

**FACULDADE DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA
LICENCIATURA EM ENGENHARIA INFORMÁTICA
SISTEMAS OPERATIVOS E PROGRAMAÇÃO CONCORRENTE**

Aula: Gestão do Processador

Grupo Docente:

- Eng^o. Délcio Chadreca (MsC)
- Dr. Alfredo Covele (MsC)

Tópicos da Aula

- ▶ Gestão do Processador
- ▶ Escaladores
- ▶ Algoritmos de escalonamento

Escalonador

Um dos componentes mais importantes da gestão de tarefas é o escalonador de tarefas (task scheduler), que decide a ordem de execução das tarefas prontas.

O algoritmo utilizado no escalonador define o comportamento do sistema operacional, permitindo obter sistemas que tratem de forma mais eficiente e rápida as tarefas a executar, que podem ter características diversas: aplicações interativas, processamento de grandes volumes de dados, programas de cálculo numérico, etc.

Escalonador

- O escalonador é a entidade do SO responsável por selecionar um processo apto para executar no processador
- O objetivo é dividir o tempo do processador de forma justa
- Típicos de sistemas multiprogramados
- Duas partes:
 - Escalonador: política de seleção
 - *Dispatcher*: efetua a troca de contexto

Objetivos do Escalonador

Maximizar a utilização do processador

Maximizar a produção do sistema

- N° de processos executados por unidade de tempo

Minimizar o tempo de execução

- Tempo total para executar um processo

Minimizar o tempo de espera

- Tempo em que o processo permanece na fila de aptos

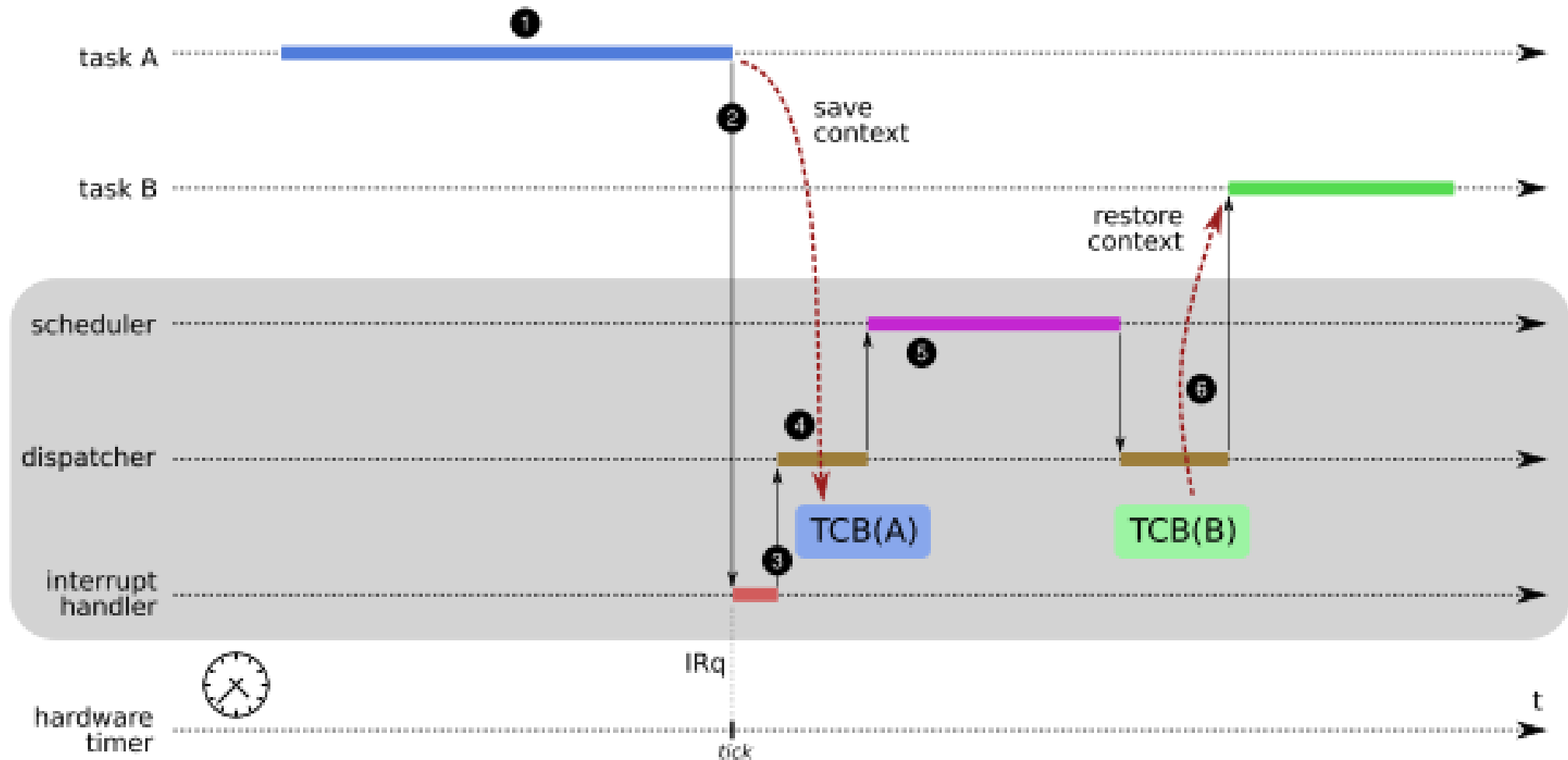
Minimizar o tempo de resposta

- Tempo decorrido entre a requisição e sua realização

Troca de Contexto

- Para que o sistema operacional possa suspender e retomar a execução de tarefas de forma transparente (sem que as tarefas o percebam), é necessário definir operações para salvar o contexto actual de uma tarefa em seu PCB e restaurá-lo mais tarde no processador. Por essa razão, o acto de suspender uma tarefa e reativar outra é denominado uma troca de contexto.
- A implementação da troca de contexto é uma operação delicada, envolvendo a manipulação de registradores e flags específicos de cada processador, sendo por essa razão geralmente é codificada em linguagem de máquina.

Diagrama temporal em uma troca de contexto



Descrição

1. Uma tarefa A está executando;
2. Ocorre uma interrupção do temporizador do hardware e a execução desvia para a rotina de tratamento, no núcleo;
3. A rotina de tratamento ativa o despachante;
4. O despachante salva o estado da tarefa A em seu PCB e atualiza suas informações de gestão;
5. Opcionalmente, o despachante consulta o escalonador para escolher a próxima tarefa a ativar (B);
6. O despachante resgata o estado da tarefa B de seu PCB e a reativa.

Quando Escalonar?

- Quando se cria um novo processo, é necessário tomar uma decisão entre executar o processo pai ou o processo filho
- Quando se termina um processo, algum outro processo deve ser escolhido entre processos prontos
- Quando um processo bloqueia para E/S
- Quando ocorre uma interrupção de E/S

Algoritmos de Escalonamento

Existem duas categorias de algoritmos de escalonamento:

- **Não preemptivos** – escolhe um processo para executar e então o deixa executar até que seja bloqueado ou até que ele voluntariamente libere a CPU
- **Preemptivos** – escolhe um processo e o deixa em execução por um tempo máximo fixado

Tipos de tarefas

- Sistemas de tempo real
- Sistema interativos
- Sistemas em lote (batch)

Além dessa das classificações acima , as tarefas também podem ser classificadas de acordo com seu comportamento no uso do processador:

- Sistemas Orientados a processamento
- Sistemas Orientados a entrada/saída

Tarefas de tempo real

- Exigem previsibilidade em seus tempos de resposta aos eventos externos, pois geralmente estão associadas ao controle de sistemas críticos, como processos industriais, tratamento de fluxos multimídia, etc.
- Cumprimento dos prazos – evitar a perda de dados
- Previsibilidade – evitar a degradação da qualidade em sistemas multimídia

Tarefas interativos

- São tarefas que recebem eventos externos (do usuário ou através da rede) e devem respondê-los rapidamente, embora sem os requisitos de previsibilidade das tarefas de tempo real.
- Esta classe de tarefas inclui a maior parte das aplicações dos sistemas desktop (editores de texto, navegadores Internet, jogos) e dos servidores de rede (e-mail, web, bancos de dados).
- Tempo de resposta – responder rapidamente às requisições. Também indica quanto tempo, em média, o usuário tem de esperar pelo fim de um trabalho
- Proporcionalidade – satisfazer às expectativas dos usuários

Tarefas em lote (batch)

- Tarefas em lote (batch): são tarefas sem requisitos temporais explícitos, que normalmente executam sem intervenção do usuário, como procedimentos de backup, varreduras de antivírus, cálculos numéricos longos, tratamentos de grandes massas de dados em lote, renderização de animações, etc.
- Vazão (*throughput*) – maximizar o n° de processos por hora ou é o número de processos por hora que o sistema termina
- Tempo de retorno – minimizar o tempo entre a submissão e o término
- Utilização de CPU – manter a CPU ocupada o tempo todo

Cont.

Tarefas orientados a processamento:

São tarefas que usam intensivamente o processador na maior parte de sua existência. Essas tarefas passam a maior parte do tempo nos estados pronta ou executando.

Tarefas orientados a entrada/saída:

São tarefas que dependem muito mais dos dispositivos de entrada/saída que do processador. Essas tarefas ficam boa parte de suas existências no estado suspenso, aguardando respostas às suas solicitações de leitura e/ou escrita de dados nos dispositivos de entrada/saída.

Exemplos desta classe de tarefas incluem editores, compiladores e servidores de rede.

Objetivos e Metricas de escalonamento

Ao se definir um algoritmo de escalonamento, deve-se ter em mente seu objetivo. Todavia, os objetivos do escalonador são muitas vezes contraditórios; o desenvolvedor do sistema tem de escolher o que priorizar, em função do perfil das aplicações a suportar.

Por exemplo, um sistema interativo voltado à execução de jogos exige valores de quantum baixos, para que cada tarefa pronta receba rapidamente o processador (provendo maior interatividade).

Todavia, valores pequenos de quantum implicam em uma menor eficiência E no uso do processador, conforme visto na Seção

Tempo de execução (ou de vida) (turnaround time)

Diz respeito ao tempo total da execução de uma tarefa, ou seja, o tempo decorrido entre a criação da tarefa e seu encerramento, computando todos os tempos de processamento e de espera.

É uma medida típica de sistemas em lote, nos quais não há interação direta com os usuários do sistema.

Não deve ser confundido com o tempo de processamento que é o tempo total de uso de processador demandado pela tarefa.

Tempo de espera (waiting time)

É o tempo total perdido pela tarefa na fila de tarefas prontas, aguardando o processador.

Deve-se observar que esse tempo não inclui os tempos de espera em operações de entrada/saída (que são inerentes à aplicação e aos dispositivos).

Tempo de resposta (response time)

- É o tempo decorrido entre a chegada de um evento ao sistema e o resultado imediato de seu processamento. Por exemplo, em um editor de textos seria o tempo decorrido entre apertar uma tecla e o caractere correspondente aparecer na tela.
- Essa medida de desempenho é típica de sistemas interativos, como sistemas desktop e de tempo real; ela depende sobretudo da rapidez no tratamento das interrupções de hardware pelo núcleo e do valor do quantum de tempo, para permitir que as tarefas interativas cheguem mais rápido ao processador quando saem do estado suspenso.

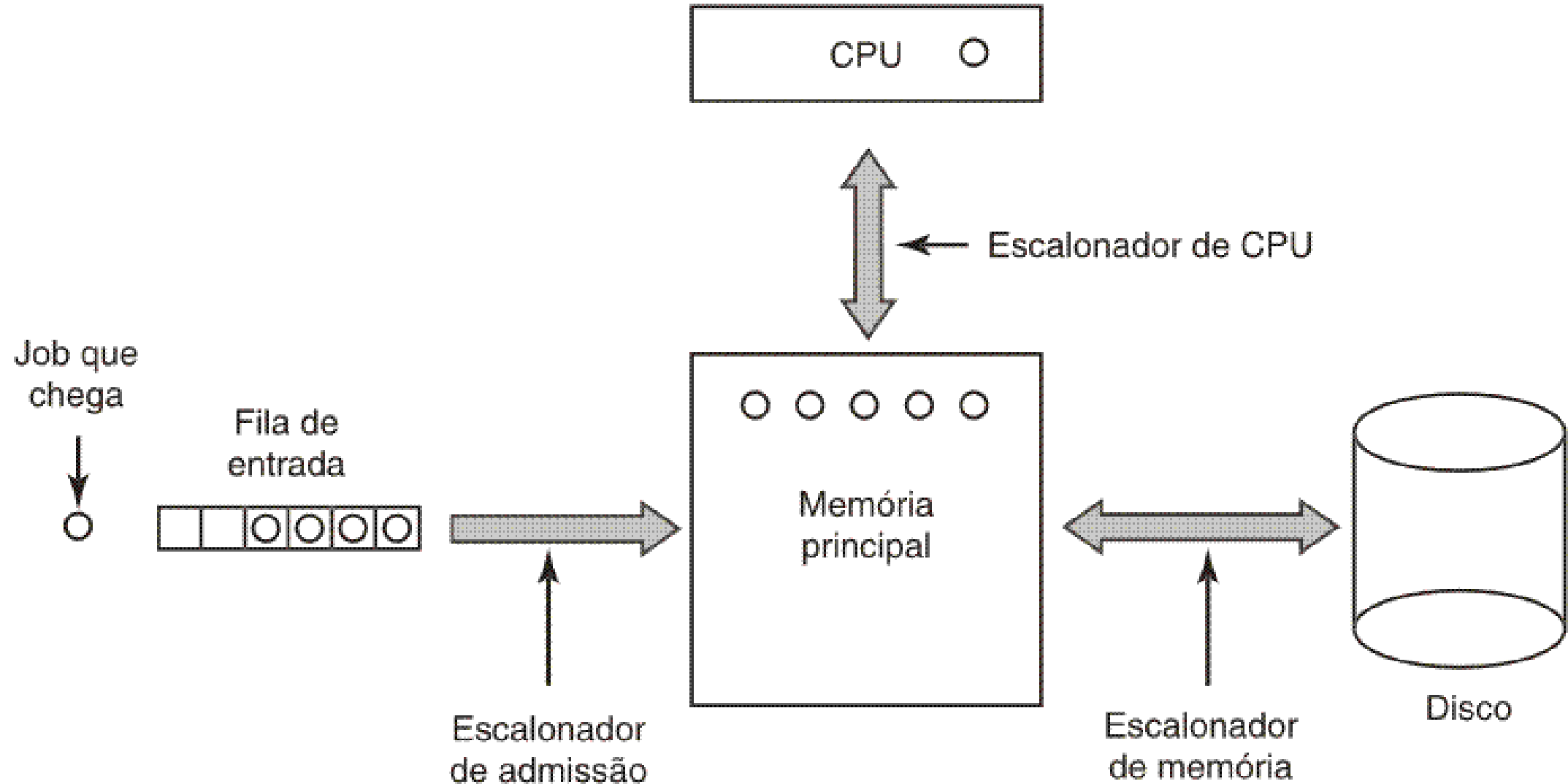
Justiça

Este critério diz respeito à distribuição do processador entre as tarefas duas tarefas de comportamento e prioridade similares devem ter durações execução similares.

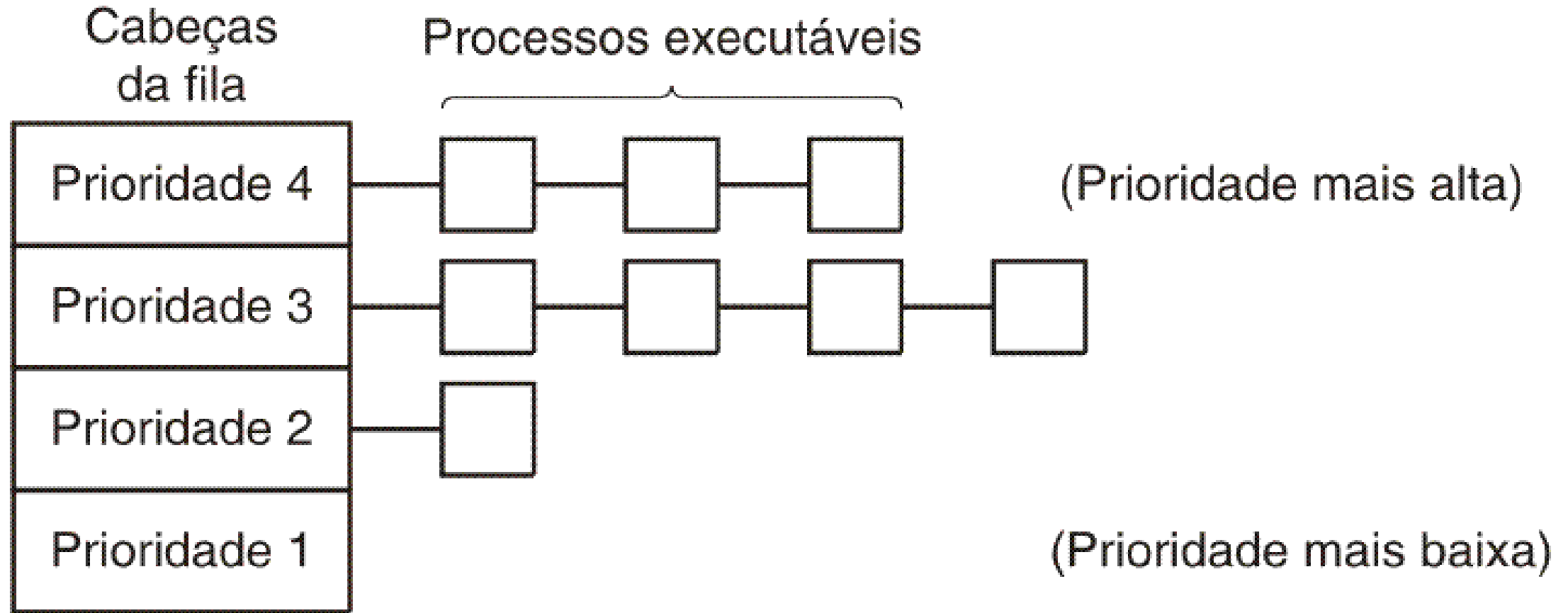
Eficiência

- Indica o grau de utilização do processador na execução das tarefas do usuário. Ela depende sobretudo da rapidez da troca de contexto e da quantidade de tarefas orientadas a entrada/saída no sistema (tarefas desse tipo geralmente abandonam o processador antes do fim do quantum, gerando assim mais trocas de contexto que as tarefas orientadas a processamento).

Escalonamento em Sistemas em Lote



Escalonamento em Sistemas Interativos



Níveis de Escalonamento

- **Longo Prazo:** pool de processos que são alocados na memória (determina quais processos no disco voltaram o estado pronto na memória principal)
- **Médio Prazo:** sistemas de tempo real, remove processos da memória, reduz o grau de multiprogramação
- **Curto Prazo:** responsável pela mudança de estados de processos que estão na memória principal (pronto, execução e espera)

Escalonador Longo Prazo

- Executado quando um novo processo é criado;
- Determina quando um processo novo passa a ser considerado no sistema;
- Controla o grau de multiprogramação do sistema
 - Quanto maior o n° de processos ativos, menor a percentagem de tempo de uso do processador por processo

Escalonador Médio Prazo

Associado a gerência de memória

- Participa do mecanismo de *swapping*

Suporte adicional a multiprogramação

- Grau de multiprogramação efetiva (diferencia aptos dos aptos suspensos)

Escalonador Curto Prazo

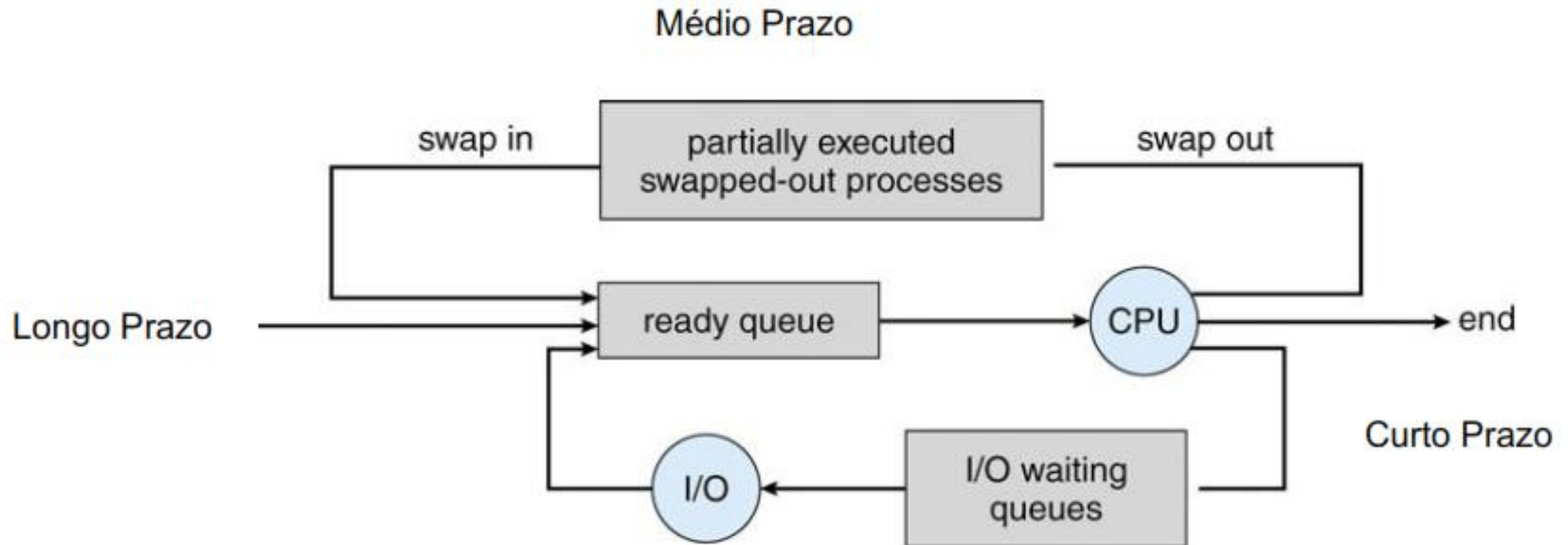
Mais importante;

Determina qual processo apto deverá utilizar o processador;

Executado sempre que ocorre eventos importantes

- Interrupção de relógio
- Interrupção de E\S
- Chamadas de sistemas
- Sinais (interrupção de software)

Diagrama de Escalonamento



Algoritmos de escalonamento

- FCFS (First Come, First Served)
- Shortest Job First.
- Shortest Remaining Time First.
- Por prioridade, sem preempção.
- Por prioridade, com preempção por prioridade.
- Round-Robin com quantum

OBRIGADO !!!