

Systemy operacyjne

Planowanie przydziału procesora

Dariusz Wawrzyniak



Systemy operacyjne



Plan wykładu

- Komponenty jądra związane z szeregowaniem
- Ogólna koncepcja planowania
- Kryteria oceny algorytmów planowania
- Algorytmy planowania

Planowanie przydziału procesora (2)

Systemy operacyjne



Komponenty jądra w planowaniu

- Planista krótkoterminowy (ang. CPU scheduler) — wyznacza wartość priorytetu procesów gotowych i wybiera proces (o najwyższym priorytecie) do wykonania.
- Ekspedytor (zwany również dyspozytorem) ang. dispatcher) — realizuje przekazanie sterowanie do procesu wybranego przez planistę (dokonuje przełączenia kontekstu).

Planowanie przydziału procesora (3)

Systemy operacyjne



Ogólna koncepcja planowania

- Tryb decyzji — określa okoliczności, w których oceniane i porównywane są priorytety procesów oraz dokonywany jest wybór procesu do wykonania.
- Funkcja priorytetu — funkcja wyznaczająca aktualny priorytet procesu na podstawie parametrów procesu i stanu systemu.
- Reguła arbitrażu — reguła rozstrzygania konfliktów w dostępie do procesora w przypadku procesów o tym samym priorytecie.

Planowanie przydziału procesora (4)

Systemy operacyjne

**Tryb decyzji**

- Schemat niewywłaszczeniowy (ang. nonpreemptive) — proces po uzyskaniu dostępu do procesora wykonywany jest do momentu zakończenia lub zgłoszenia żądania obsługi do systemu.
- Schemat wywłaszczeniowy (ang. preemptive) — proces może zostać zatrzymany i umieszczony w kolejce procesów gotowych, a procesor zostaje przydzielony procesowi o wyższym (lub równym) priorytecie.

Planowanie przydziału procesora (5)

Systemy operacyjne

**Podjęmowanie decyzji o wywłaszczeniu**

- Utworzenie i przyjęcie nowego procesu
- Obudzenie procesu w wyniku otrzymania komunikatu, sygnału gotowości urządzenia (przerwanie) lub sygnału wynikającego z synchronizacji
- Upłynięcie kwantu czasu odmierzanego przez czasomierz
- Wzrost priorytetu innego procesu w stanie gotowy powyżej priorytetu procesu wykonywanego — możliwe w systemie ze zmiennymi priorytetami

Planowanie przydziału procesora (6)

Systemy operacyjne

**Funkcja priorytetu**

- Argumentami funkcji priorytetu są wybrane składowe stanu procesu oraz stanu systemu.
- Priorytet procesu w danej chwili jest wartością wynikową funkcji priorytetu dla bieżących wartości parametrów stanu danego procesu i aktualnego stanu systemu.

Planowanie przydziału procesora (7)

Systemy operacyjne

**Argumenty funkcji priorytetu ⁽¹⁾**

- Czas oczekiwania — czas spędzony w kolejce procesów gotowych (czas spędzony w stanie gotowości)
- Czas obsługi — czas, przez który proces był wykonywany (wykorzystywał procesor) od momentu przyjęcia do systemu
- Rzeczywisty czas przebywania w systemie — czas spędzony w systemie od momentu przyjęcia (czas obsługi + czas oczekiwania + czas realizacji żądań zasobowych)
- Czasowa linia krytyczna — czas, po którym wartość wyników spada (nawet do zera, np. przy przewidywaniu pogody)

Planowanie przydziału procesora (8)

Systemy operacyjne

Argumenty funkcji priorytetu ⁽²⁾

- Priorytet zewnętrzny — składowa priorytetu, która pozwala wyróżnić procesy ze względu na klasy użytkowników lub rodzaj wykonywanych zadań
- Wymagania odnośnie wielkości przestrzeni adresowej pamięci
- Obciążenie systemu — liczba procesów przebywających w systemie i ubiegających się (potencjalnie) o przydział procesora lub innych zasobów, zajętość pamięci

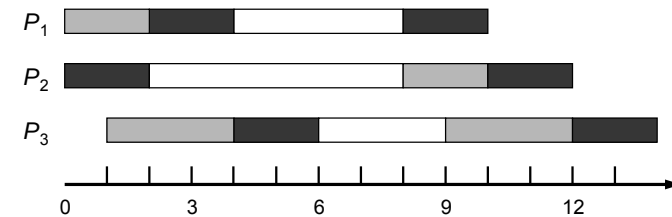
Planowanie przydziału procesora (9)

Systemy operacyjne



Przykład realizacji przetwarzania

Poniższy diagram przedstawia zmiany stanu 3 procesów w czasie i ma na celu zobrazowanie parametrów czasowych.



Planowanie przydziału procesora (10)

Systemy operacyjne



Reguła arbitrażu

- Losowo — możliwe w przypadku, gdy liczba procesów o tym samym priorytecie jest niewielka
- Cyklicznie — cykliczny przydział procesora kolejnym procesom
- Chronologicznie — w kolejności przyjmowania procesów do systemu (w kolejności FIFO)

Planowanie przydziału procesora (11)

Systemy operacyjne

Kryteria oceny algorytmów planowania ⁽¹⁾

- Efektywność z punktu widzenia systemu
 - wykorzystanie procesora (processor utilization) — procent czasu, przez który procesor jest zajęty pracą
 - przepustowość (throughput) — liczba procesów kończonych w jednostce czasu
- Inne aspekty z punktu widzenia systemu
 - sprawiedliwość (fairness) — równe traktowanie procesów
 - respektowanie zewnętrznych priorytetów procesów
 - równoważenie obciążenia wykorzystania zasobów

Planowanie przydziału procesora (12)

Systemy operacyjne

Kryteria oceny algorytmów planowania ⁽²⁾

- Efektywność z punktu widzenia użytkownika
 - czas cyklu przetwarzania (turnaround time) — czas pomiędzy przedłożeniem zadania, a zakończeniem jego wykonywania (rzeczywisty czas przebywania w systemie w momencie zakończenia procesu),
 - czas odpowiedzi (reakcji, response time) — czas pomiędzy przedłożeniem żądania, a rozpoczęciem przekazywania odpowiedzi.
 - czas opóźnienia — czas od linii krytycznej do momentu zakończenia wykonywania
- Inne aspekty z punktu widzenia użytkownika
 - przewidywalność — realizacja przetwarzania w zbliżonym czasie niezależnie od obciążenia systemu.

Planowanie przydziału procesora (13)

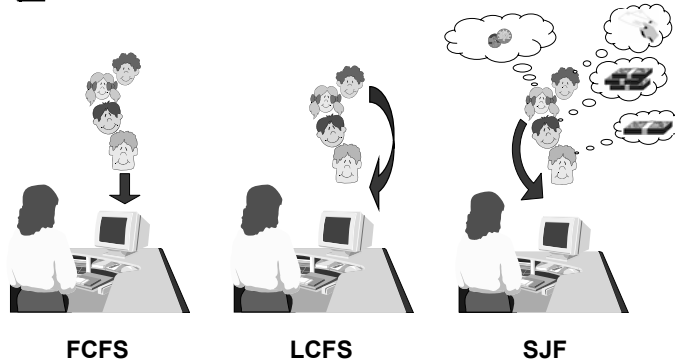
Systemy operacyjne

Algorytmy planowania niewywłaszczającego ⁽¹⁾

- FCFS (First Come First Served) — pierwszy zgłoszony, pierwszy obsługany
- LCFS (Last Come First Served) — ostatni zgłoszony, pierwszy obsługany
- SJF (SJN, SPF, SPN, Shortest Job/Process First/Next) — najpierw najkrótsze zadanie

Planowanie przydziału procesora (14)

Systemy operacyjne

Algorytmy planowania niewywłaszczającego ⁽²⁾

Planowanie przydziału procesora (15)

Systemy operacyjne

Algorytmy planowania wywłaszczającego ⁽¹⁾

- Planowanie rotacyjne (ang. Round Robin, RR) — po ustalonym kwancie czasu proces wykonywany jest przerywany i trafia do kolejki procesów gotowych.
- SRT (Shortest Remaining Time) — najpierw zadanie, które ma najkrótszy czas do zakończenia.

Planowanie przydziału procesora (16)

Systemy operacyjne

Algorytmy planowania wywłaszczającego⁽²⁾

RR **SRT**

Planowanie przydziału procesora (17)

Systemy operacyjne

Podstawowe algorytmy planowania a funkcja priorytetu

- Podstawowe algorytmy planowania można uzyskać przez odpowiednią definicję funkcji priorytetu.
- Parametrami funkcji priorytetu dla podstawowych algorytmów planowania są następujące atrybuty czasowe procesów:
 - a — bieżący (dotychczasowy) czas obsługi
 - r — rzeczywisty czas w systemie
 - t — całkowity wymagany czas obsługi (czas obsługi do momentu zakończenia)

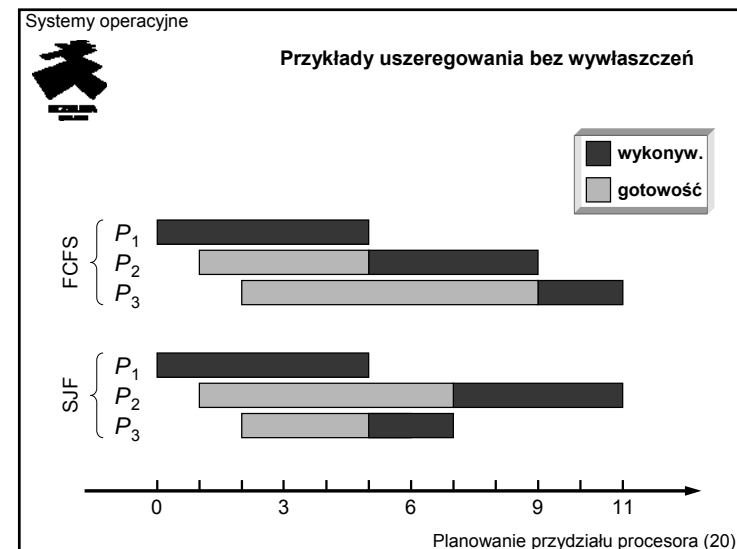
Planowanie przydziału procesora (18)

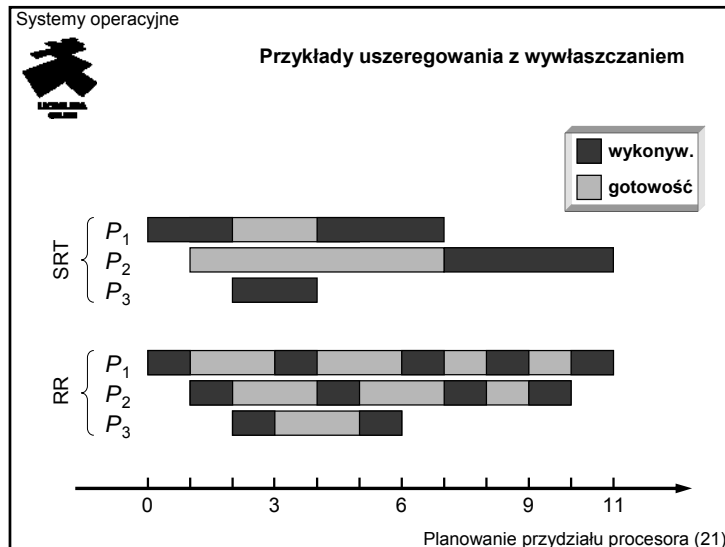
Systemy operacyjne

Własności algorytmów planowania

algorytm	priorytet	tryb decyzji	arbitraż
FIFO	r	niewywłaszczeniowy	losowy
LIFO	$-r$	niewywłaszczeniowy	losowy
SJF	$-t$	niewywłaszczeniowy	losowy lub chronologiczny
SRT	$a - t$	wywłaszczeniowy	losowy lub chronologiczny
RR	stały	wywłaszczeniowy	cykliczny

Planowanie przydziału procesora (19)





Systemy operacyjne

Algorytmy SJF/SRT — estymacja czasu obsługi

- Średnia arytmetyczna

$$S_{n+1} = \frac{1}{n} \cdot \sum_{i=1}^n T_i \Leftrightarrow S_{n+1} = \frac{T_n + (n-1) \cdot S_n}{n}$$
- Średnia wykładnicza

$$S_{n+1} = \alpha \cdot T_n + (1-\alpha) \cdot S_n$$

$$\Downarrow$$

$$S_{n+1} = \sum_{i=0}^{n-1} (1-\alpha)^i \cdot \alpha \cdot T_{n-i} =$$

$$= \alpha T_n + (1-\alpha) \alpha T_{n-1} + \dots + (1-\alpha)^{n-1} \alpha T_1$$

Planowanie przydziału procesora (22)

Systemy operacyjne

Algorytm RR — dobór kwantu czasu

- Krótki kwant czasu oznacza zmniejszenie czasu cyklu przetwarzania procesów krótkich, ale zwiększa narzut czasowy związany z przełączaniem kontekstu.
- Z punktu widzenia interakcji z użytkownikiem kwant czasu powinien być trochę większy, niż czas odpowiedzi (reakcji).

Planowanie przydziału procesora (23)

Systemy operacyjne

Dobór kwantu czasu, a czas odpowiedzi systemu

Legend:
 ■ wykonywanie
 □ oczekiwanie
 ■ gotowość

kwant czasu > czas interakcji

czas odpowiedzi

kwant czasu

kwant czasu < czas interakcji

kwant czasu

czas odpowiedzi

Planowanie przydziału procesora (24)

Systemy operacyjne

Inne algorytmy planowania ⁽¹⁾

- Planowanie priorytetowe — oparte na priorytecie zewnętrznym
- Planowanie wielokolejkowe — w systemie jest wiele kolejek procesów gotowych i każda z kolejek może być inaczej obsługiwana.
- Planowanie przed liniami krytycznymi — zakończenie zadania przed czasową linią krytyczną lub możliwie krótko po tej linii

Planowanie przydziału procesora (25)

Systemy operacyjne

Inne algorytmy planowania ⁽²⁾

priorytetowe

wielokolejkowe

przed liniami krytycznymi

Planowanie przydziału procesora (26)

Systemy operacyjne



Szeregowanie procesów, ograniczonych wejściem-wyjściem

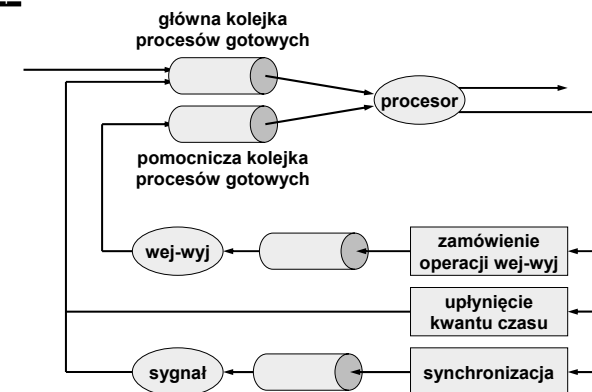
- Procesy ograniczone wejściem-wyjściem potrzebują niewiele czasu procesora, większość czasu w systemie spędzając na oczekiwaniu na urządzenia zewnętrzne.
- Opóźnianie przydziału procesora dla tego typu procesów powoduje zmniejszenie wykorzystania urządzeń zewnętrznych, a przydział — ze względu na nie długą fazę procesora — nie powoduje istotnego zwiększenia czasu oczekiwania innych procesów.
- Właściwym algorytmem byłby SJF lub SRT.
- Bezwzględna preferencja dla procesów oczekujących na gotowość urządzeń może spowodować głodzenie procesów ograniczonych procesorem.

Planowanie przydziału procesora (27)

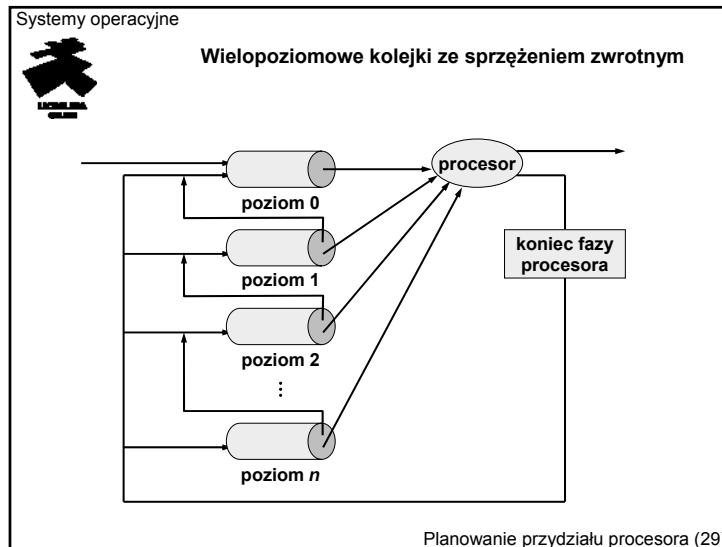
Systemy operacyjne



Wirtualne planowanie rotacyjne (VRR)



Planowanie przydziału procesora (28)



Systemy operacyjne

Implementacja algorytmów planowania

- Z punktu widzenia przetwarzania użytkowego przełączanie kontekstu jest marnotrawstwem czasu procesora.
- Decyzja planisty musi zapaść w możliwie krótkim czasie.
- Struktury danych muszą być tak zaprojektowane, żeby ułatwić dokonanie szybkiego wyboru procesu o najwyższym priorytecie zgodnie z polityką planowania przydziału procesora (modelem matematycznym).

Planowanie przydziału procesora (30)

Systemy operacyjne

Implementacja algorytmu FCFS

- Struktura danych dla kolejki procesów gotowych → kolejka FIFO
- Umieszczenie procesu w kolejce procesów gotowych → dopisanie procesu na końcu kolejki FIFO
- Wybór procesu do wykonania → pobranie procesu z czoła kolejki FIFO
- Czy taki algorytm realizuje dokładnie założenia modelu matematycznego?

Planowanie przydziału procesora (31)

Systemy operacyjne

Kolejki priorytetowe

- Kolejka priorytetowa jest wielopoziomową kolejką ze sprzężeniem zwrotnym, w której każdy poziom odpowiada pewnej wartości priorytetu lub pewnemu zakresowi wartości.
- Umieszczenie procesu w kolejce priorytetowej sprowadza się do wyznaczenia pozycji odpowiedniej dla priorytetu procesu, a następnie umieszczeniu procesu na końcu kolejki na tej pozycji.
- Wybór procesu o najwyższym priorytecie sprowadza się do zlokalizowania pierwszej niepustej w kolejności malejących priorytetów i wybrania pierwszego procesu z tej kolejki.

Planowanie przydziału procesora (32)