được hiện thực với cả hai chiến lược: trung dụng và không trung dụng



Shortest-Job-First (SJF) (tt)

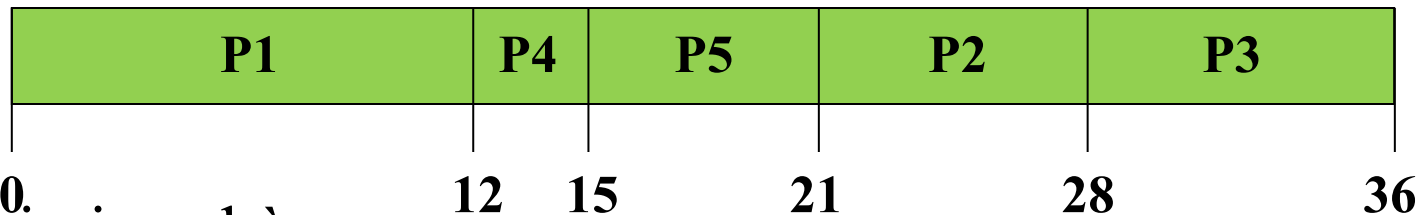
- SJF ở chế độ không trung dụng:
 - Hàm chọn lựa được thực thi khi CPU trống
 - Khi tiến trình được cấp CPU thì sẽ thực thi cho đến khi kết thúc
 - Khi một tiến trình kết thúc, một tiến trình khác có thời gian thực thi ngắn nhất sẽ được chọn



Ví dụ: SJF ở chế độ không trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian chờ:

☐ $P1 = 0, P2 = 19, P3 = 23, P4 = 3, P5 = 3$

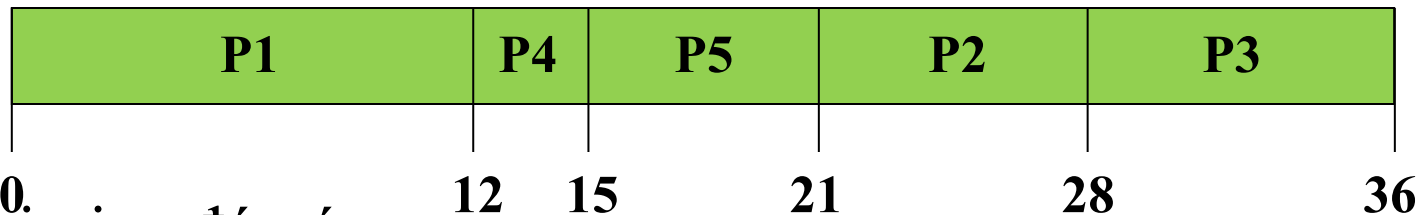
☐ Thời gian chờ trung bình: $(0 + 19 + 23 + 3 + 3)/5 = 9.6$



Ví dụ: SJF ở chế độ không trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian đáp ứng:

☐ $P1 = 0, P2 = 19, P3 = 23, P4 = 3, P5 = 3$

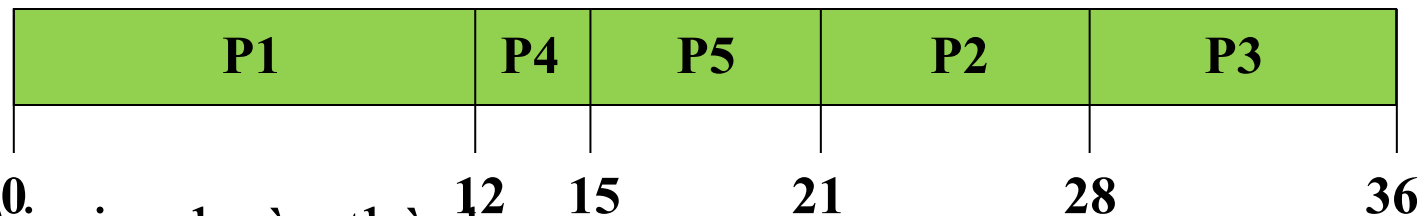
☐ Thời gian đáp ứng trung bình: $(0 + 19 + 23 + 3 + 3)/5 = 9.6$



Ví dụ: SJF ở chế độ không trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian hoàn thành:

☐ $P1 = 12, P2 = 26, P3 = 31, P4 = 6, P5 = 9$

☐ Thời gian hoàn thành trung bình: $(12 + 26 + 31 + 6 + 9)/5 = 16.8$



Shortest-Job-First (SJF) (tt)

■ SJF ở chế độ trung dụng

- Hàm chọn lựa được thực thi khi có tiến trình mới xuất hiện hoặc có tiến trình kết thúc
- Khi có tiến trình mới xuất hiện với *CPU-burst* nhỏ hơn thời gian yêu cầu còn lại (remaining time) của tiến trình đang thực thi, tiến trình mới sẽ được chọn và tiến trình đang thực thi sẽ bị dừng lại.
- Khi một tiến trình kết thúc, một tiến trình khác có *CPU-burst* (hoặc thời gian yêu cầu còn lại) nhỏ nhất sẽ được chọn tiếp theo.
- SJF ở chế độ trung dụng còn được gọi là Shortest-Remaining-Time-First (SRTF)

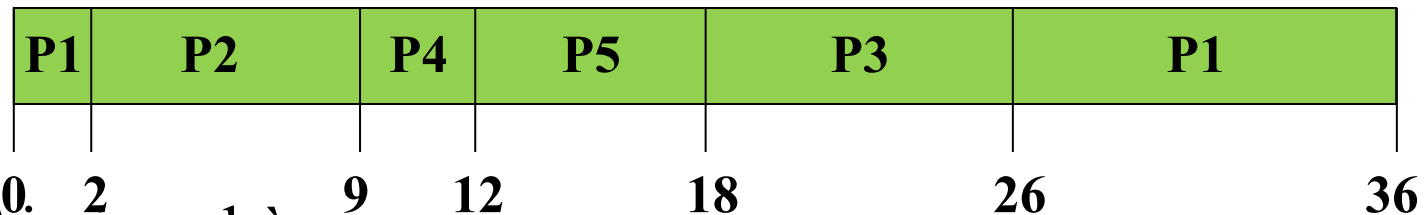
■ SJF là tối ưu về thời gian đợi: có thời gian chờ đợi trung bình ngắn nhất với một tập tiến trình cho trước



Ví dụ: SJF ở chế độ trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian chờ:

☐ $P1 = 24, P2 = 0, P3 = 13, P4 = 0, P5 = 0$

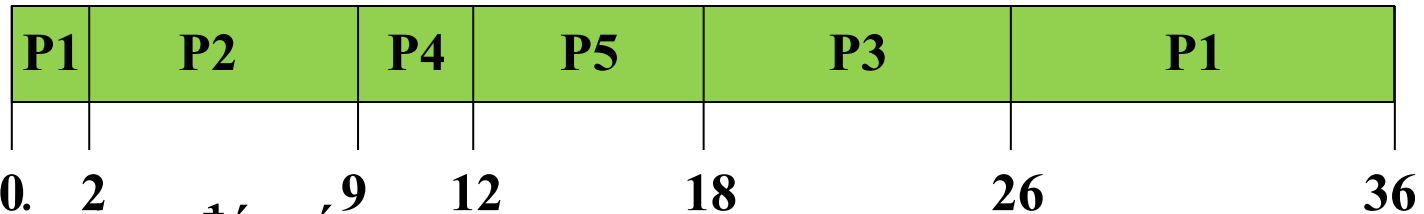
☐ Thời gian chờ trung bình: $(24 + 0 + 13 + 0 + 0)/5 = 7.4$



Ví dụ: SJF ở chế độ trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian đáp ứng:

☐ $P1 = 0, P2 = 0, P3 = 13, P4 = 0, P5 = 0$

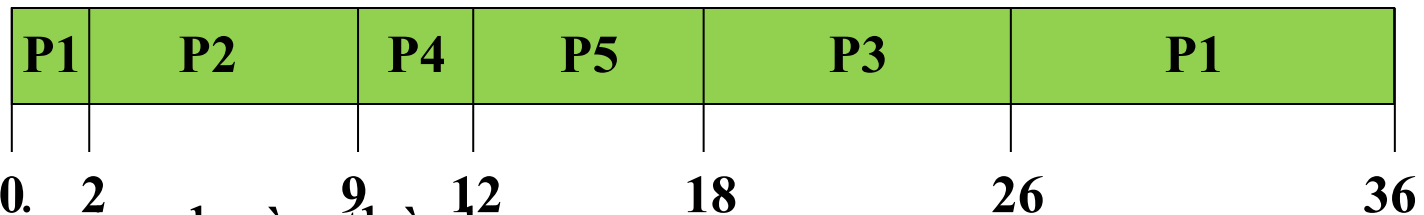
☐ Thời gian đáp ứng trung bình: $(0 + 0 + 13 + 0 + 0)/5 = 2.6$



Ví dụ: SJF ở chế độ trung dụng

Process	Arrival Time	Burst Time
P1	0	12
P2	2	7
P3	5	8
P4	9	3
P5	12	6

■ Giải đồ Gantt



■ Thời gian hoàn thành:

☐ $P1 = 36, P2 = 7, P3 = 21, P4 = 3, P5 = 6$

☐ Thời gian hoàn thành trung bình: $(36 + 7 + 21 + 3 + 6)/5 = 14.6$



Nhận xét về giải thuật SJF

- Có thể xảy ra tình trạng “đói” tài nguyên (starvation) đối với các tiến trình có *CPU-burst* lớn nếu có nhiều tiến trình với *CPU-burst* nhỏ (liên tục) xuất hiện trong hệ thống.
- Cơ chế không trung dụng không phù hợp cho hệ thống time sharing (interactive).
- Giải thuật SJF ngầm định rằng độ ưu tiên được xác định dựa theo độ dài *CPU-burst*.
- Các tiến trình hướng CPU (CPU-bound) có độ ưu tiên thấp hơn so với tiến trình hướng I/O (I/O-bound)
- Tuy nhiên, khi một tiến trình hướng CPU được thực thi thì nó độc chiếm CPU cho đến khi kết thúc.



Nhận xét về giải thuật SJF (tt)

- Ưu điểm: SJF tối ưu trong việc giảm thời gian đợi trung bình
- Hạn chế: Cần phải ước lượng thời gian cần CPU tiếp theo của process
- Giải pháp cho vấn đề này?

[1] P. Samal, S. Jha and R. K. Goyal, "**CPU Burst-Time Estimation using Machine Learning**," 2022 *IEEE Delhi Section Conference (DELCON)*, New Delhi, India, 2022, pp. 1-6, doi: 10.1109/DELCON54057.2022.9753639.

[2] Panda, A.R., Sirmour, S., Mallick, P.K. (2022). **Real-Time CPU Burst Time Prediction Approach for Processes in the Computational Grid Using ML**. In: Mohanty, M.N., Das, S. (eds) *Advances in Intelligent Computing and Communication. Lecture Notes in Networks and Systems*, vol 430. Springer, Singapore.

[3] Prakram, Parul, and Mohit Prakram. "**A Fuzzy Built Methodology towards Predicting the Next CPU-Burst For SJF Scheduling Algorithm**." *Think India Journal* 22, no. 17 (2019): 3851-3859.

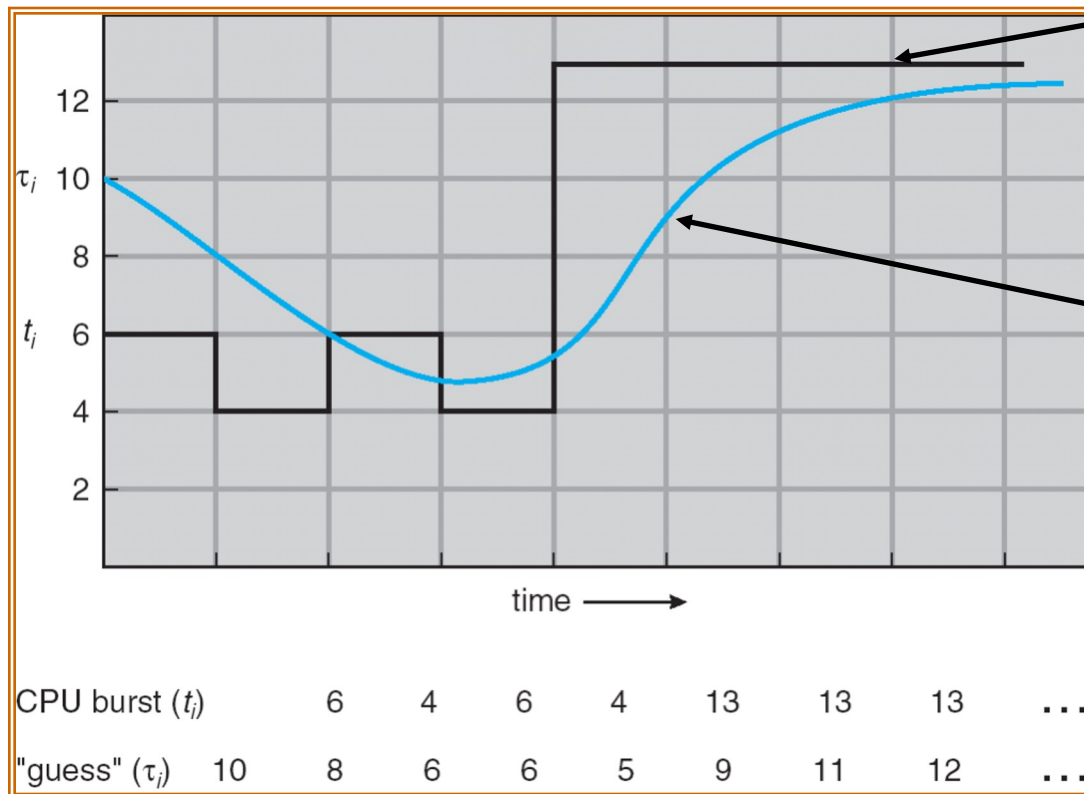


Nhận xét về giải thuật SJF (tt)

- Thời gian sử dụng CPU chính là độ dài của CPU burst
 - Trung bình tất cả các CPU burst đo được trong quá khứ
 - Nhưng thông thường những CPU burst càng mới càng phản ánh đúng hành vi của process trong tương lai
- Một kỹ thuật thường dùng là sử dụng trung bình hàm mũ (exponential averaging)
 - $\tau_{n+1} = \alpha t_n + (1 - \alpha)\tau_n, 0 \leq \alpha \leq 1$
 - $\tau_{n+1} = \alpha t_n + (1 - \alpha)\alpha t_{n-1} + \dots + (1 - \alpha)^j \alpha t_{n-j} + \dots + (1 - \alpha)^{n+1} \tau_0$
 - Nếu chọn $\alpha = \frac{1}{2}$ thì có nghĩa là trị đo được t_n và trị dự đoán τ_n được xem quan trọng như nhau.



Dự đoán thời gian sử dụng CPU



Độ dài CPU burst
đo được

Độ dài CPU burst
dự đoán, với
 $\alpha = \frac{1}{2}$ và $\tau_0 = 10$



Định thời theo độ ưu tiên - Priority Scheduling

- Mỗi tiến trình sẽ được gán một độ ưu tiên (thường biểu diễn bởi một con số).
- CPU sẽ được cấp cho tiến trình có độ ưu tiên cao nhất theo các giá trị số được gán (có thể theo thứ tự tăng dần hay giảm dần).
- Định thời sử dụng độ ưu tiên có thể:
 - Preemptive hoặc
 - Non-preemptive



Cách gán độ ưu tiên cho tiến trình

- SJF là một giải thuật định thời sử dụng độ ưu tiên được xác định dựa vào thời gian sử dụng CPU (giá trị được ước lượng)
- Ngoài ra, việc gán độ ưu tiên còn có thể dựa vào:
 - ☐ Yêu cầu về bộ nhớ
 - ☐ Số lượng file được mở
 - ☐ Tỷ lệ thời gian dùng cho I/O trên thời gian sử dụng CPU
 - ☐ Các yêu cầu bên ngoài ví dụ như: số tiền người dùng trả khi thực thi công việc



Hạn chế của định thời theo độ ưu tiên

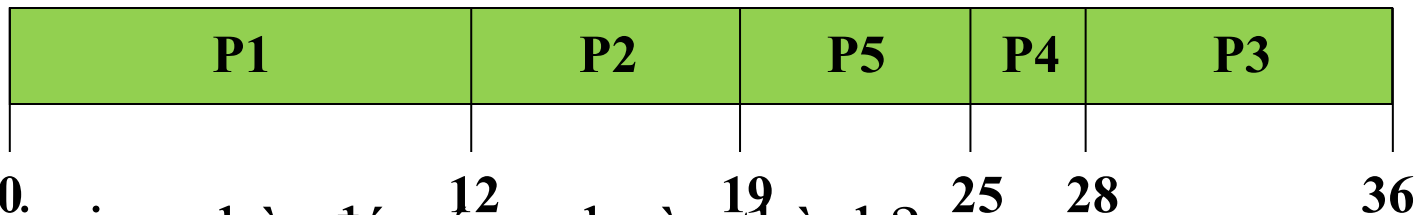
- Vấn đề trì hoãn vô hạn định: tiến trình có độ ưu tiên thấp có thể không bao giờ được thực thi (do có những tiến trình độ ưu tiên cao hơn liên tục xuất hiện).
- Giải pháp: làm mới (aging) – độ ưu tiên của tiến trình sẽ tăng theo thời gian



Ví dụ định thời theo độ ưu tiên

Process	Arrival Time	Burst Time	Priority
P1	0	12	2
P2	2	7	1
P3	5	8	5
P4	9	3	4
P5	12	6	3

■ Giải quyết bằng



■ Thời gian chờ, đáp ứng, hoàn thành?



Tóm tắt lại nội dung buổi học

- Các khái niệm cơ bản về định thời
- Các bộ định thời
- Các tiêu chuẩn định thời CPU
- Các giải thuật định thời
 - First-Come, First-Served (FCFS)
 - Shortest Job First (SJF)
 - Shortest Remaining Time First (SRTF)
 - Priority Scheduling



Bài tập 1

- Sử dụng các giải thuật FCFS, SJF, SRTF, Priority để tính các giá trị thời gian đợi, thời gian đáp ứng và thời gian hoàn thành trung bình

Process	Arrival Time	Burst Time	Priority
P1	0	4	3
P2	5	1	2
P3	3	8	1
P4	10	2	4
P5	8	7	3



Bài tập 2

- Sử dụng các giải thuật FCFS, SJF, SRTF, Priority để tính các giá trị thời gian đợi, thời gian đáp ứng và thời gian hoàn thành trung bình

Thread	Priority	Burst	Arrival
P_1	40	20	0
P_2	30	25	25
P_3	30	25	30
P_4	35	15	60
P_5	5	10	100
P_6	10	10	105



COMPUTER ENGINEERING



UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

THẢO LUẬN

