

Sistemas Operacionais

Processos e Threads

Prof. Marcos Monteiro, MBA

<http://www.marcosmonteiro.com.br>

contato@marcosmonteiro.com.br

Estrutura de um Sistema Operacional

Gerente de Processos

Gerente de Memória

Gerente de Arquivos

Gerente de I/O

Sistema de Proteção

Interpretador de Comandos

GERÊNCIA DE PROCESSOS

- Um processo é um programa em execução. Sua constituição é:
 - 1. Código executável;
 - 2. Pilha de execução;
 - 3. Contador de programa;
 - 4. Ponteiro de pilha;
 - 5. Valores dos registradores;
- O Sistema Operacional tem como principais funções:
 - Criar e excluir os processos do usuário e do sistema;
 - Suspender e retornar processos;
 - Fornecer mecanismos para sincronizar os processos;
 - Fornecer mecanismos para comunicação de processos;
- As informações de cada processo são armazenadas na Tabela de Processos.
- Um processo pode criar outros processos. Dessa maneira, um processo que gera outro processo é chamado processo-pai. Já o processo que foi gerado a partir de outro processo é chamado processo-filho.

ESTADOS DE UM PROCESSO

- Novo – o processo está sendo criado;
- Em execução (Run) – as instruções estão sendo executadas;
- Em espera (Blocked) – o processo está bloqueado esperando a ocorrência de um evento para poder voltar a ser executado. Muitas vezes ele fica esperando o término de uma operação de I/O ou um sinal de algum processo;
- Pronto (Ready) – o processo está aguardando ser selecionado para uso do processador;
- Encerrado – o programa terminou sua execução.

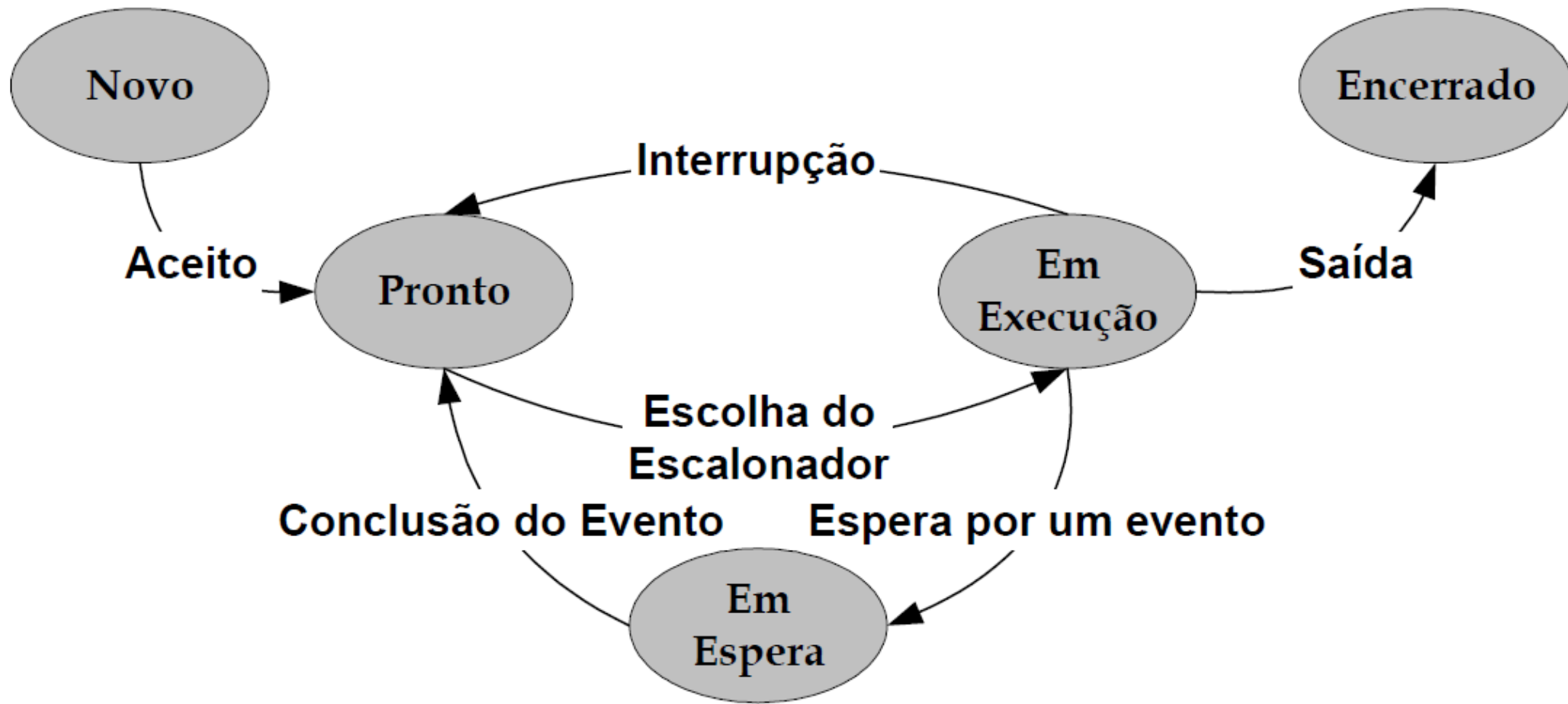


Figura 4-2 - Diagrama de Transição de Estados de um Processo

Espera Ocupada x Bloqueio

Na espera ocupada, o processo está constantemente gastando CPU (testando uma determinada condição).

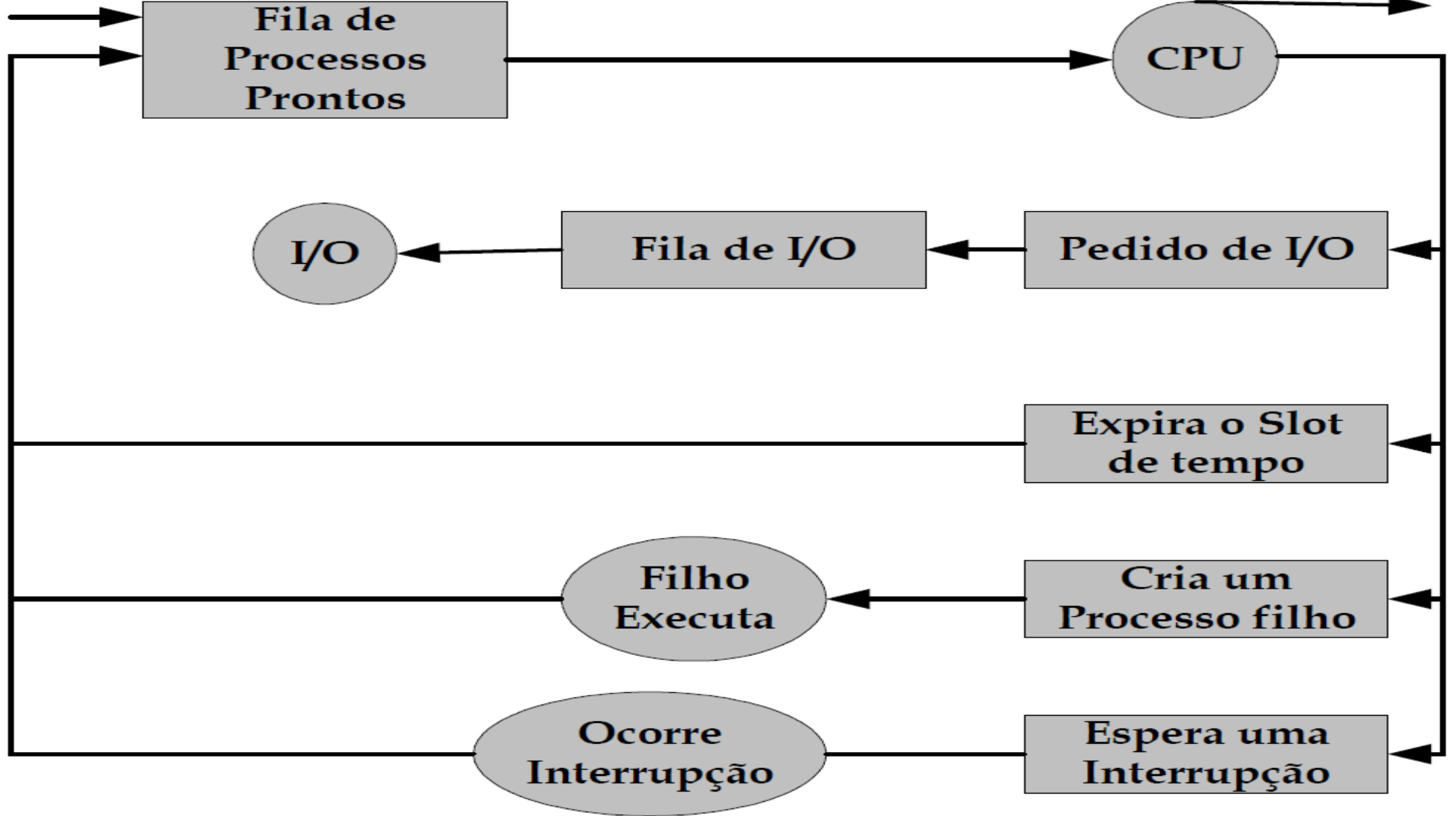
No bloqueio, o processo desiste de usar a CPU e é acordado quando a condição desejada se torna verdadeira.

BLOCO DE CONTROLE DE PROCESSO

- Cada processo é representado no sistema operacional por um bloco de controle de processo (PCB – Process Control Block).
- Estado do Processo – informa o estado atual do processo;
- Contador de Programa – indica o endereço da próxima instrução a ser executada para este processo;
- Registradores de CPU – informa os registradores a serem utilizados pelo sistema. A quantidade e categorias de registradores variam de acordo com a arquitetura do computador em uso. As informações contidas nos registradores devem ser salvas quando ocorrer uma interrupção;
- Informações de Escalonamento de CPU – indica a prioridade do processo, ponteiros para filas de escalonamento e demais informações do algoritmo de escalonamento;

ESCALONAMENTO DE PROCESSOS

- O objetivo da multiprogramação é ter processos em execução o tempo todo, visando maximizar o uso do processador. Para isto o tempo de uso do processador é compartilhado entre todos os processos em execução. A viabilidade está no fato de que nenhum processo usa os mesmos recursos ao mesmo tempo, ou mesmo que algumas operações de I/O demoradas deixam o processador ocioso.



Prontos e de Dispositivos. Os círculos são os recursos e as elipses eventos. As setas representam os fluxos. Todo novo processo é alocado na Fila de Processos Prontos, onde aguarda ser selecionado para execução. Podem ocorrer os seguintes eventos:

- O processo emite uma requisição de uso de um dispositivo de I/O e é alocado na fila do mesmo;
- O processo pode criar um outro filho, chamado processo filho, e espera o seu término;
- O processo pode ser removido a força da CPU como resultado de uma interrupção e ser

alocado na fila de processos prontos;

ESCALONADORES

- Nos sistemas multitarefa, vários processos são submetidos à execução.
- Como existe apenas um único processador, aqueles que estão aptos a serem executados imediatamente são colocados em um spool (espaço de armazenagem em disco), onde são mantidos até serem executados. O escalonador de longo prazo seleciona alguns processos para ocuparem a memória. O escalonador de curto prazo seleciona um dos processos da memória para utilizar o processador, por isso ele também é chamado de escalonador de CPU.

- A diferença entre esses escalonadores está na frequência de execução.
- Enquanto o segundo é executado com altíssima frequência devido ao curto de período de tempo de uso de um processo por vez, o primeiro é executado sem muita frequência. É tarefa do escalonador de longo prazo formar o banco de processos a serem escalonados. Conseqüentemente, é importante que ele faça uma cuidadosa seleção dos processos. Para fazer esta seleção, foi feita uma nova divisão de categoria de processos:
 - **Processos Limitados por I/O (I/O Bound)**– são processos que predominantemente fazem operações de I/O durante a sua execução;
 - **Processos Limitados por CPU (CPU Bound)** – predominantemente, eles fazem operações com a CPU.

- É altamente recomendado que o escalonador de longo prazo faça uma combinação com processos de ambas categorias. Esta recomendação tem por objetivo maximizar o uso dos recursos computacionais. A título de exemplificação, toma-se a hipótese de que há uma grande quantidade de processos limitados por I/O na memória. Existe uma grande probabilidade de não haver processos na fila de processos prontos para execução, conseqüentemente o processador estará sendo subutilizado.
- Por outro lado, se houver muitos processos limitados por CPU, a probabilidade dos recursos ficarem ociosos é grande. Além disso, pode haver processos que poderiam utilizá-los e não o fazem por não estarem na memória.

PROCESSOS COOPERATIVOS

- Os processos concorrentes que estão executando no sistema operacional podem ser concorrentes ou cooperativos. Um processo é independente se não afeta a execução de outros processos. Basicamente, qualquer processo que não compartilha dados ou recursos é independente. Caso um processo afete outro, então ele é cooperativo. Existem várias razões para construir um ambiente favorável aos processos cooperativos:
 - **Compartilhamento de Informações** – vários usuários podem estar interessados na mesma informação. É preciso fornecer mecanismos para que possam utilizá-la sem causar danos uns aos outros ou à própria informação;
 - **Velocidade de Computação** – dividir uma tarefa em partes menores pode fazer com que ela seja executada mais rápida. É importante que cada parte não utilize os mesmos recursos;
 - **Modularidade** – a construção do sistema em forma modular facilita a manutenção do mesmo e até a gerência de memória e processos do sistema operacional;

- As decisões de escalonamento de CPU podem ocorrer de quatro maneiras:
 - 1. Quando o processo passa do estado de execução para o estado de espera (bloqueado);
 - 2. Quando um processo passa do estado de execução para o estado pronto;
 - 3. Quando um processo passa do estado bloqueado para o estado pronto;
 - 4. Quando um processo termina.

DISPATCHER

- O dispatcher é o processo que fornece o controle da CPU para o processo selecionado pelo escalonador. Ele deve fazer a mudança de contexto, ou seja, verificar e salvar quais os valores de variáveis, registradores, posição no programa e recursos do processo que está deixando a CPU e acionar o contexto exigido pelo novo processo. O tempo de mudança de contexto é chamado de latência de dispatcher.

processos independentes, subprocessos e threads

Nos *processos independentes* não existe vínculo entre o processo criado e o seu criador. Cada processo possui seu próprio contexto de hardware, contexto de software e espaço de endereçamento.

Já os *subprocessos* são criados dentro de uma hierarquia, onde existe uma dependência entre o processo criador e o o subprocesso. Caso o processo pai deixe de existir o processo filho deixará também, apesar de cada processo possuir sua própria PCB.

Diferentemente dos dois exemplos anteriores, os *threads* compartilham o espaço de endereçamento e o contexto de software, porém cada thread possui seu próprio contexto de hardware.

RESUMO

Tempo de processador é o tempo que um processo leva no estado de execução durante o seu processamento.

Tempo de espera é o tempo total que um processo permanece na fila de pronto durante o seu processamento, aguardando ser executado.

Tempo de turnaround é o tempo que um processo leva desde sua criação até o seu término.

Tempo de resposta é o tempo decorrido entre uma requisição ao sistema ou à aplicação e o instante em que a resposta é exibida.

O escalonador seleciona processos que estejam no estado de pronto e os coloca em execução, obedecendo os critérios estabelecidos (tempo, prioridade)

Algoritmo de Escalonamento

FIRST-COME, FIRST-SERVED (PRIMEIRO A CHEGAR, PRIMEIRO SERVIDO) -FIFO

JOB MAIS CURTO PRIMEIRO (Caso haja dois curtos iguais=FIFO)

ESCALONAMENTO POR PRIORIDADE

prioridade pode ser definida de forma interna ou externamente. Para definir internamente são necessários parâmetros para análise como, por exemplo:

limite de tempo, requisitos de memória, quantidade de arquivos abertos pelo processo e o surto médio de CPU do processo.

ROUD ROBIN

Um quantum pode ter duração de 10 a 100 milisegundos, novamente depende da implementação do sistema operacional.











