

Sistemas Operacionais

Gerência de Arquivos

Prof. José Paulo G. de Oliveira
Eng. da Computação, UPE
jpgo@ecomp.poli.br

Conteúdo

- Motivação
- Definição
- Objetivos
- Conceitos fundamentais
- Estrutura do sistema de arquivos
- Alocação de arquivos

Conteúdo

- **Motivação**
- Definição
- Objetivos
- Conceitos fundamentais
- Estrutura do sistema de arquivos
- Alocação de arquivos

Arquivos para quê?

- Necessidade de armazenar informações para uso posterior
 - Programas e dados

Arquivos para quê?

- Necessidade de armazenar informações para uso posterior
 - Programas e dados
- Função do computador: recuperar, processar e apresentar informações previamente armazenadas
 - documentos, fotografias, músicas e vídeos

Arquivos para quê?

- Necessidade de armazenar informações para uso posterior
 - Programas e dados
- Função do computador: recuperar, processar e apresentar informações previamente armazenadas
 - documentos, fotografias, músicas e vídeos
- Para simplificar/padronizar o armazenamento e busca de informações
 - Conceito de ARQUIVO

Conteúdo

- Motivação
- **Definição**
- Objetivos
- Conceitos fundamentais
- Estrutura do sistema de arquivos
- Alocação de arquivos

O conceito de arquivo

Um **arquivo** é basicamente um conjunto de **informações armazenadas** em um dispositivo físico **não-volátil**, com um **nome** ou outra **referência** que permita sua **localização** posterior.

Texto, imagem, áudio, código executável, etc.

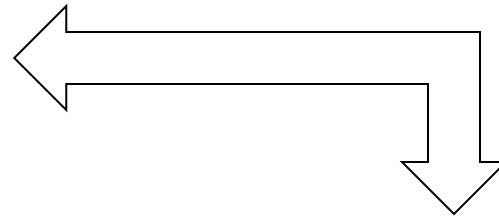
Conteúdo

- Motivação
- Definição
- **Objetivos**
- Conceitos fundamentais
- Estrutura do sistema de arquivos
- Alocação de arquivos



Gerência de Arquivos para quê?

Gerência de Arquivos para quê?

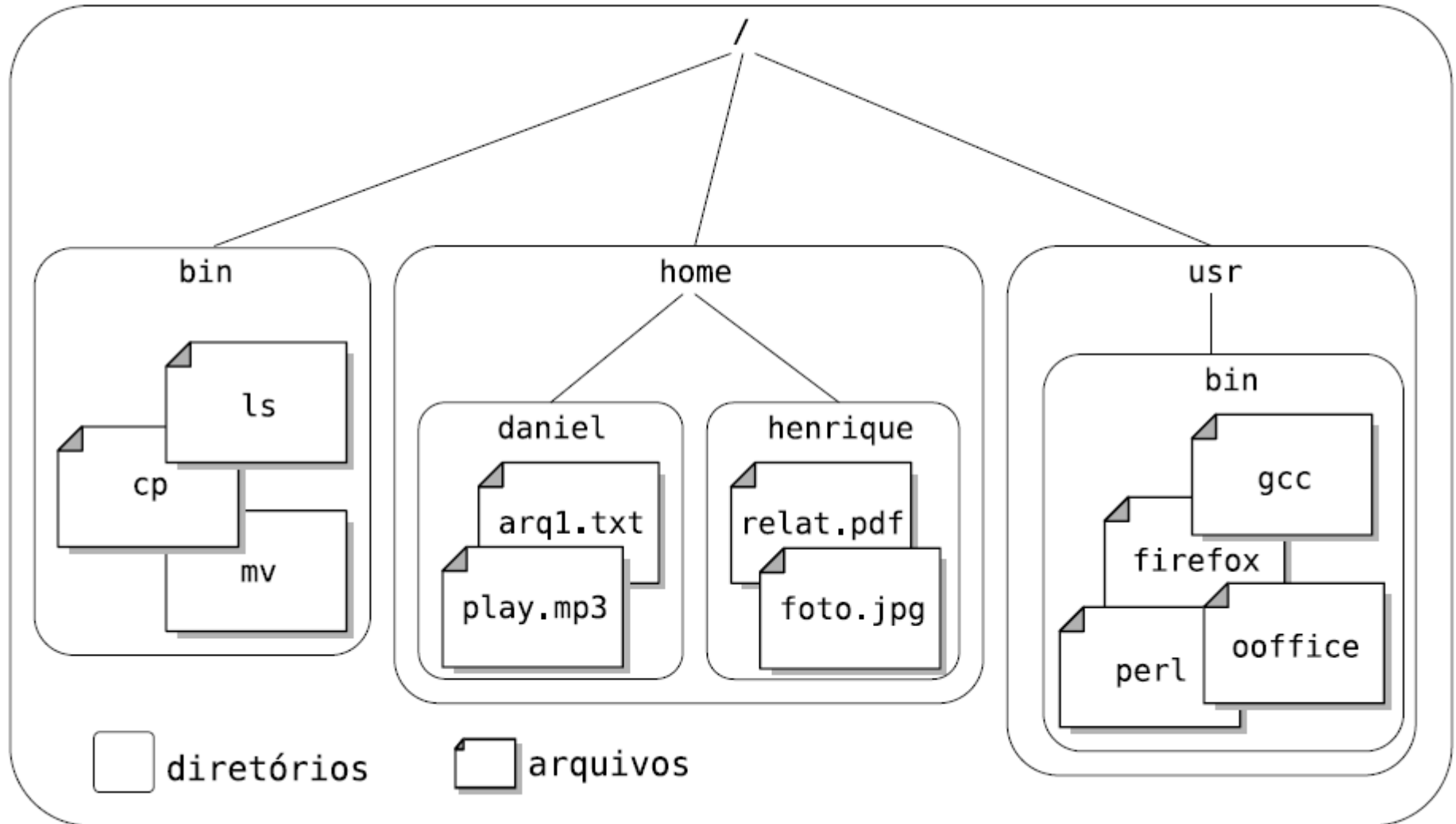


[illegible]

Gerência de Arquivos para quê?



Organização de arquivos



Organização de arquivos

- Um **dispositivo de armazenamento** pode conter **milhões** de arquivos
 - Arquivos são organizados em estruturas **hierárquicas** denominadas ***diretórios***

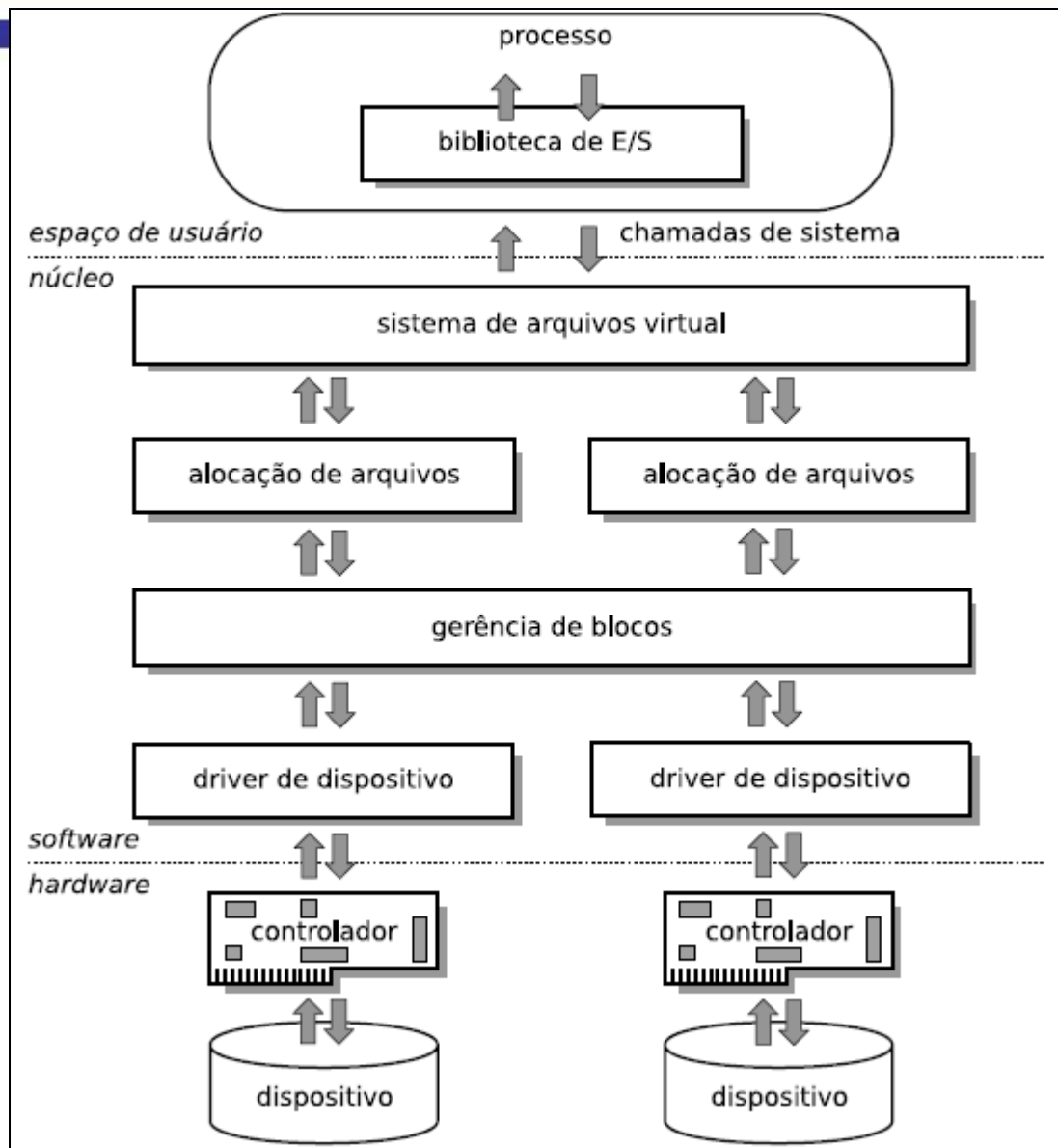
Organização de arquivos

- Um **dispositivo de armazenamento** pode conter **milhões** de arquivos
 - Arquivos são organizados em estruturas **hierárquicas** denominadas ***diretórios***
- A organização **física e lógica** dos arquivos e diretórios dentro de um dispositivo é definida pelo ***sistema de arquivos***

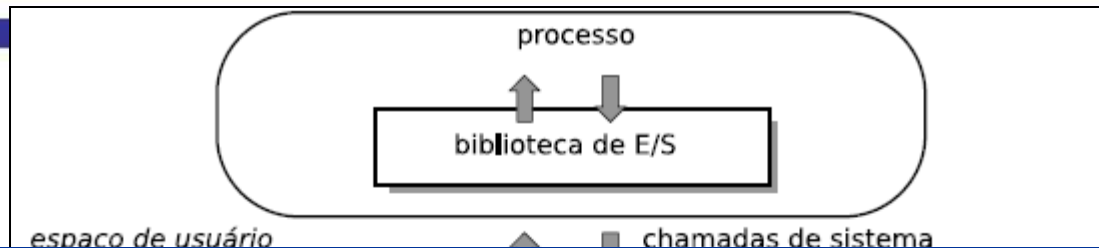
Gerência de Arquivos para que mais?

- Definir estrutura dos arquivos
 - Formato
 - Tipo
- Modos de acesso
- Controle de acesso
 - Segurança
 - Integridade da informação
- Interface homogênea para processos
 - Acesso homogêneo ao dispositivo (HD, fita magnética, disco óptico, flash)

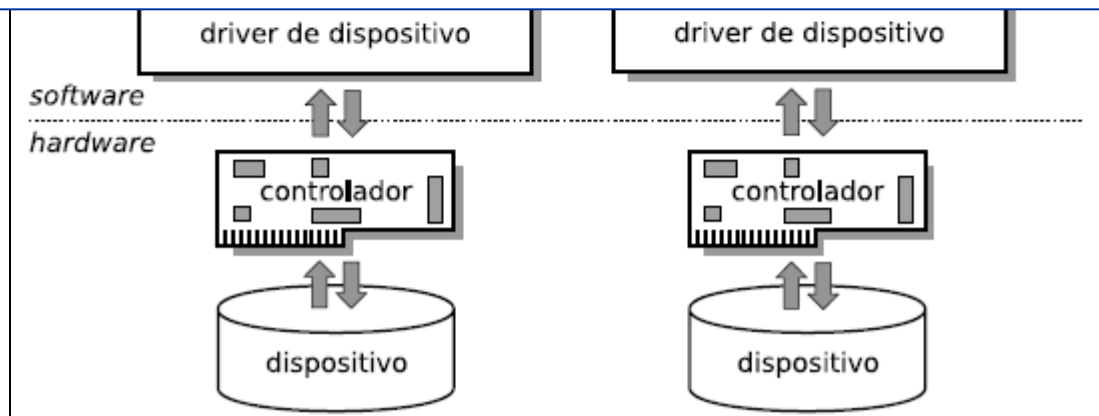
Sistemas de arquivos



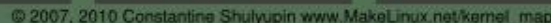
Sistemas de arquivos



Estrutura Complexa



3.8.20



Conteúdo

- Motivação
- Definição
- Objetivos
- **Conceitos fundamentais**
- Estrutura do sistema de arquivos
- Alocação de arquivos



Gerência de arquivos

- Atributos
- Formatos
- Operações
- Método de Acesso
- Compartilhamento



Gerência de arquivos

Atributos

Atributos

Cada **arquivo** é **caracterizado** por um conjunto de ***atributos***, que podem variar de acordo com o **sistema de arquivos** utilizado

Atributos

Cada **arquivo** é **caracterizado** por um conjunto de ***atributos***, que podem variar de acordo com o **sistema de arquivos** utilizado

Também chamados de *meta dados*

Atributos mais usuais

Nome: Série de caracteres que **identifica** o arquivo para o usuário

Ex.: “foto1.jpg”, “relatório.pdf”, “hello.c” etc.

Tipo: Indica **formato** dos dados contidos no arquivo

Muitos sistemas operacionais usam parte do **nome** do arquivo para identificar o **tipo** de seu conteúdo

Atributos mais usuais

Tamanho: indicação do **tamanho** do conteúdo do arquivo, em bytes ou registros

Datas: para fins de gerência; datas mais importantes relacionadas ao arquivo; datas de **criação**, de último **acesso** e de última **modificação** do conteúdo

Proprietário: em sistemas **multi-usuários**, cada arquivo tem um **proprietário**, que deve estar corretamente **identificado**

Atributos mais usuais

Permissões de acesso: indicam quais **usuários** têm **acesso** àquele arquivo e que **formas** de acesso são permitidas

- leitura, escrita, remoção, execução, etc.

Localização: indicação do **dispositivo** físico onde o arquivo se encontra e da **posição** do arquivo dentro dele

Gerência de arquivos

- Atributos
- Formatos
- Operações
- Método de Acesso
- Compartilhamento

Formatos

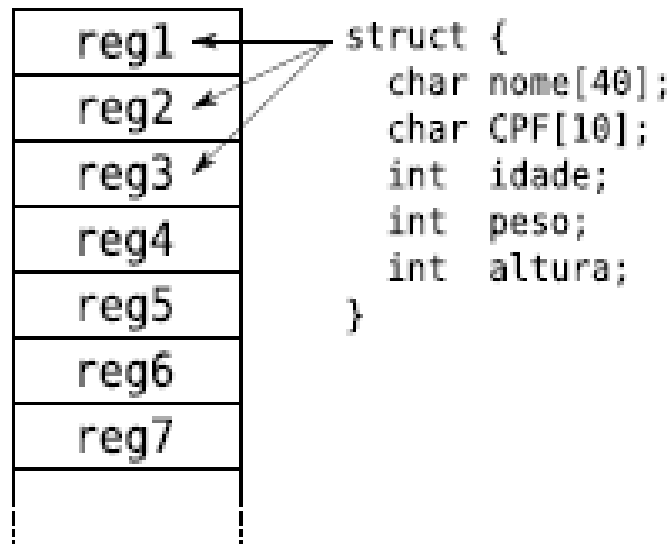
- Em sua **forma** mais simples, um arquivo contém basicamente uma **sequência de bytes**

Formatos

- Em sua **forma** mais simples, um arquivo contém basicamente uma **sequência de bytes**
- que pode estar **estruturada** de diversas formas para representar **diferentes** tipos de informação
 - Registro
 - Texto
 - Executável

Formatos - Arquivos de registro

Alguns núcleos de SO: **arquivos com estruturas internas** que vão além da simples sequência de bytes



Registros em sequência

nome (chave)	telefone (valor)
daniel	9977-1173
marina	9876-5432
henrique	8781-9750
gabriel	8858-8286
renata	9663-9293
andressa	8779-5538
guilherme	9979-4166

Registros indexados

Formatos - Arquivos de texto

- Um formato de arquivo de uso muito frequente é o arquivo de *texto puro* (ou *plain text*)
- É muito usado para armazenar informações textuais simples
 - códigos-fonte de programas
 - arquivos de configuração
 - páginas HTML, etc.

Formatos - Arquivos de texto

Formado por linhas de caracteres ASCII (*American Standard Code for Information Interchange*) de tamanho variável, separadas por **caracteres de controle**

Exemplo:

```
int main()
{
    printf("Hello, world\n");
    exit(0);
}
```

Formatos - Arquivos de texto

O arquivo de texto **hello.c** seria armazenado da seguinte forma em um ambiente UNIX:

0000	69	6e	74	20	6d	61	69	6e	28	29	0a	7b	0a	20	20	70
	i	n	t		m	a	i	n	()	\n	{	\n			p
0010	72	69	6e	74	66	28	22	48	65	6c	6c	6f	2c	20	77	6f
	r	i	n	t	f	("	H	e	l	l	o	,		w	o
0020	72	6c	64	5c	6e	22	29	3b	0a	20	20	65	78	69	74	28
	r	l	d	\	n	")	;	\n			e	x	i	t	(
0030	30	29	3b	0a	7d	0a										
	0)	;	\n	}	\n										

New Line ("\\n")

Formatos - Arquivos de texto

E em um ambiente DOS/WINDOWS:

0000	69	6e	74	20	6d	61	69	6e	28	29	0d	0a	7b	0d	0a	20
	i	n	t		m	a	i	n	()	\r	\n	{	\r	\n	
0010	20	70	72	69	6e	74	66	28	22	48	65	6c	6c	6f	2c	20
		p	r	i	n	t	f	("	H	e	l	l	o	,	
0020	77	6f	72	6c	64	5c	6e	22	29	3b	0d	0a	20	20	65	78
	w	o	r	l	d	\	n	")	;	\r	\n			e	x
0030	69	74	28	30	29	3b	0d	0a	7d	0d	0a					
	i	t	(0)	;	\r	\n	}	\r	\n					

Carriage Return (“\r”)

New Line (“\n”)

Formatos - Arquivos executáveis

São divididos internamente em várias seções

- Código
- Tabelas de símbolos (variáveis e funções)
- Listas de dependências (bibliotecas necessárias)
- Outras informações de configuração

Formatos - Arquivos executáveis

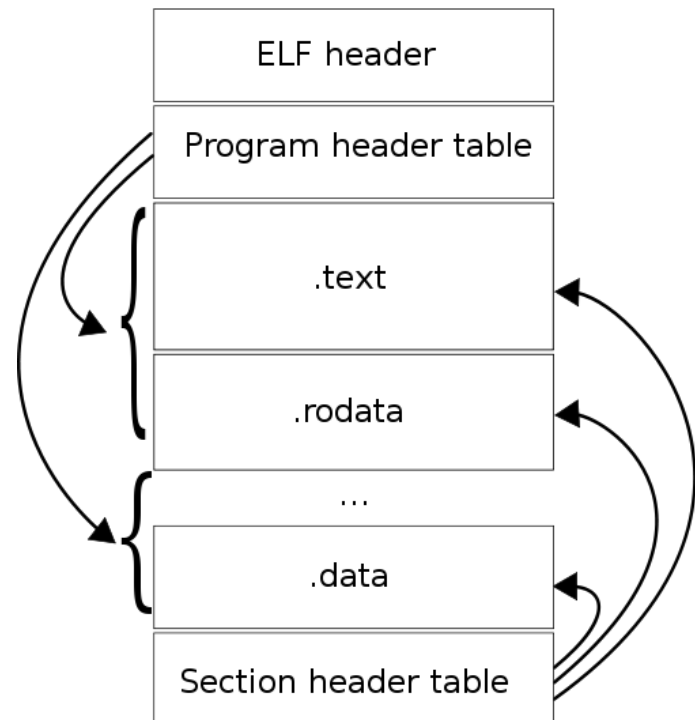
Formatos mais populares:

- **ELF** (*Executable and Linking Format*): formato para programas executáveis e bibliotecas nas plataformas UNIX modernas
- **PE** (*Portable Executable*): formato para executáveis e bibliotecas na plataforma Windows

Formatos - Arquivos executáveis

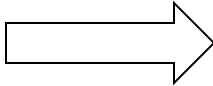
ELF

É composto por um cabeçalho e várias **seções**, contendo código executável, tabelas de símbolos e informações de relocação de código (definição de endereços de símbolos)



Formatos - Arquivos não convencionais

Além do armazenamento de código e dados, arquivos também podem ser “tratados” como:

- Diretórios
 - Atalhos (*links*)
- 
- Abordados depois

Formatos - Arquivos não convencionais

Além do armazenamento de código e dados, arquivos também podem ser “tratados” como:

- Diretórios
 - Atalhos (*links*)
 - Dispositivos físicos
 - Configurações do núcleo
 - Estruturas de comunicação do núcleo:
 - *Sockets*
 - *Pipes*
 - Filas de mensagens
- ⇒ Abordados depois

Arquivos especiais

Abstração de dispositivos de baixo nível: os sistemas UNIX costumam mapear as interfaces de acesso de vários dispositivos físicos (E/S) em arquivos dentro do diretório /dev

Ex.:

/dev/audio: placa de som

/dev/ttyS0: porta de comunicação serial COM1

Arquivos especiais

Abstração de interfaces do núcleo: em sistemas UNIX, os diretórios **/proc** e **/sys** permitem consultar e/ou modificar informações internas do **núcleo do sistema operacional**, dos processos em execução e dos *drivers* de dispositivos

Arquivos especiais

Abstração de interfaces do núcleo: em sistemas UNIX, os diretórios **/proc** e **/sys** permitem consultar e/ou modificar informações internas do **núcleo do sistema operacional**, dos processos em execução e dos *drivers* de dispositivos

Canais de comunicação: conexão TCP é apresentada aos dois processos envolvidos como um arquivo, sobre o qual eles podem **escrever (enviar)** e **ler (receber)** dados entre si

Identificação de conteúdo



Identificação de conteúdo

Já que um arquivo de dados pode ser visto como uma simples **sequência de bytes**, como é possível saber que tipo de informação essa sequência representa?

- DOS/WINDOWS: Indicar o tipo do conteúdo como parte do nome do arquivo → “**praia.jpg**”
- UNIX: Uso de alguns bytes no início de cada arquivo para a definição de seu tipo → “**números mágicos**”

Identificação de conteúdo

Ex.: UNIX

Números mágicos de alguns tipos de arquivos

Tipo de arquivo	bytes iniciais	Tipo de arquivo	bytes iniciais
Documento PostScript	%!	Documento PDF	%PDF
Imagem GIF	GIF89a	Imagem JPEG	0xFFD8
Música MIDI	MThd	Classes Java (JAR)	0xCAFEBAE

Identificação de conteúdo

Padrão **MIME** é usado para identificar arquivos transferidos como anexos de e-mail e conteúdos recuperados de páginas web

Multipurpose Internet Mail Extensions

Tipo MIME	Significado
application/java-archive	Arquivo de classes Java (JAR)
application/msword	Documento do Microsoft Word
application/vnd.oasis.opendocument.text	Documento do OpenOffice
audio/midi	Áudio em formato MIDI
audio/mpeg	Áudio em formato MP3
image/jpeg	Imagem em formato JPEG
image/png	Imagem em formato PNG
text/csv	Texto em formato CSV (<i>Comma-separated Values</i>)
text/html	Texto HTML
text/plain	Texto puro
text/rtf	Texto em formato RTF (<i>Rich Text Format</i>)
text/x-csrc	Código-fonte em C
video/quicktime	Vídeo no formato <i>Quicktime</i>



Gerência de arquivos

- Atributos
- Formatos
- Operações
- Método de Acesso
- Compartilhamento

Uso de arquivos

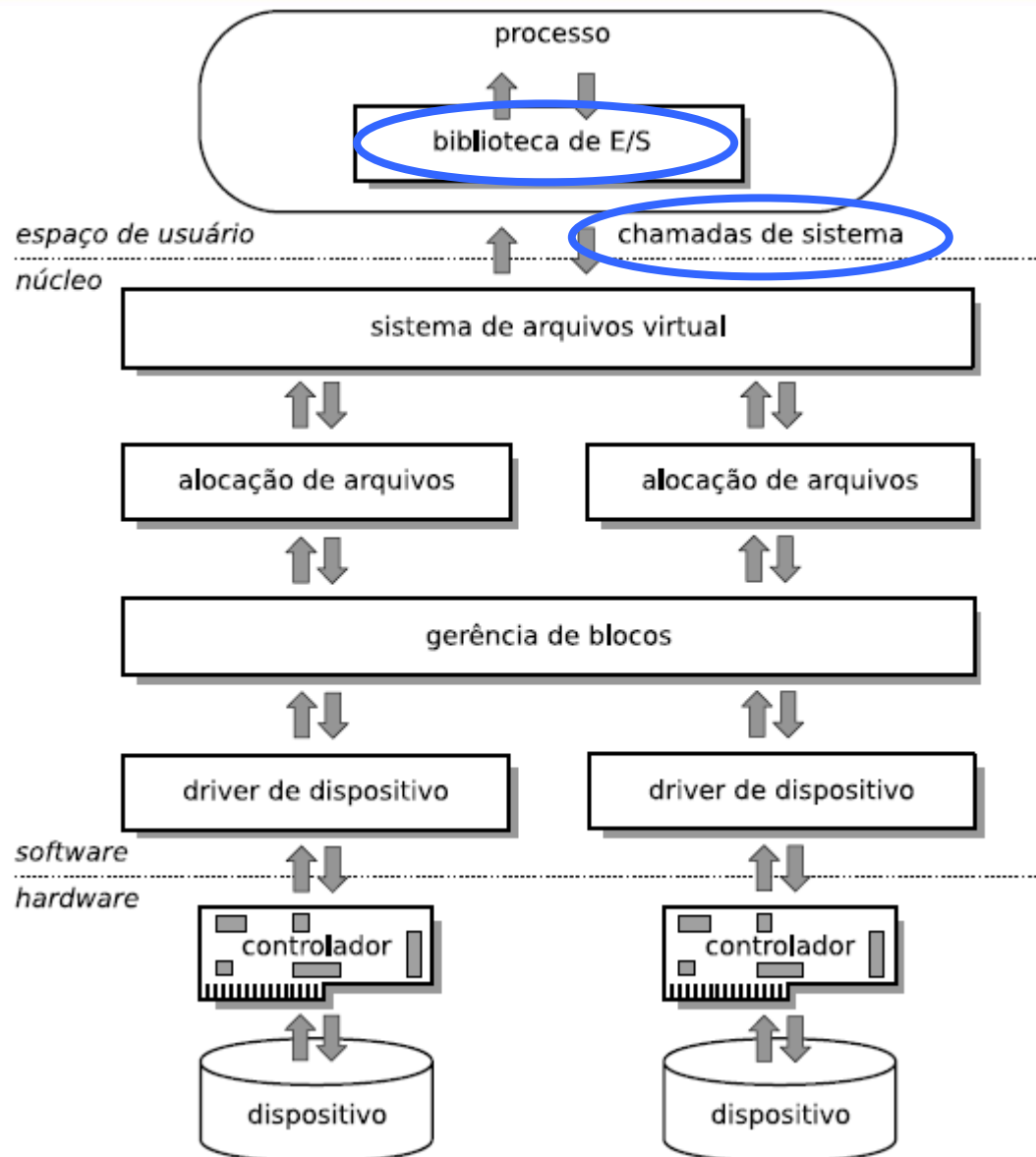


Operações

O **uso** dos arquivos é feito por meio de um conjunto de **operações**

Geralmente implementadas sob a forma de **chamadas de sistema** e **funções de bibliotecas**

Relembrando!



Uso de arquivos

- Arquivos são usados por **processos** para ler e escrever dados de forma não-volátil

Uso de arquivos

- Arquivos são usados por **processos** para ler e escrever dados de forma não-volátil
- **Processos** têm à sua disposição uma *interface de acesso*

Uso de arquivos

- Arquivos são usados por **processos** para ler e escrever dados de forma não-volátil
- **Processos** têm à sua disposição uma *interface de acesso*
- Essa **interface** normalmente é composta por uma **referência** ao arquivo e por um **conjunto de funções** para realizar **operações** sobre esses arquivos

Uso de arquivos

- Arquivos são usados por **processos** para ler e escrever dados de forma não-volátil
- **Processos** têm à sua disposição uma *interface de acesso*
- Essa **interface** normalmente é composta por uma **referência** ao arquivo e por um **conjunto de funções** para realizar **operações** sobre esses arquivos
- Por meio dessa interface, os processos podem:
 - **localizar** arquivos no disco
 - **criar arquivos, ler e modificar** seu conteúdo

Operações



Criar: **alocar espaço** no dispositivo de armazenamento e definir seus atributos

Operações

Criar: **alocar espaço** no dispositivo de armazenamento e definir seus atributos

Abrir: para **aplicação** ler ou escrever em um arquivo deve solicitar ao sistema operacional sua “abertura”

Operações

Criar: **alocar espaço** no dispositivo de armazenamento e definir seus atributos

Abrir: para **aplicação** ler ou escrever em um arquivo deve solicitar ao sistema operacional sua “abertura”

Ler: permite **transferir** dados presentes no arquivo para uma área de memória da **aplicação**

Operações

Escrever: transferir dados na memória da aplicação para o dispositivo físico

- Novos dados podem ser adicionados no final do arquivo ou sobrescrever dados já existentes
- Dados existentes podem ser removidos

Operações

Escrever: transferir dados na memória da aplicação para o dispositivo físico

- Novos dados podem ser adicionados no final do arquivo ou sobrescrever dados já existentes
- Dados existentes podem ser removidos

Mudar atributos: modificar características do arquivo

- nome, localização, proprietário, permissões, etc.

Operações

Fechar: ao concluir o uso do arquivo, a **aplicação** deve informar ao sistema operacional que ele não é mais necessário

- **Libera** as estruturas de gerência do arquivo na memória do núcleo

Operações

Fechar: ao concluir o uso do arquivo, a **aplicação** deve informar ao sistema operacional que ele não é mais necessário

- **Libera** as estruturas de gerência do arquivo na memória do núcleo

Remove: para **eliminar** o arquivo do dispositivo, descartando seus dados e liberando o espaço ocupado por ele

Abertura de arquivo

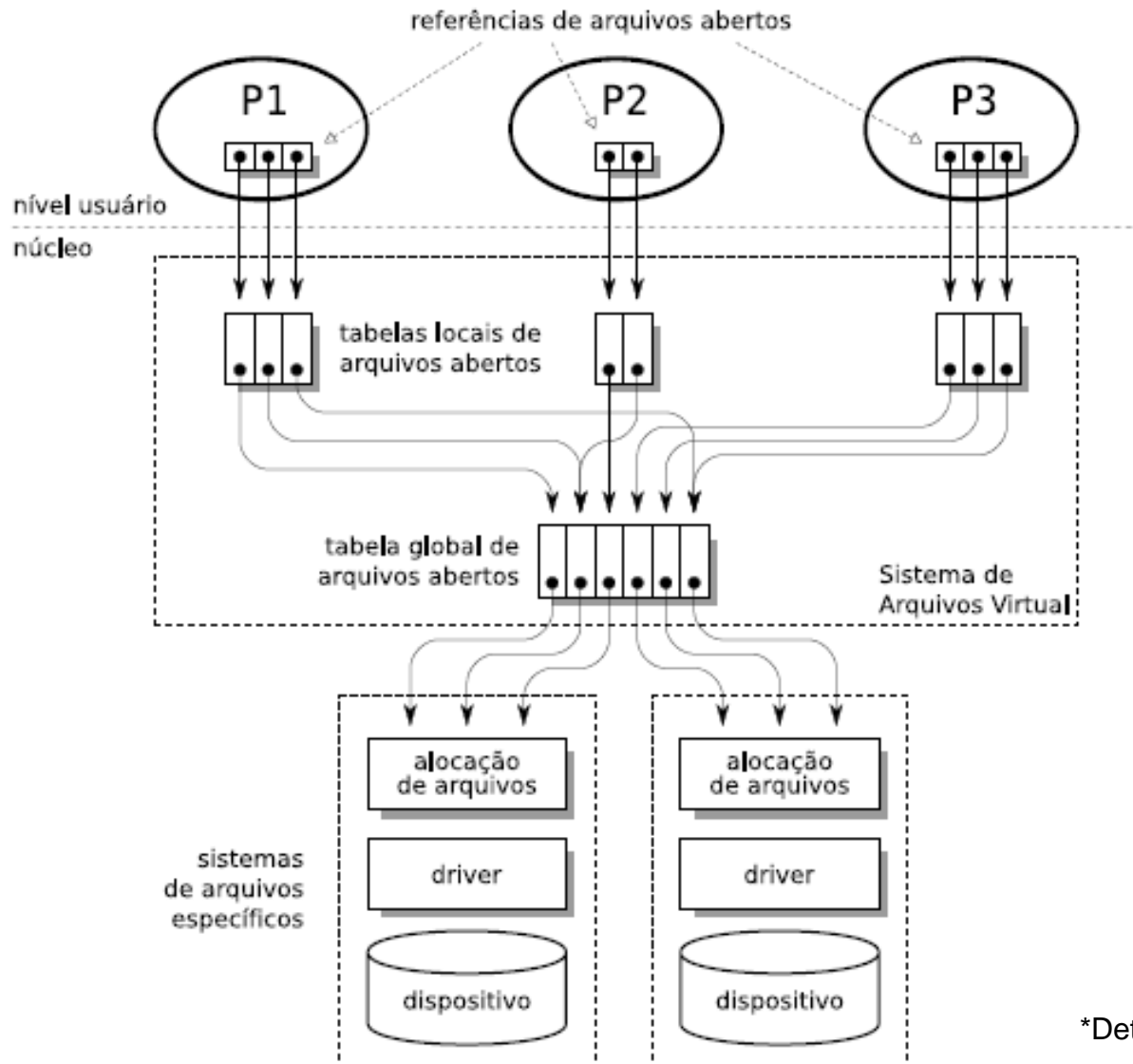


Uso de arquivos - Abertura

O núcleo do SO realiza as seguintes operações:

1. Localizar o arquivo no dispositivo físico, usando seu nome e *caminho de acesso*
2. Verificar se a aplicação tem **permissão** para usar aquele arquivo da forma desejada (leitura, escrita, remoção)
3. Criar uma **estrutura na memória** do núcleo para representar o arquivo aberto
4. Inserir uma **referência** a essa estrutura na lista de arquivos abertos mantida pelo sistema, para fins de gerência
5. Devolver à aplicação a **referência** a essa estrutura, para ser usada nos acessos subsequentes ao arquivo recém-aberto

Uso de arquivos - Abertura



*Detalhado nas próximas aulas



Gerência de arquivos

- Atributos
- Formatos
- Operações
- **Método de Acesso**
- Compartilhamento

Uso de arquivos – Formas de acesso

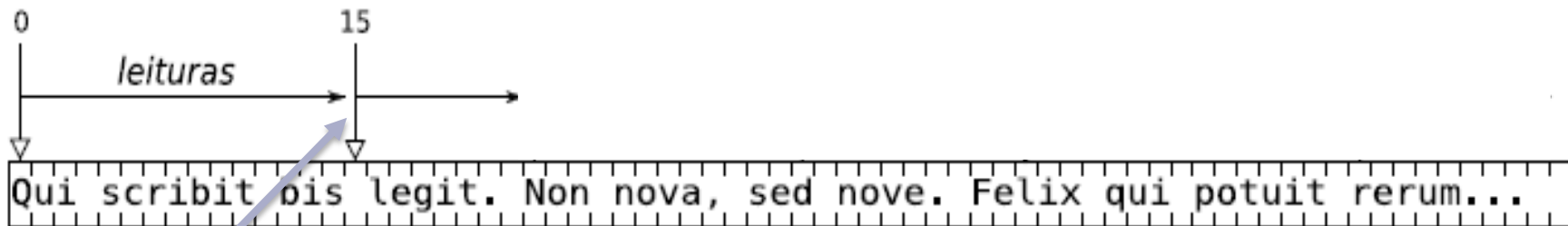
Uma vez aberto um arquivo, a aplicação pode **acessar** os dados contidos nele, **modificá-los** ou **escrever** novos dados



Duas formas de acesso são usuais: o **acesso sequencial** e o **acesso direto**

Formas de acesso - Sequencial

Os dados são sempre lidos e/ou escritos em sequência, do início ao final do arquivo



Ponteiro

Formas de acesso - Sequencial

Quando o ponteiro atinge o final do arquivo (*EoF - End-of-File*):

- Leituras não são mais permitidas
- Escritas ainda o são
 - Dados podem ser acrescentados ao final
 - Desde que haja espaço no dispositivo!

Formas de acesso - Direto

Pode-se indicar a **posição** no arquivo onde cada leitura ou escrita deve ocorrer

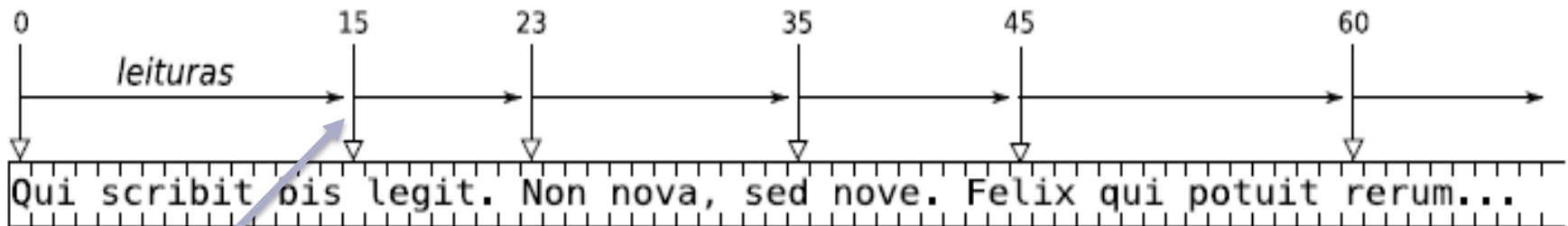
Formas de acesso - Direto

Pode-se indicar a **posição** no arquivo onde cada leitura ou escrita deve ocorrer

*Na prática, a maioria dos sistemas operacionais usa o **acesso sequencial** como modo **básico** de operação*

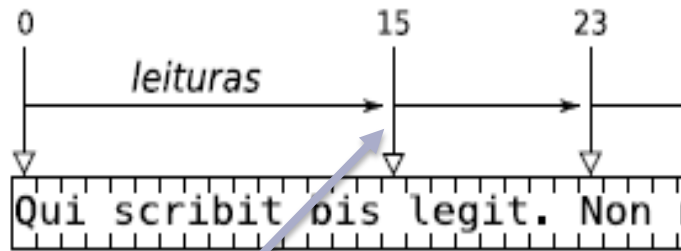
*Mas, oferece operações para mudar a posição do ponteiro do arquivo caso necessário, o que permite então o acesso **direto** a qualquer registro do arquivo*

Formas de acesso - Direto

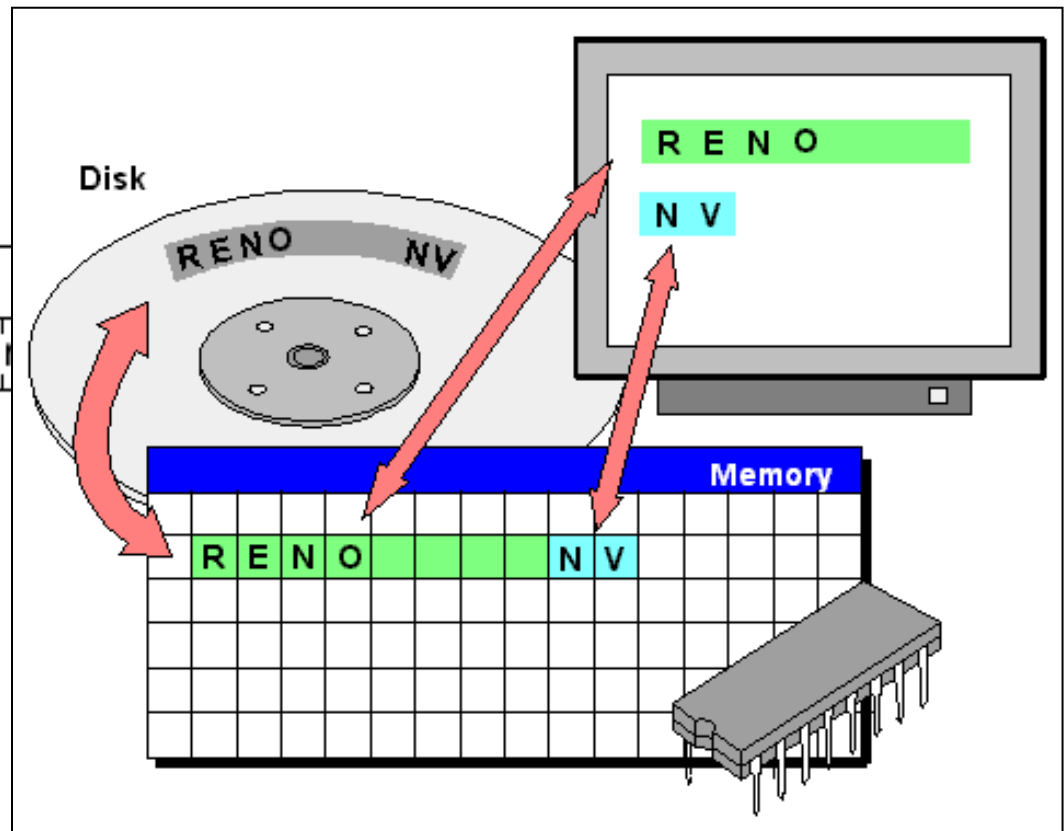


Ponteiro

Formas de acesso - Direto



Ponteiro



Controle de acesso



Controle de acesso

É importante definir claramente o **proprietário** de cada arquivo e quais **operações** ele e outros usuários do sistema podem efetuar

Controle de acesso

Proprietário: identifica o usuário dono do arquivo, geralmente aquele que o criou

Permissões de acesso: define que operações cada usuário do sistema pode efetuar sobre o arquivo

Controle de acesso

Existem muitas formas de se definir permissões de acesso a recursos em um sistema computacional

- No caso de **arquivos**:
 - Listas de controle de acesso (ACL - *Access Control Lists*)

Controle de acesso

Existem muitas formas de se definir permissões de acesso a recursos em um sistema computacional

– No caso de **arquivos**:

- Listas de controle de acesso (ACL - *Access Control Lists*)

```
1 arq1.txt : (João: ler), (José: ler, escrever), (Maria: ler, remover)
2 video.avi : (José: ler), (Maria: ler)
3 musica.mp3: (Daniel: ler, escrever, apagar)
```

Controle de acesso

Ex.: Considerando a seguinte listagem de diretório em um sistema UNIX

9 bits

```
1 host:~> ls -l
2 d rwx --- --- 2 maziero prof 4096 2008-09-27 08:43 figuras
3 - rwx r-x --- 1 maziero prof 7248 2008-08-23 09:54 hello-unix
4 - rw- r-- r-- 1 maziero prof 54 2008-08-23 09:54 hello-unix.c
5 - rw- --- --- 1 maziero prof 59 2008-08-23 09:49 hello-windows.c
6 - rw- r-- r-- 1 maziero prof 195780 2008-09-26 22:08 main.pdf
7 - rw- --- --- 1 maziero prof 40494 2008-09-27 08:44 main.tex
```


Controle de acesso

Ex.: Considerando a seguinte listagem de diretório em um sistema UNIX

```
1 host:~> ls -l
2 d rwx --- --- 2 maziero prof 4096 2008-09-27 08:43 figuras
3 - rwx r-x --- 1 maziero prof 7248 2008-08-23 09:54 hello-unix
4 - rw- r-- r-- 1 maziero prof 54 2008-08-23 09:54 hello-unix.c
5 - rw- --- --- 1 maziero prof 59 2008-08-23 09:49 hello-windows.c
6 - rw- r-- r-- 1 maziero prof 195780 2008-09-26 22:08 main.pdf
7 - rw- --- --- 1 maziero prof 40494 2008-09-27 08:44 main.tex
```

Permissões do proprietário

Permissões do grupo

Permissões dos demais usuários



Gerência de arquivos

- Atributos
- Formatos
- Operações
- Método de Acesso
- **Compartilhamento**

Compartilhamento de Arquivos



Compartilhamento de Arquivos

Em um sistema **multi-tarefas**, é frequente ter arquivos acessados por **mais de um processo**, ou mesmo **mais de um usuário**, caso as permissões de acesso ao mesmo o permitam

O acesso concorrente em **leitura** a um arquivo não acarreta problemas, mas a possibilidade de **escritas** e leituras simultâneas tem de ser prevista e tratada de forma adequada

Compartilhamento – Travas de acesso

A solução mais simples e mais frequentemente utilizada para gerenciar o acesso compartilhado a arquivos é o uso de **travas de exclusão mútua** (*mutex locks*)

Compartilhamento – Travas de acesso

As travas oferecidas pelo sistema operacional podem ser **obrigatórias** (*mandatory locks*) ou **recomendadas** (*advisory locks*)

- **Travas obrigatórias** são impostas pelo núcleo de forma incontornável
- **Travas recomendadas** não são impostas pelo núcleo do sistema operacional

Compartilhamento – Travas de acesso

As travas sobre arquivos também podem ser **exclusivas** ou **compartilhadas**

- **Trava exclusiva (escrita)**: garante acesso exclusivo ao arquivo: enquanto uma trava exclusiva estiver ativa, nenhum outro processo poderá obter uma trava sobre aquele arquivo

Compartilhamento – Travas de acesso

As travas sobre arquivos também podem ser **exclusivas** ou **compartilhadas**

- **Trava exclusiva (escrita)**: garante acesso exclusivo ao arquivo: enquanto uma trava exclusiva estiver ativa, nenhum outro processo poderá obter uma trava sobre aquele arquivo
- **Trava compartilhada (leitura)**: impede outros processos de criar travas exclusivas sobre aquele arquivo, mas permite a existência de outras travas compartilhadas

Compartilhamento – Semântica de acesso

A forma como os **dados escritos por um processo** são percebidos pelos demais processos que abriam aquele arquivo é chamada de ***semântica de compartilhamento***.

- *UNIX*
- *Sessão*
- *Imutável*

Compartilhamento – Semântica de acesso

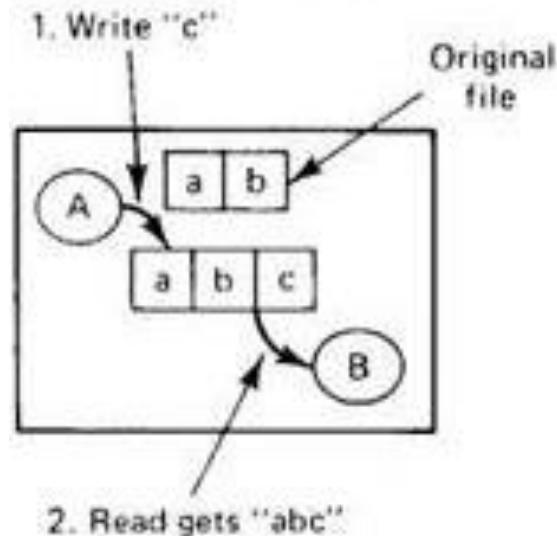
Semântica UNIX: toda modificação em um arquivo é imediatamente **visível a todos** os processos que mantêm aquele arquivo aberto

- Mais comum em sistemas de arquivos locais

Compartilhamento – Semântica de acesso

Semântica UNIX: toda modificação em um arquivo é imediatamente **visível a todos** os processos que mantêm aquele arquivo aberto

- Mais comum em sistemas de arquivos locais



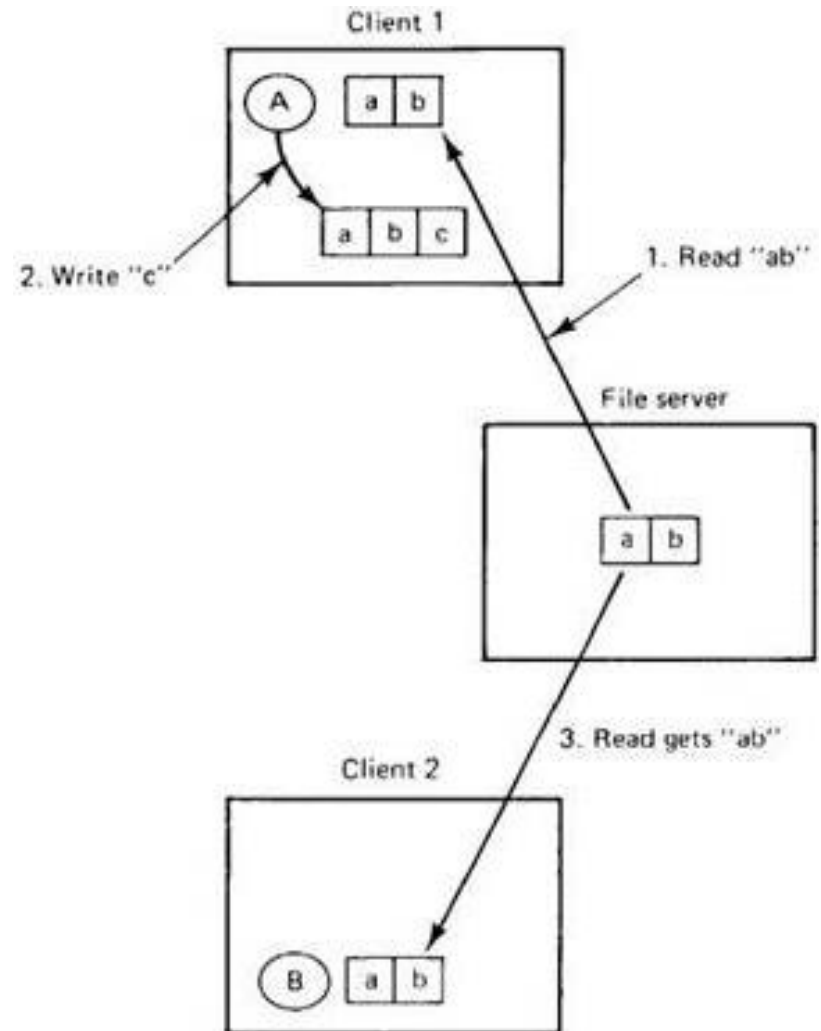
Compartilhamento – Semântica de acesso

Semântica de sessão: Modificações em um arquivo feitas em uma sessão somente são visíveis na **mesma sessão** e pelas sessões que **iniciarem depois** do seu encerramento

- Normalmente aplicada a sistemas de arquivos de rede

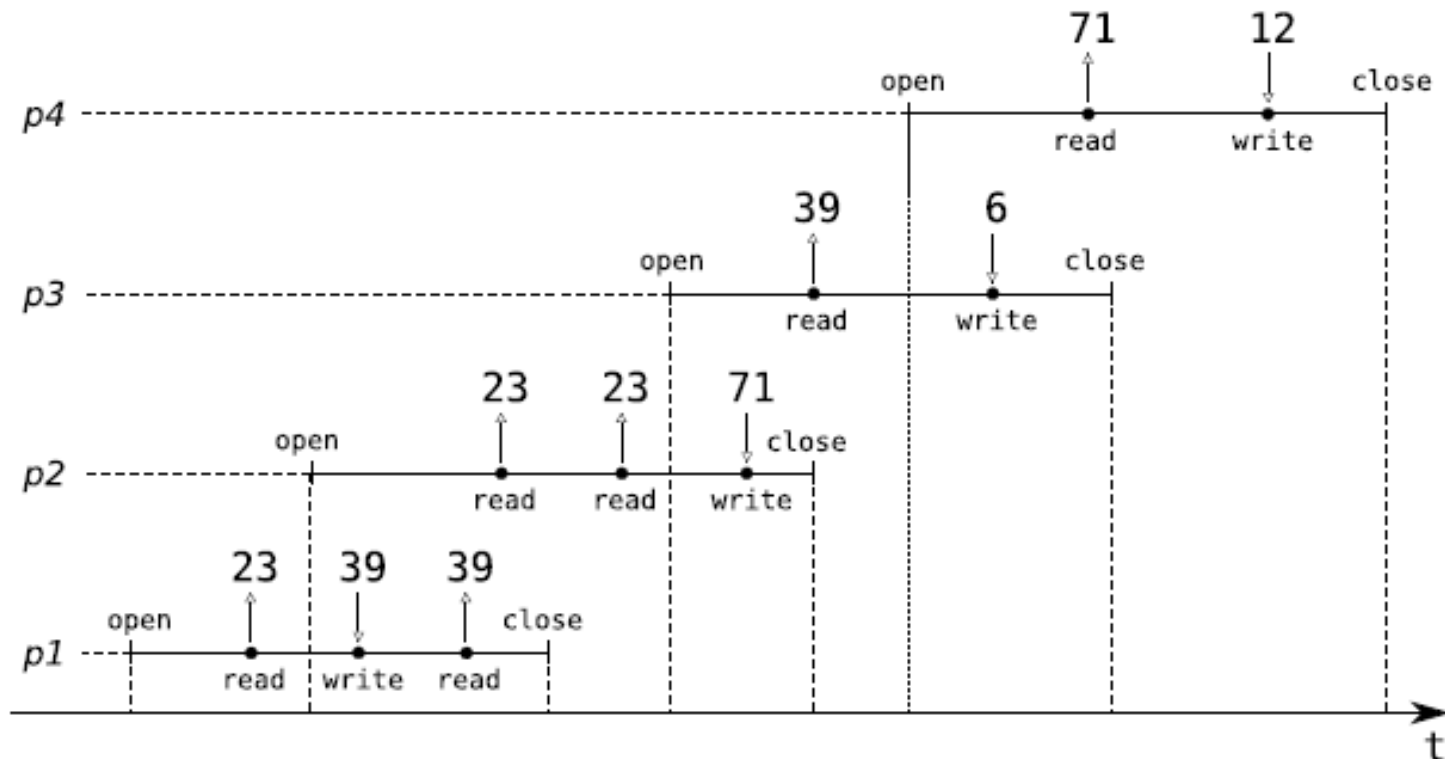
Compartilhamento – Semântica de acesso

Semântica de sessão



Compartilhamento – Semântica de acesso

Semântica de sessão



Compartilhamento – Semântica de acesso

Semântica imutável: Se um arquivo pode ser compartilhado por vários processos, ele é marcado como imutável, ou seja, seu conteúdo não pode ser modificado

- Usada em alguns sistemas de arquivos distribuídos*

Conteúdo

- Motivação
- Definição
- Objetivos
- Conceitos fundamentais
- **Estrutura do sistema de arquivos**
- Alocação de arquivos

Conteúdo

- Estrutura do sistema de arquivos
 - Organização de volume
 - Diretórios
 - Arquivos
- Alocação de arquivos
 - Tabela de arquivos
 - Estrutura *i-node*