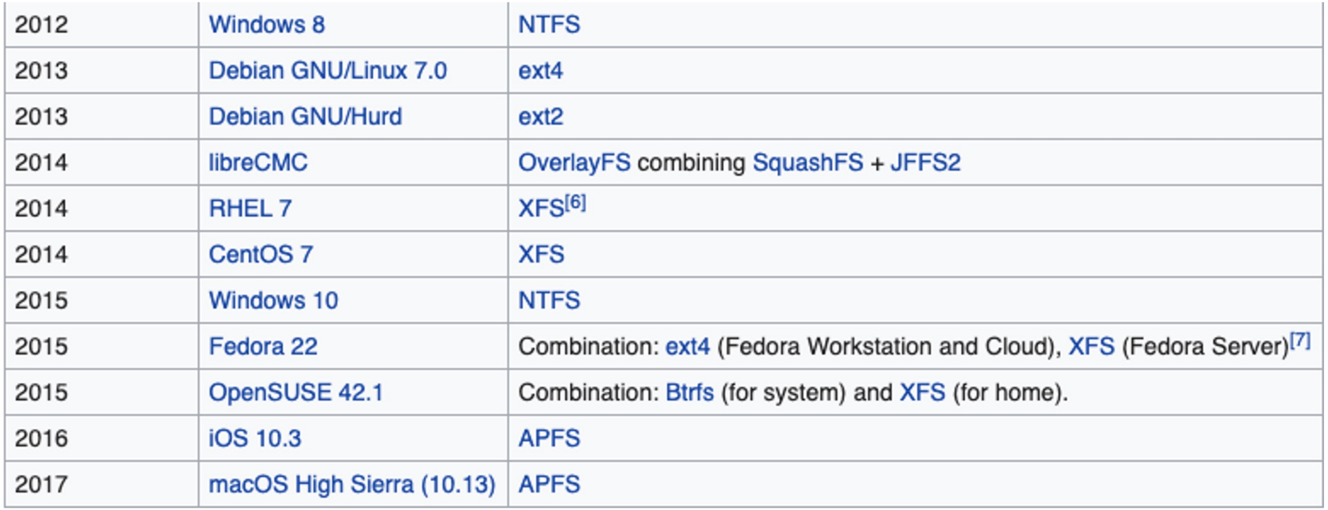
377 Operating Systems

File System Implementation

File System Implementation

- File systems are generally pure software in an OS
- A file system is an on-disk data structure(s) that the OS interacts with using open(), read(), write(), etc. files
- File systems get pretty complicated, so we will look at a simplified one that shares aspects of real ones

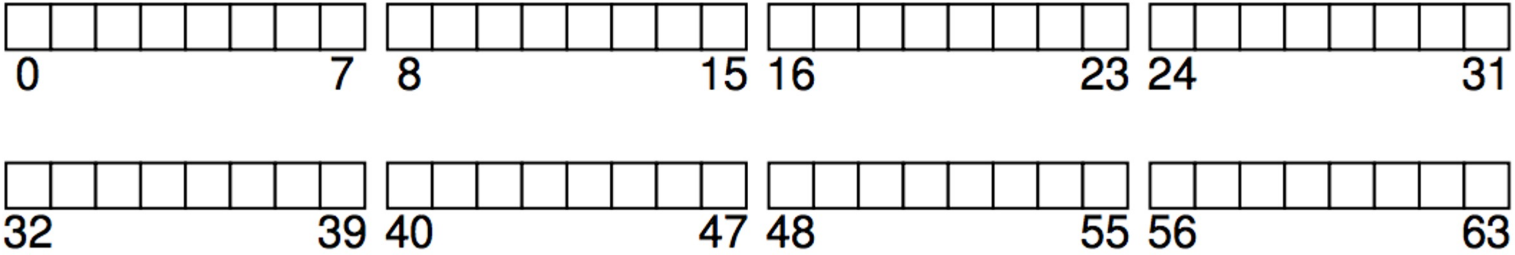
Lots of Different FSs



2012	Windows 8	NTFS
2013	Debian GNU/Linux 7.0	ext4
2013	Debian GNU/Hurd	ext2
2014	libreCMC	OverlayFS combining SquashFS + JFFS2
2014	RHEL 7	XFS[6]
2014	CentOS 7	XFS
2015	Windows 10	NTFS
2015	Fedora 22	Combination: ext4 (Fedora Workstation and Cloud), XFS (Fedora Server)[7]
2015	OpenSUSE 42.1	Combination: Btrfs (for system) and XFS (for home).
2016	iOS 10.3	APFS
2017	macOS High Sierra (10.13)	APFS

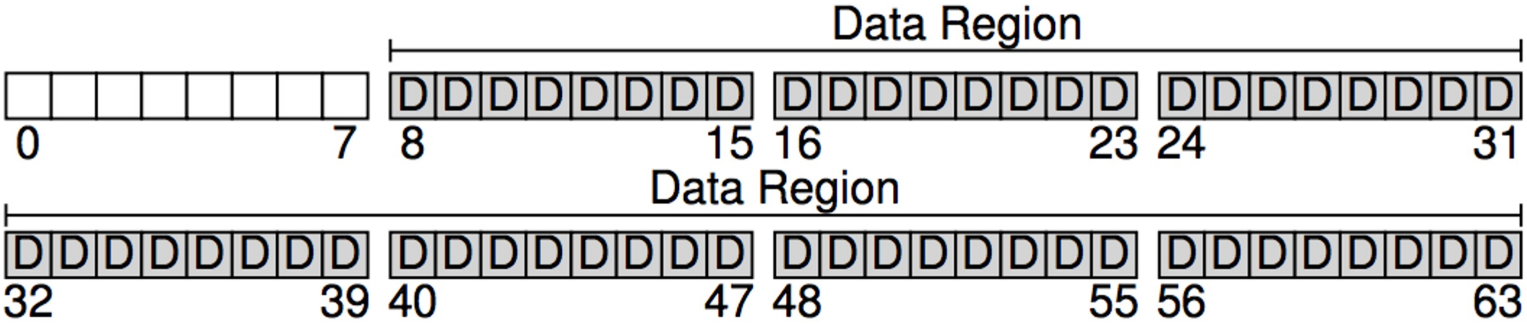
Blocks

- To build our file system we divide up the disk into a series of blocks (not the same as the disk sectors, which may be smaller!)
- Let’s start with 4KB blocks (fairly common size)
- For example a small FS with 64 blocks



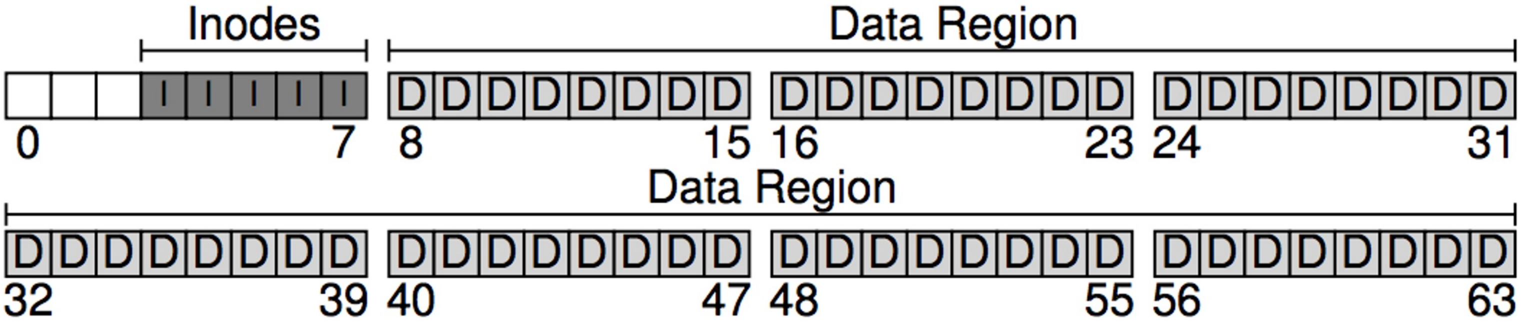
Data Region

- Let’s reserve most of the file system for the users’ data
- We will call that the Data Region


Data Region
D D D D D D D D D D D D D D D D D D D D D D D D
0
7
8
15
16
23
24
31
Data Region
D D D D D D D D D D D D D D D D D D D D D D D D D D D D D D D D
32
39
40
47
48
55
56
63

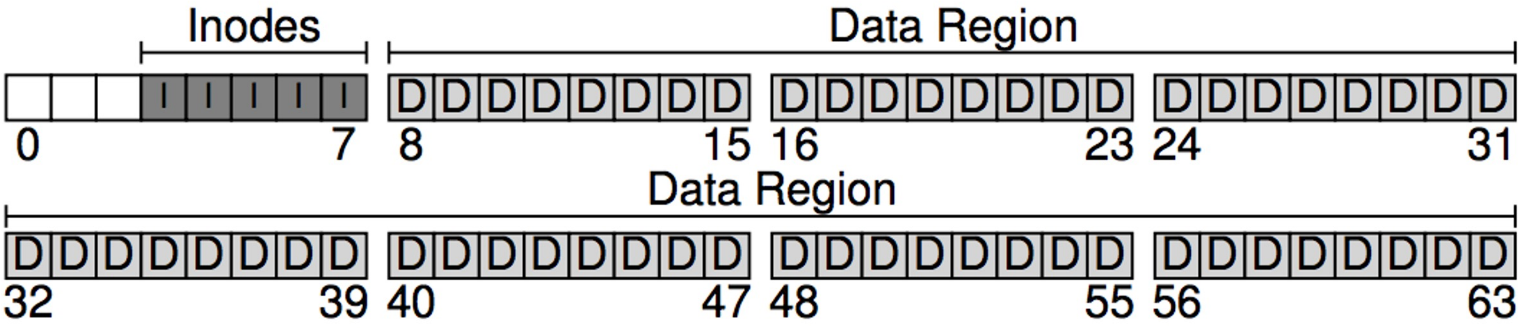
Inodes

- inodes are a structure that contain information about the size of the file, its creation data, etc.
- That data is called metadata
- We will reserve 5 blocks for inodes



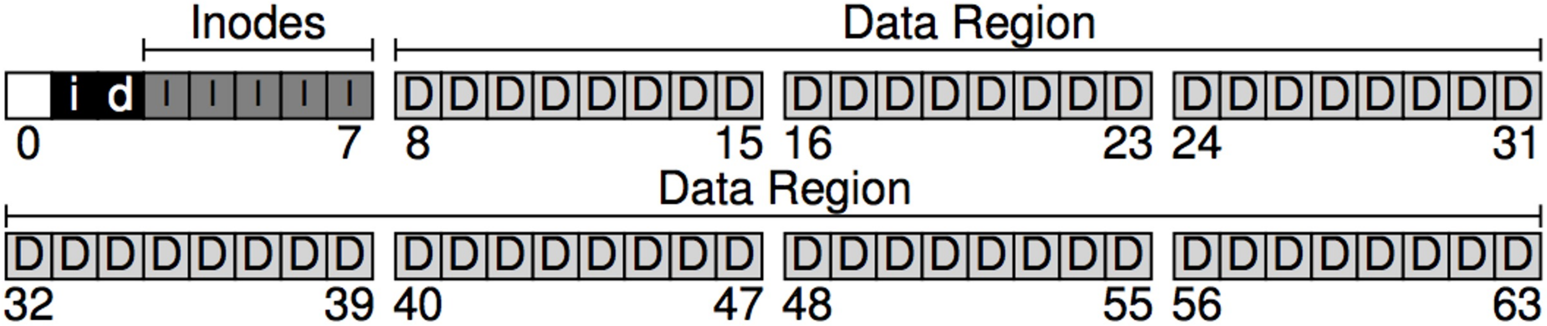
Inodes

- Inodes don’t need a whole block for the metadata for each file.
- 256 bytes should be good.
- So, if each file is identified by an inode, how many files does this file system support?



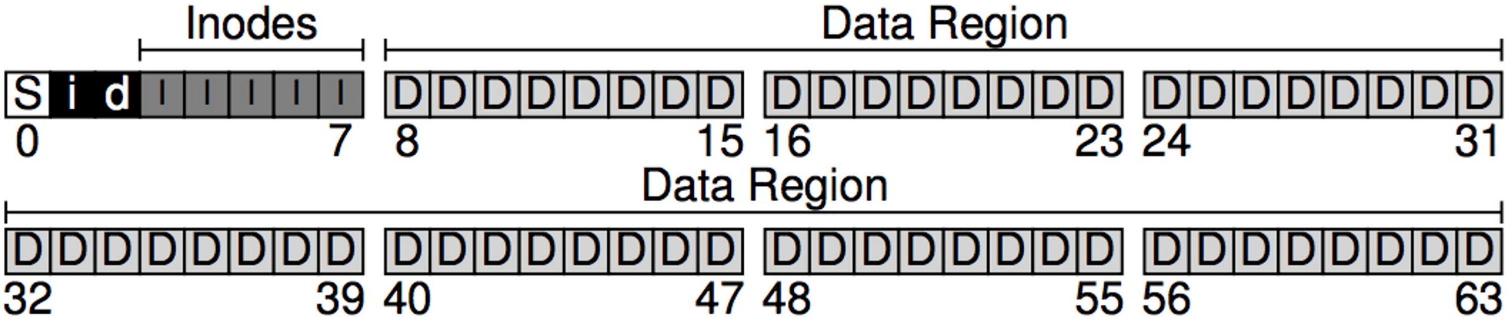
Free Space

- We need to track which inodes and which data blocks are free/used.
- A simple structure to track such things is a bitmap
- One bit per inode (80 bits), and one bit per data block (56 bits). But let's be lazy and use a whole block for each



Superblock

- When mounting a file system, the OS needs to know which kind of file system is on the disk, how many inodes, etc.
- This is the metadata about the file system itself
- It is stored in the 0th block. It is the **superblock**

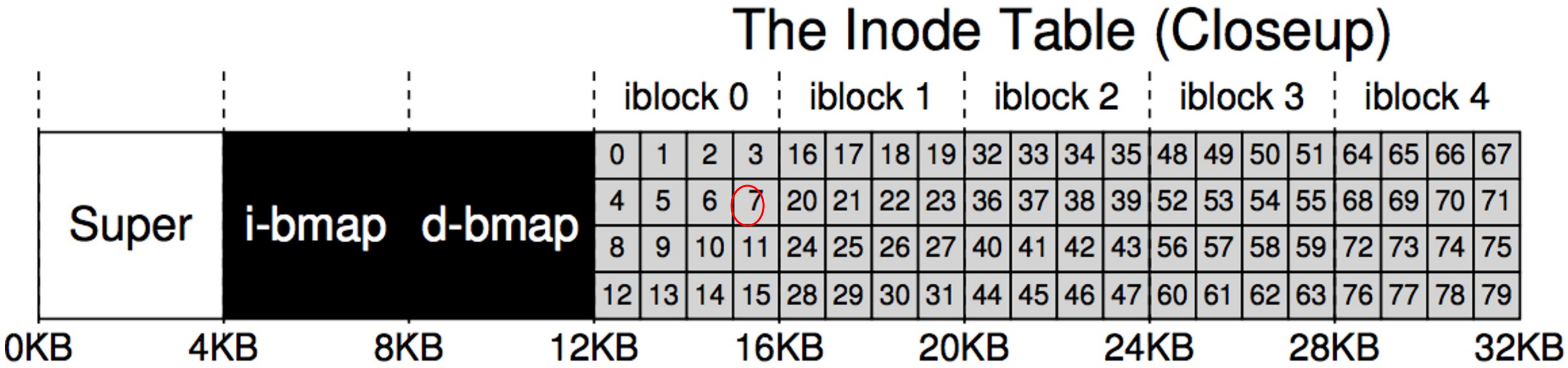


Finding an inode

- Each inode on the system has a number (recall these will be important when reading directories)

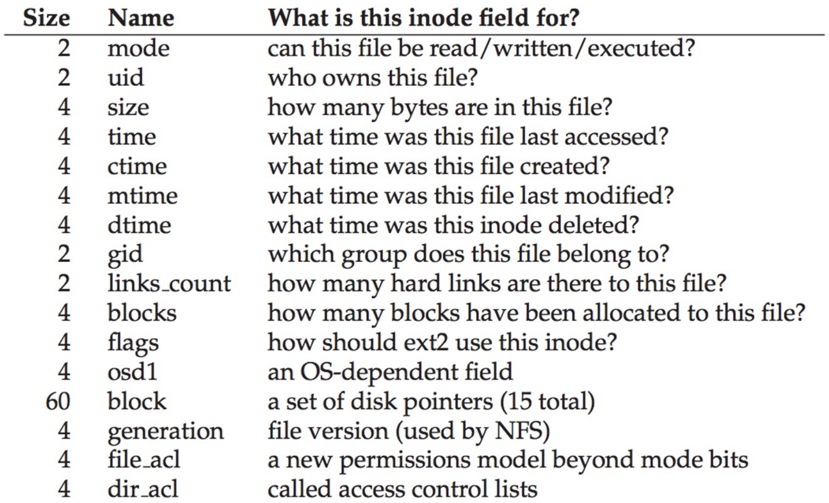
- If you are looking for inode 32 you need to know that address on disk:

32 * sizeof(inode) + start of inode region = 8192 + 12kB = 20kB



iblock 0				iblock 1				iblock 2				iblock 3				iblock 4			
0	1	2	3	16	17	18	19	32	33	34	35	48	49	50	51	64	65	66	67
4	5	6	7	20	21	22	23	36	37	38	39	52	53	54	55	68	69	70	71
8	9	10	11	24	25	26	27	40	41	42	43	56	57	58	59	72	73	74	75
12	13	14	15	28	29	30	31	44	45	46	47	60	61	62	63	76	77	78	79

What is in the inode??



Size	Name	What is this inode field for?
2	mode	can this file be read/written/executed?
2	uid	who owns this file?
4	size	how many bytes are in this file?
4	time	what time was this file last accessed?
4	ctime	what time was this file created?
4	mtime	what time was this file last modified?
4	dtime	what time was this inode deleted?
2	gid	which group does this file belong to?
2	links_count	how many hard links are there to this file?
4	blocks	how many blocks have been allocated to this file?
4	flags	how should ext2 use this inode?
4	osd1	an OS-dependent field
60	block	a set of disk pointers (15 total)
4	generation	file version (used by NFS)
4	file_acl	a new permissions model beyond mode bits
4	dir_acl	called access control lists

- Simplified ext2fs node
- Notice: **size**, and **block** pointers

Direct Pointers

- If the inode has a fixed number of pointers, this defines the max size of the file: **pointers * block size**
- If there are 12 direct pointers, then the maximum size of a file on this file system is **48KB**.

Direct Pointers

- If the inode has a fixed number of pointers, this defines the max size of the file: **pointers * block size**

- If there are 12 direct pointers, then the maximum size of a file on this file system is **48KB**.

This is rather limiting.
Any ideas on how to allow for bigger files?
Think about how we solved our page table issues...

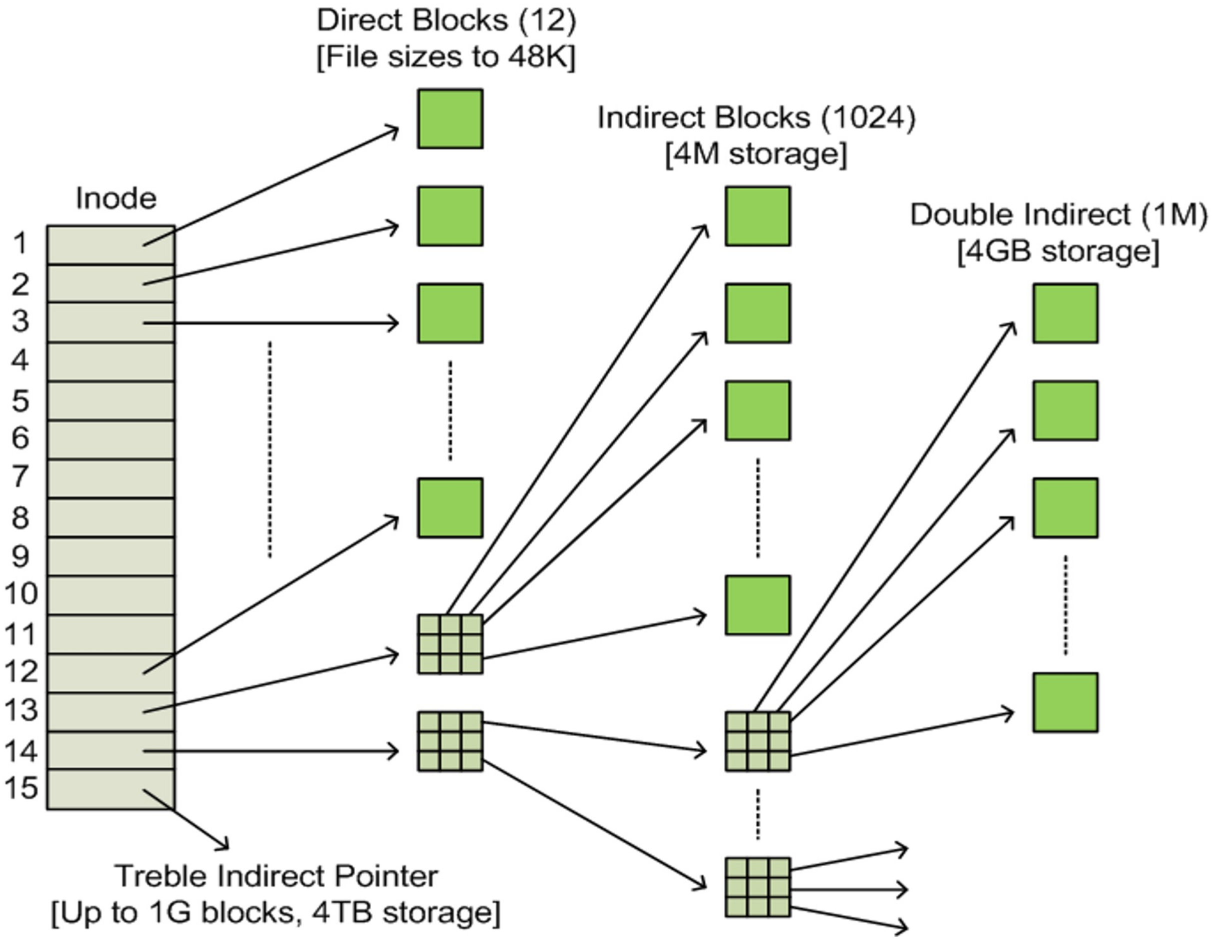
Indirect Pointers

- If we need bigger files, we can use an additional **indirect pointer** which is a pointer to a data block, filled with pointers

- A 4KB block with 4-byte pointers = 1024 pointers in a block

- So (12+1024)*4kB = 4144KB

- Or you can use an additional double indirect pointers: each indirect block points to an indirect block: (12+1024*1024)*4KB = 4GB



http://richiervs.blogspot.com/2012/07/understanding-unix-like-file-systems.html

Triple indirect?

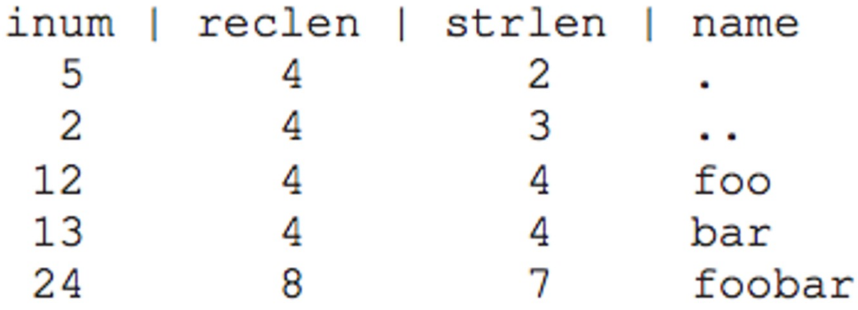
- If you want to go even larger, you can add in a triple indirect pointer!
- This totally unbalanced tree is a bit mad but works reasonably since almost all files are small.

Aside: Modern FSs

- Modern file systems (XFS, ext4fs) use a B+Tree with “extents” instead of the unbalanced tree
- Extents are a structure that describes a range of blocks (starting block + number of blocks).
- This leads to great compression of the block map

Storing Directories

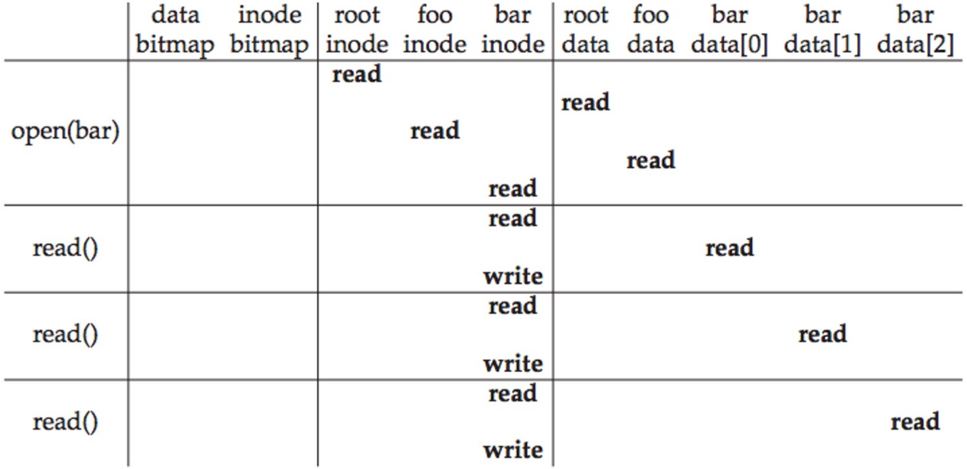
- Directories are often treated like special files
- They are just a set of data blocks containing directory entries that are names + inodes
- Its parent directory contains an entry that points to the inode for the directory



inum	reclen	strlen	name
5	4	2	.
2	4	3	..
12	4	4	foo
13	4	4	bar
24	8	7	foobar

Example:Reading

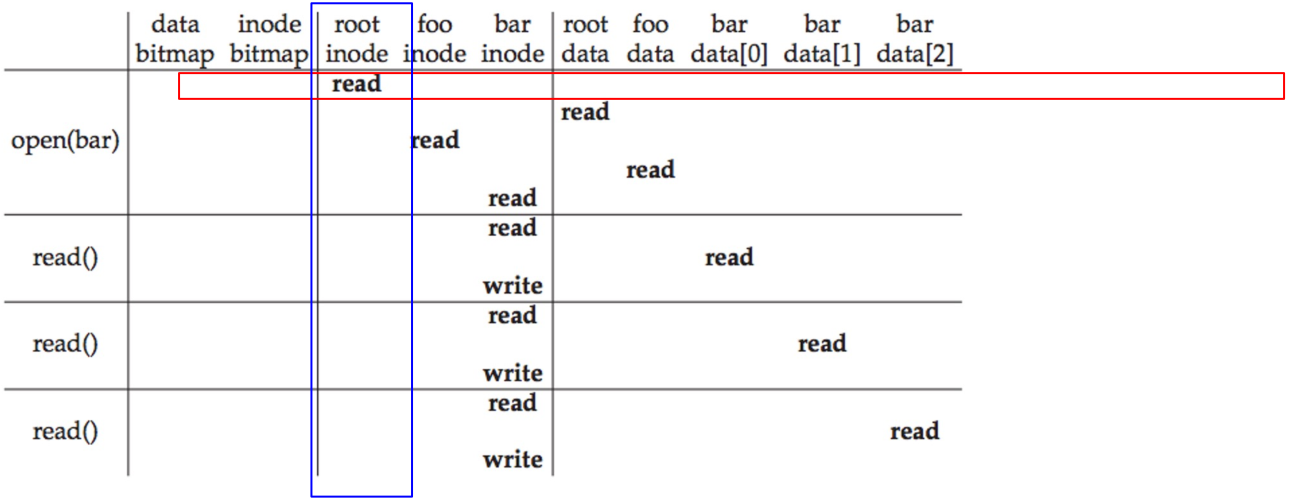
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
open(bar)			**read**							
						read				
				read						
							read			
					read					
read()					**read**					
								read		
					write					
read()					**read**					
									read	
					write					
read()					**read**					
										read
					write					

Example:Reading

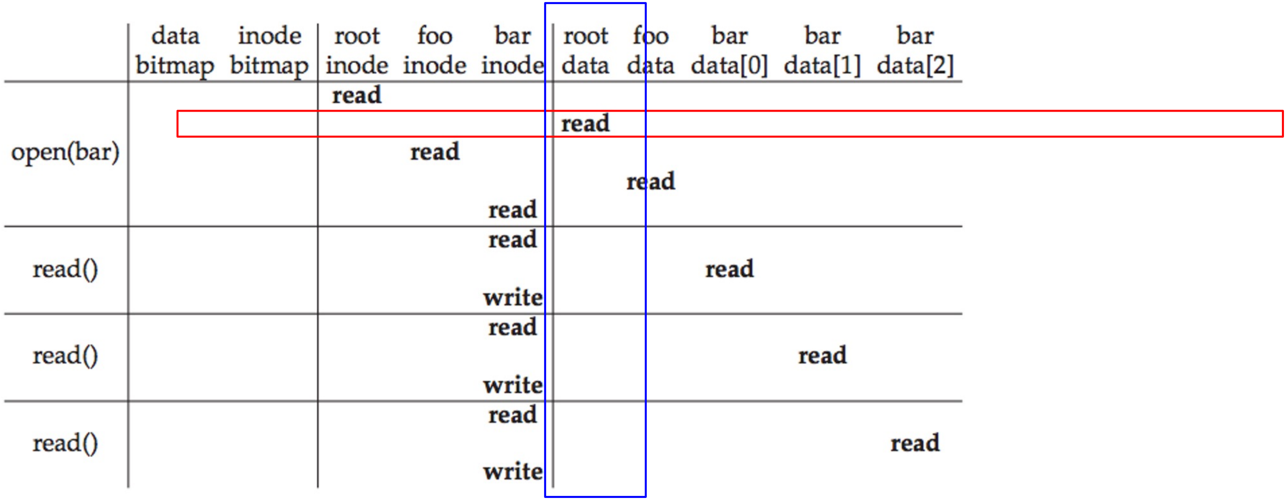
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

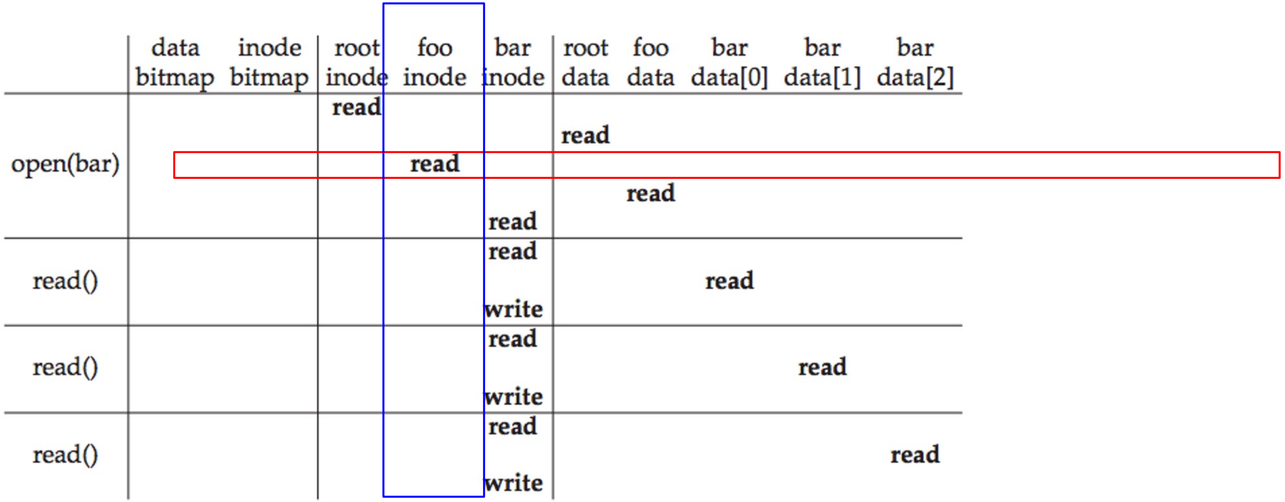
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			**read**							
						read				
				read						
							read			
					read					
read()					**read**					
								read		
					write					
read()					**read**					
									read	
					write					
read()					**read**					
										read
					write					

Example:Reading

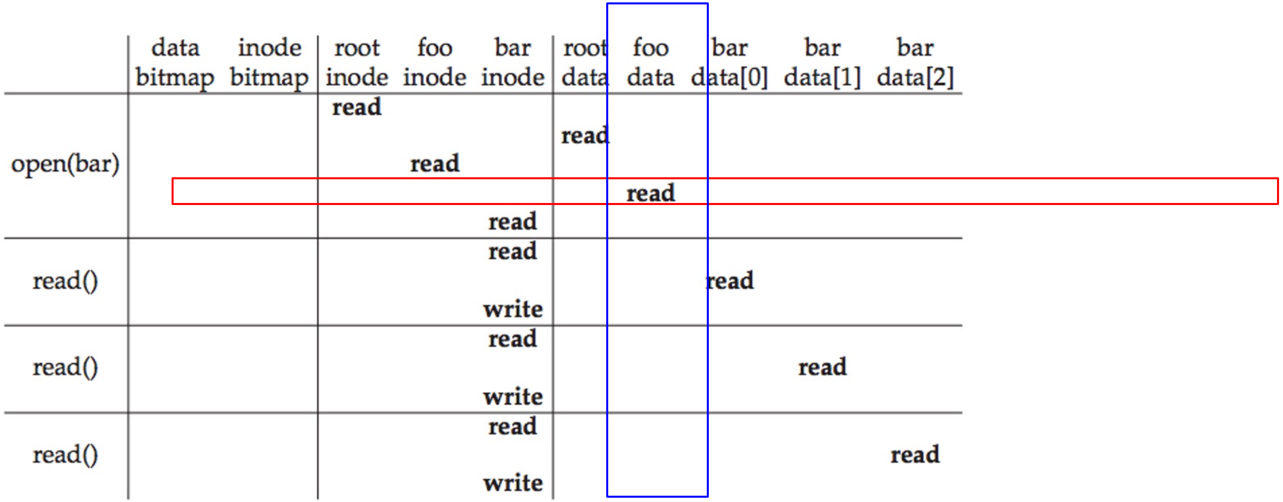
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

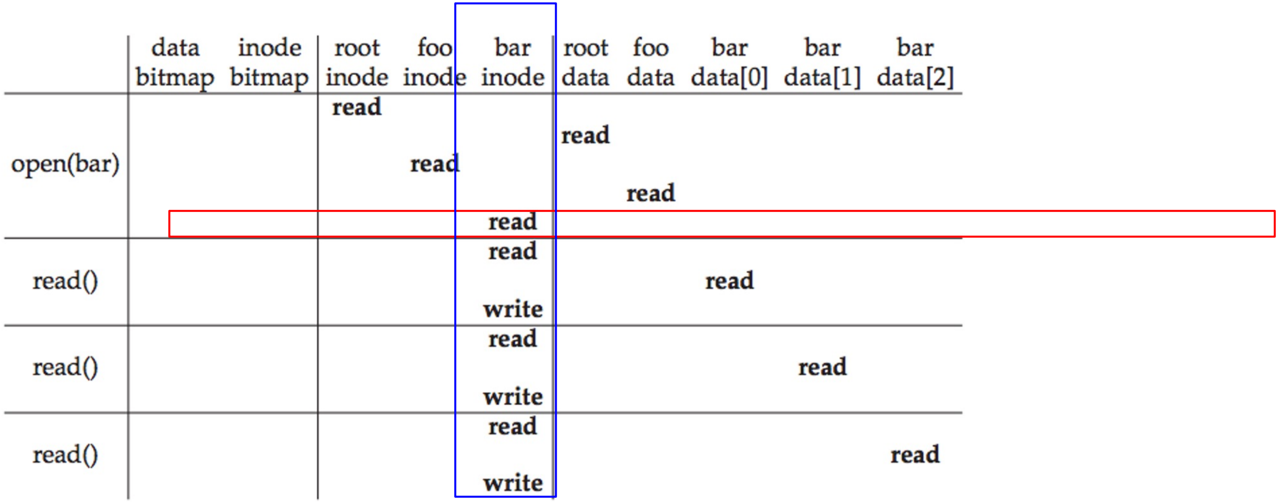
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

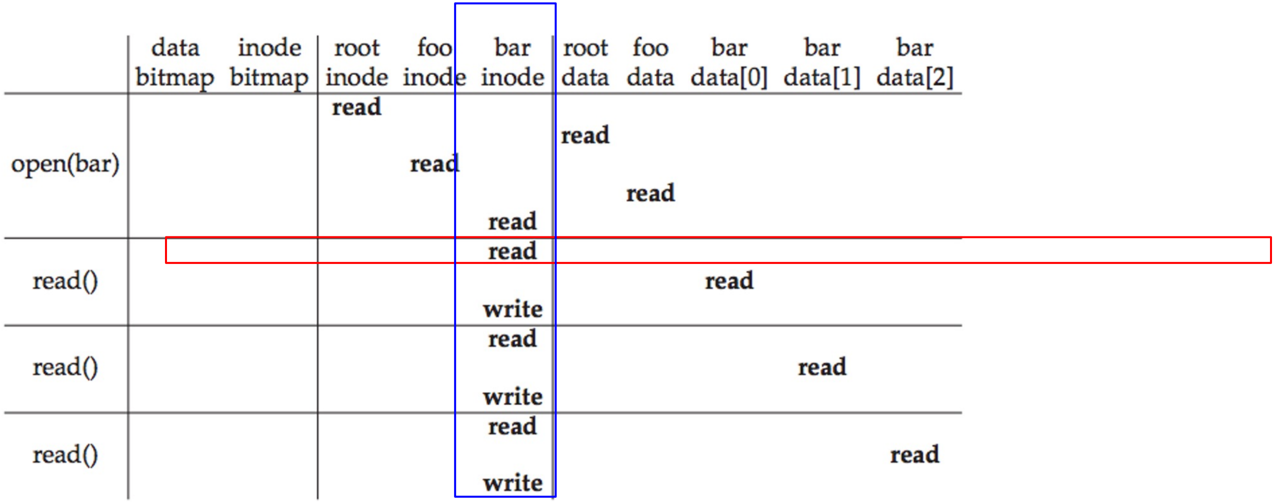
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

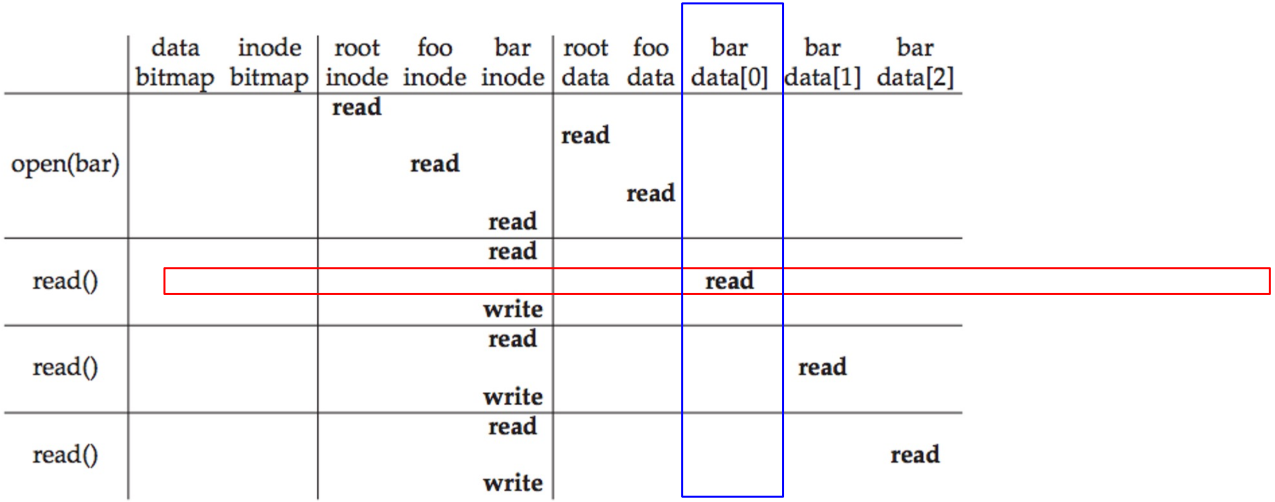
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

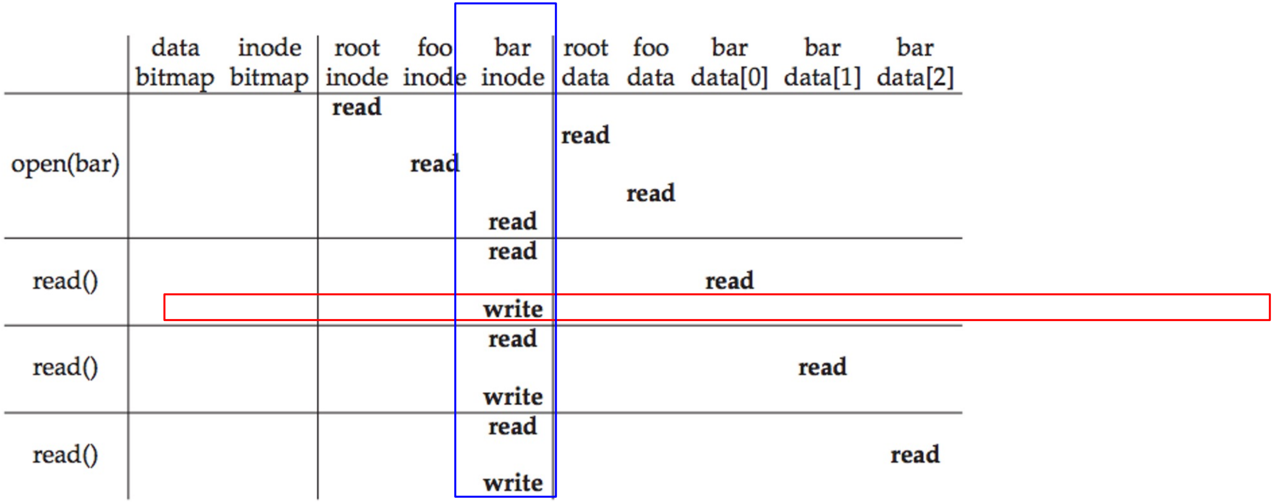
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

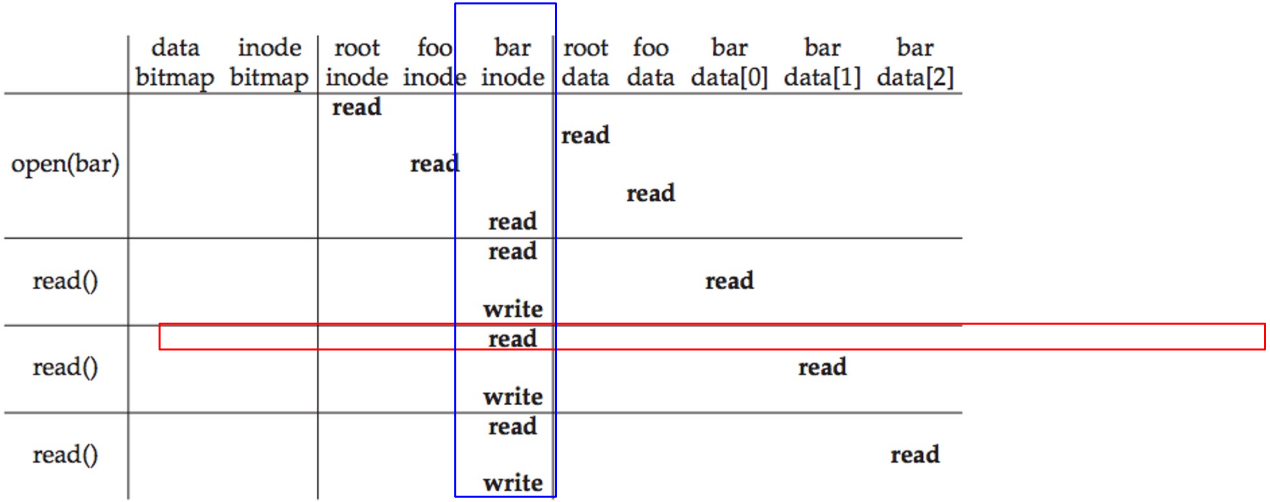
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

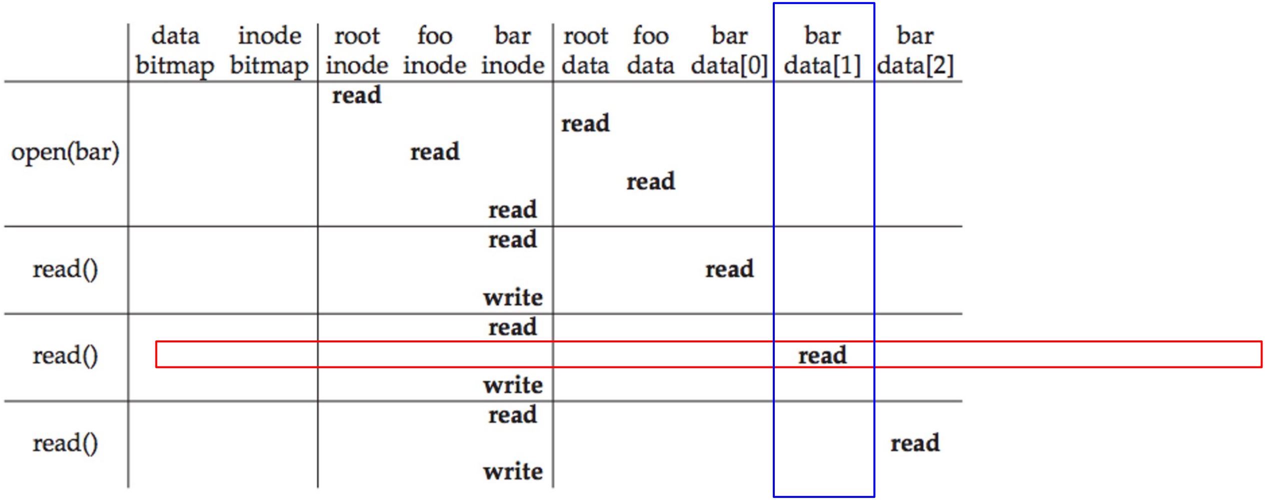
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			**read**							
						read				
				read						
							read			
					read					
read()					**read**					
								read		
					write					
read()					**read**					
									read	
					write					
read()					**read**					
										read
					write					

Example:Reading

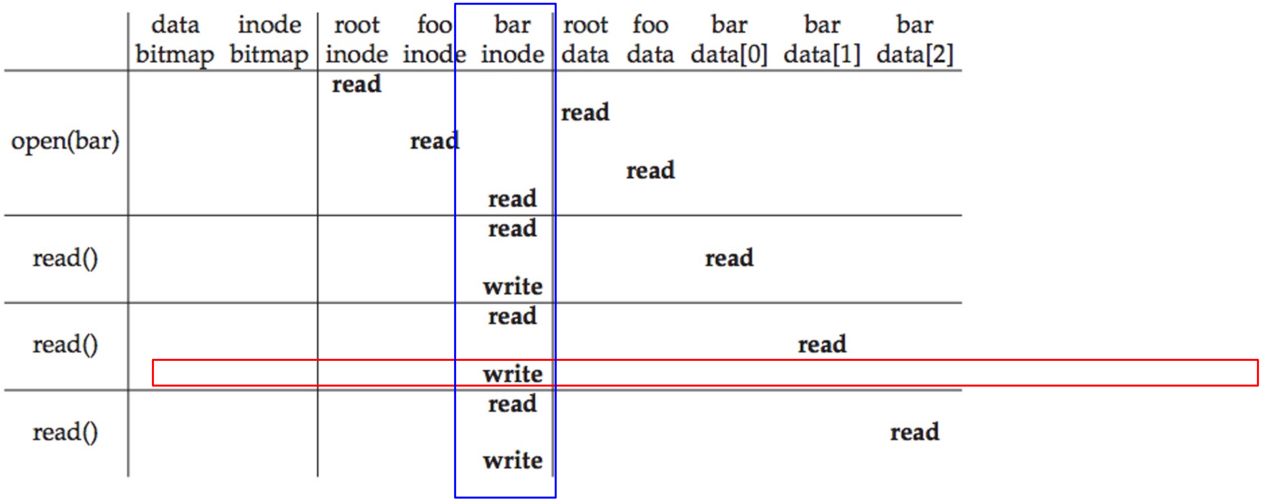
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



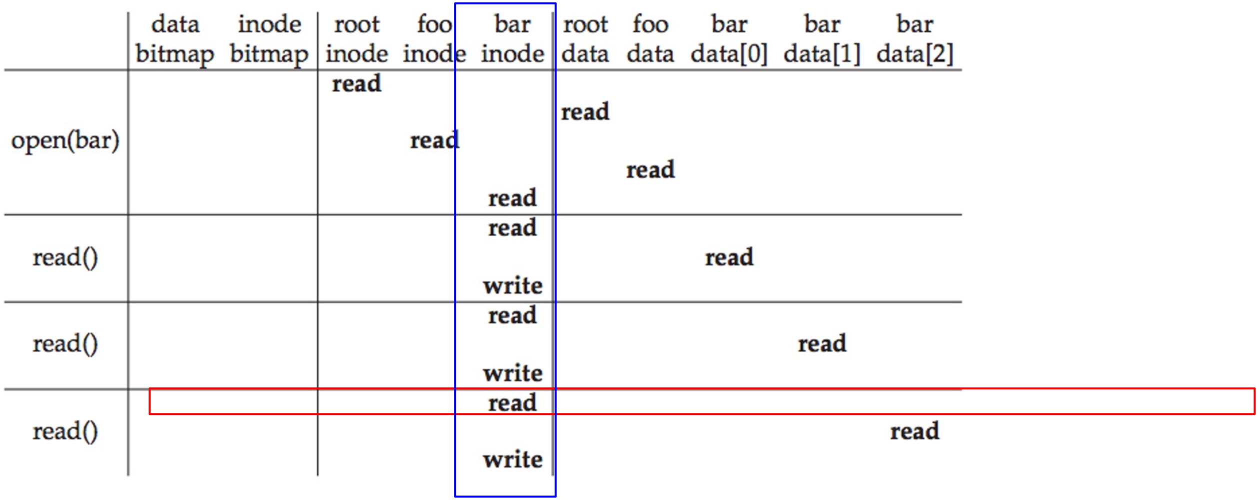
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

- Let’s try and read a file: /foo/bar

- We have to traverse directories and inodes

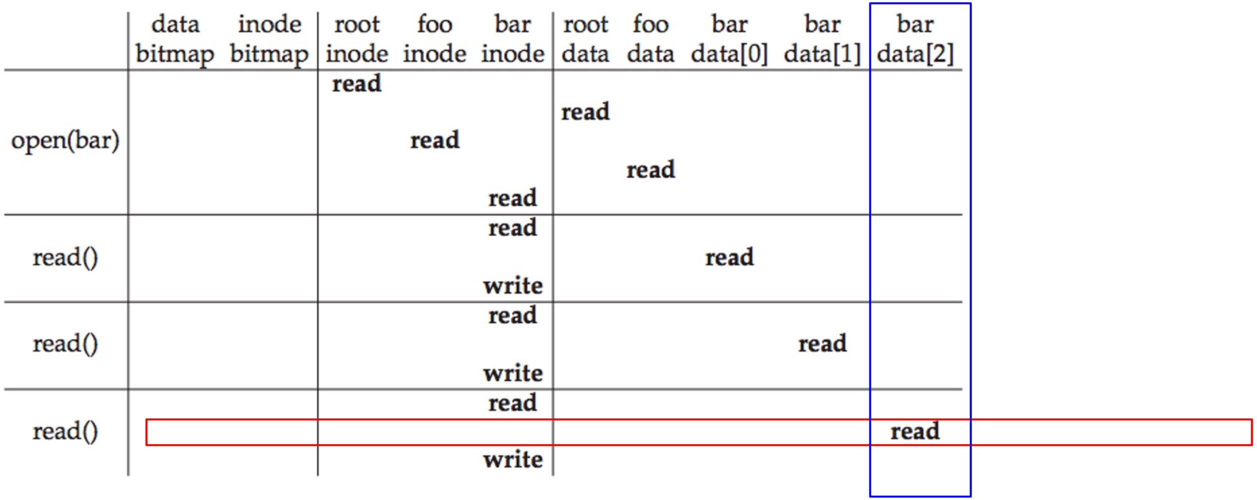
- We have to also write the last accessed time..



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

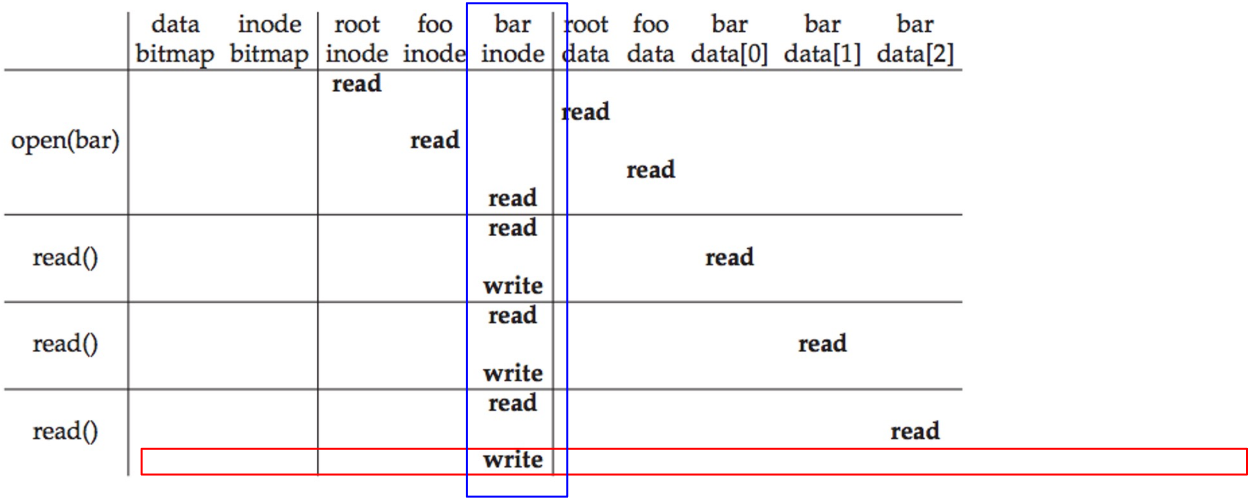
- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..



	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
open(bar)			read							
						read				
				read						
							read			
					read					
read()					read					
								read		
					write					
read()					read					
									read	
					write					
read()					read					
										read
					write					

Example:Reading

- Let’s try and read a file: /foo/bar
- We have to traverse directories and inodes
- We have to also write the last accessed time..

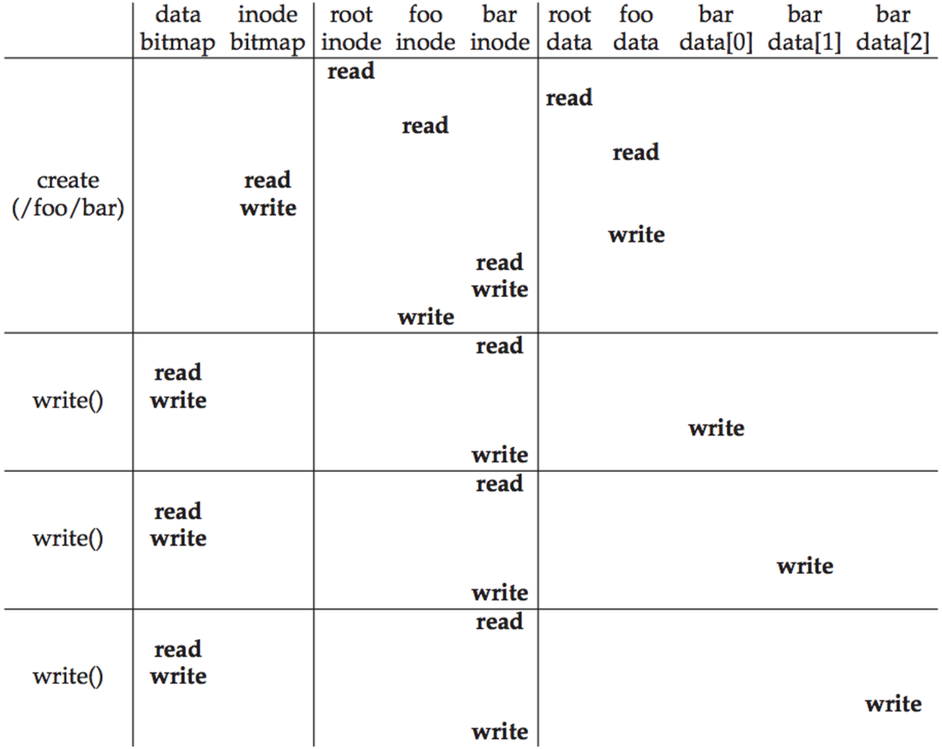


	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
			read							
						read				
open(bar)				read						
							read			
					read					
					read					
read()								read		
					write					
					read					
read()									read	
					write					
					read					
read()										read
					write					

Example: Creating/Writing a File

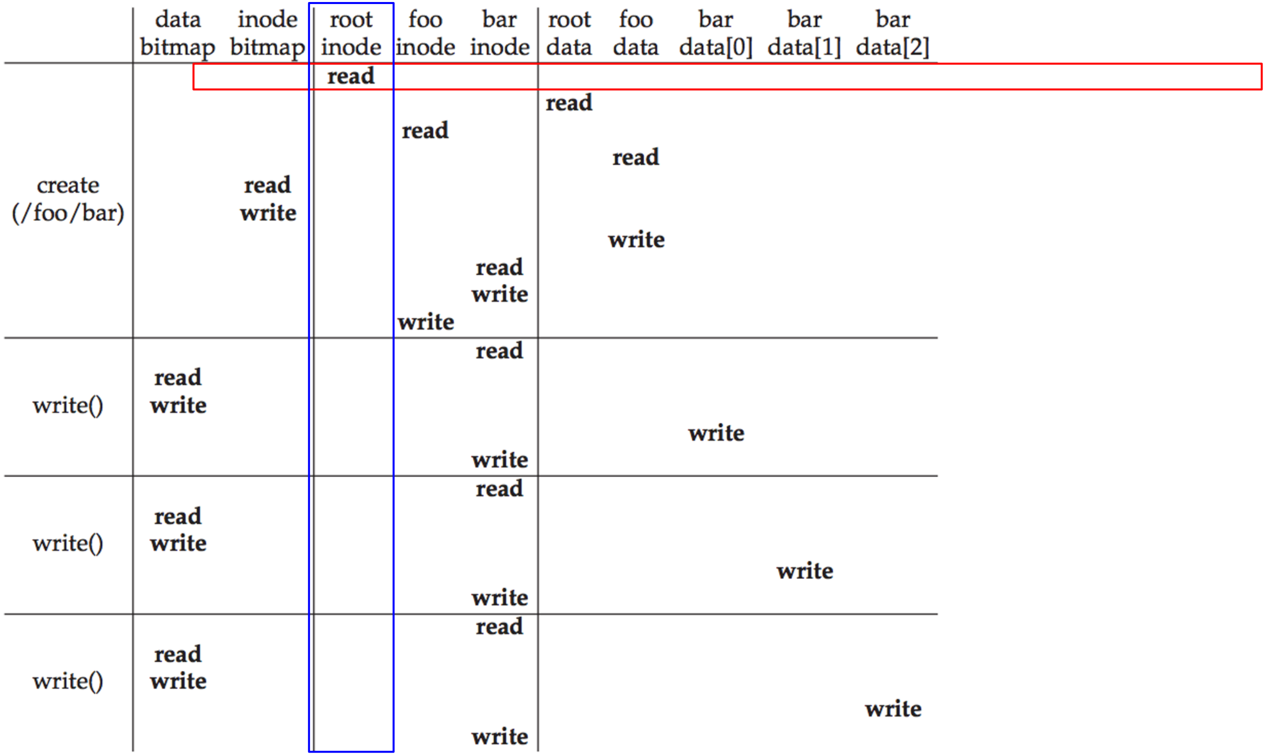
- When creating we must do lots of writes!
- We must write to the inode allocation bitmap and to the directory, etc. etc.
- We also must allocate data blocks for the file we want to write and update the inode with that mapping as we go.

Example: Creating/Writing a File



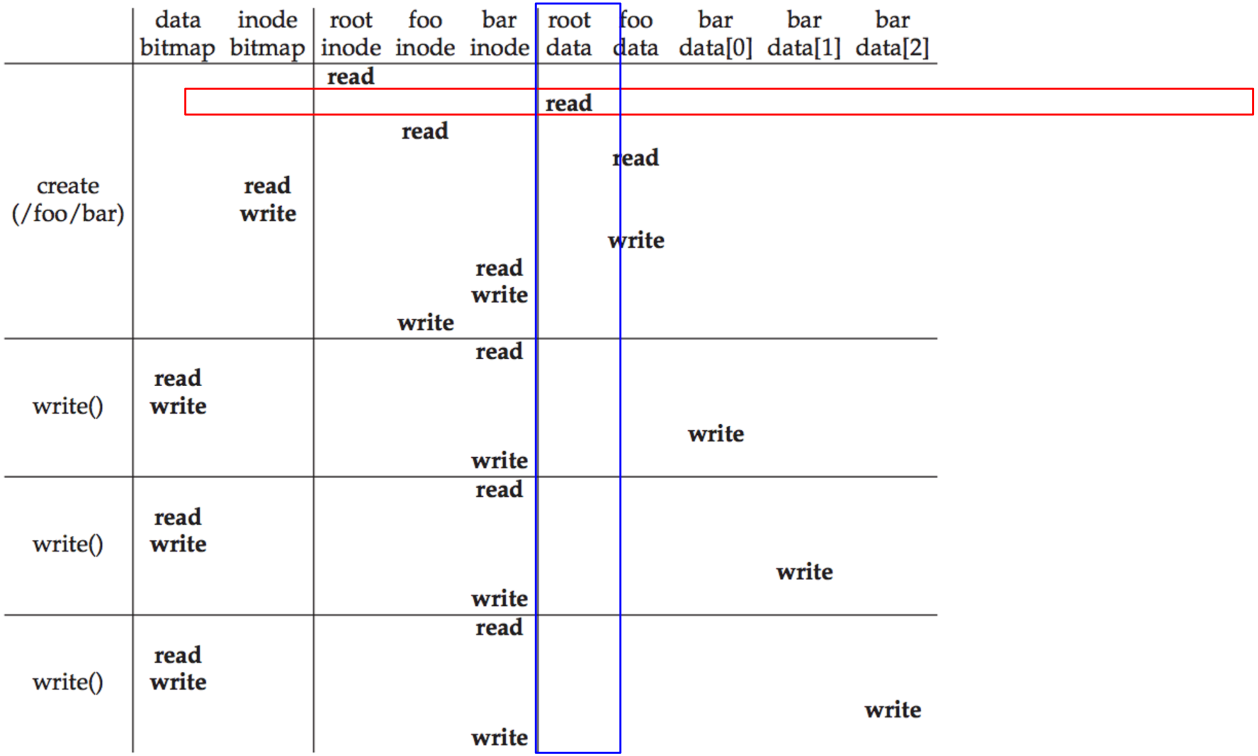
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



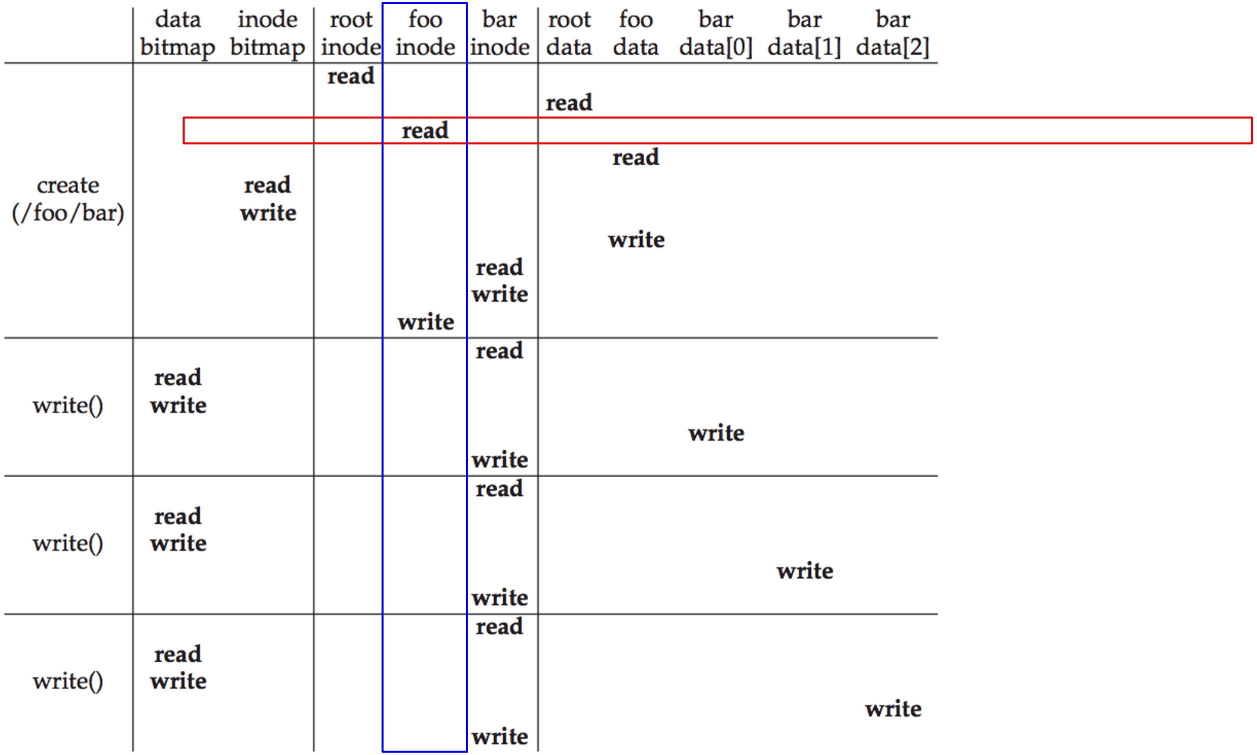
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



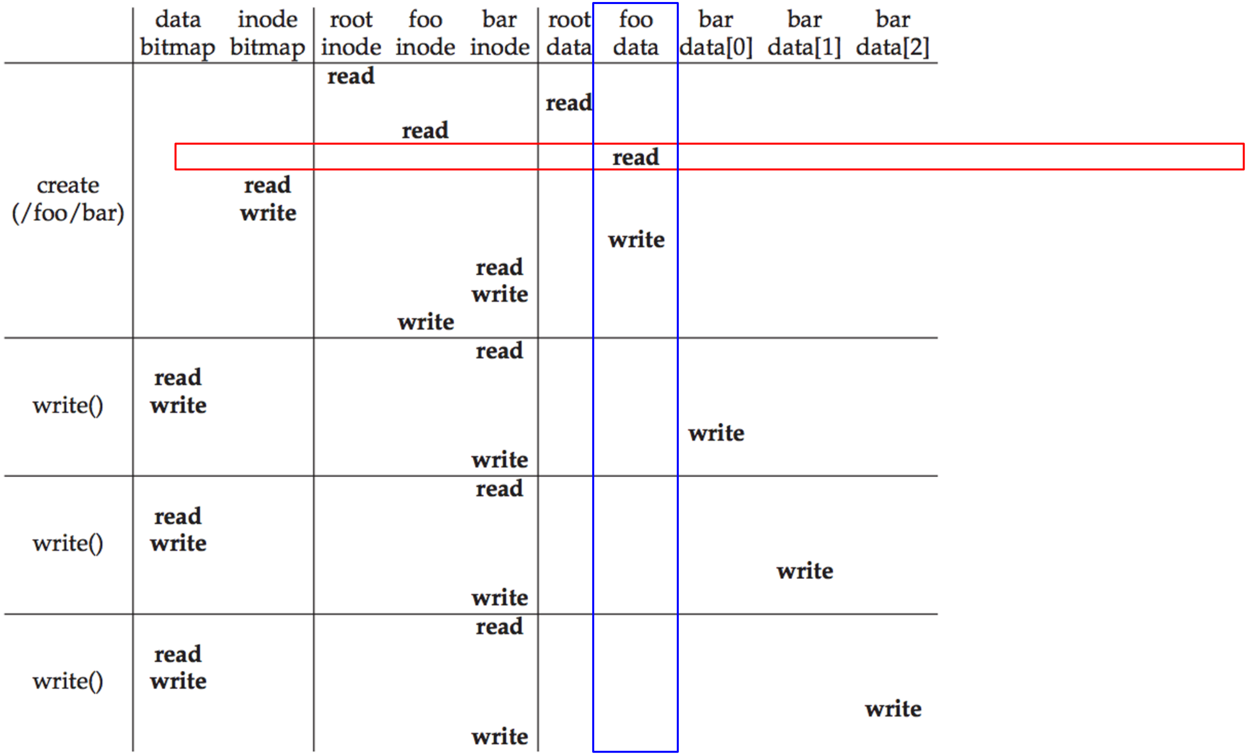
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



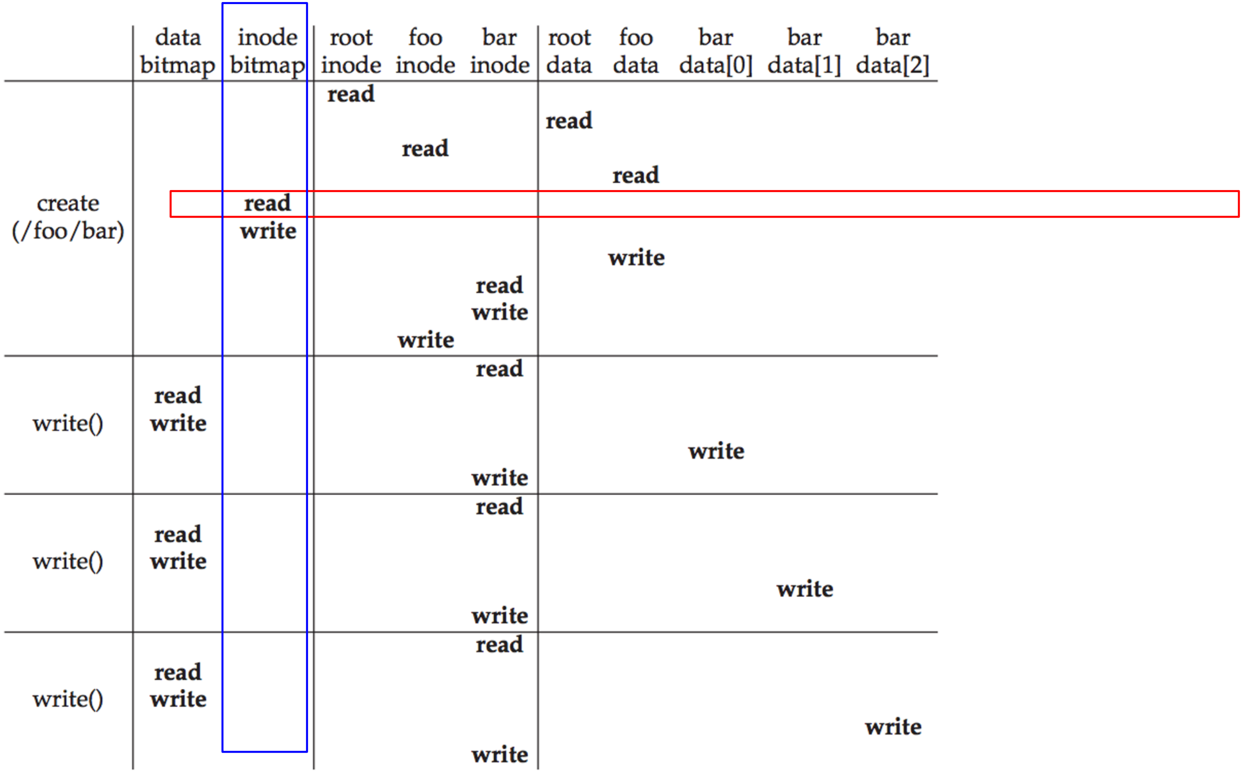
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



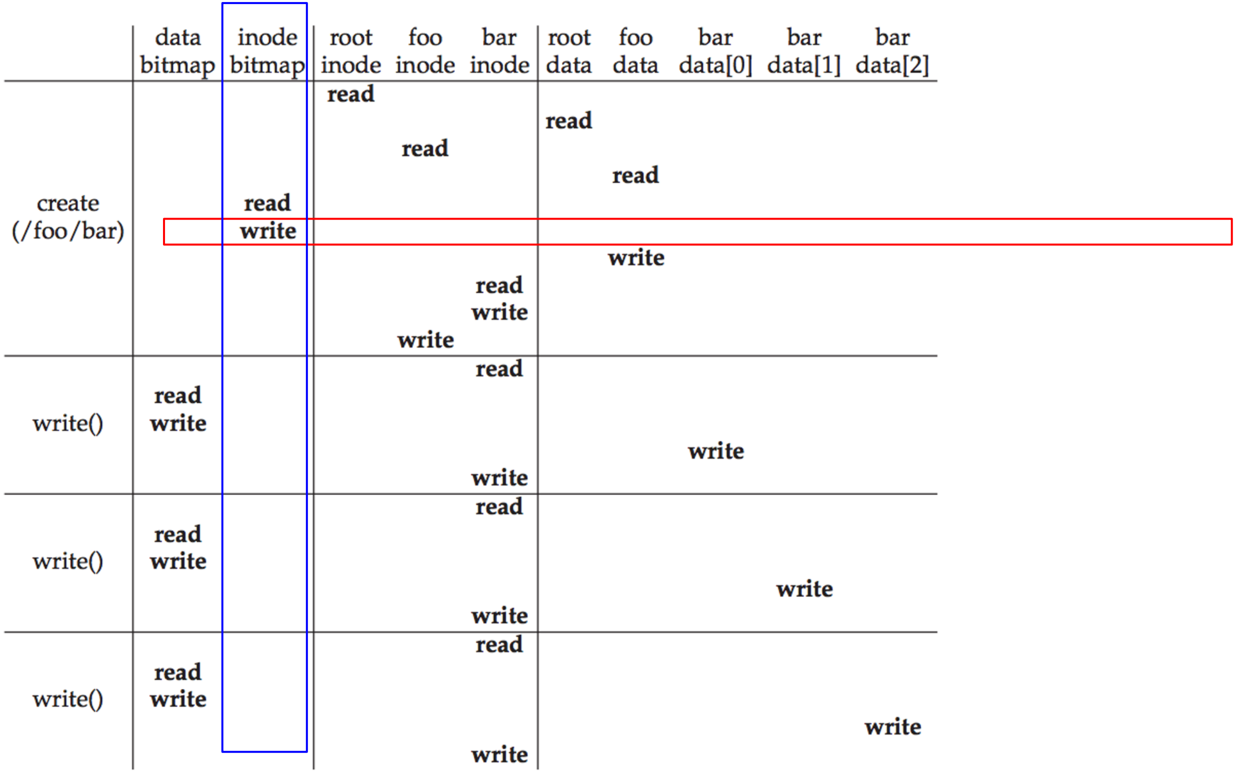
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



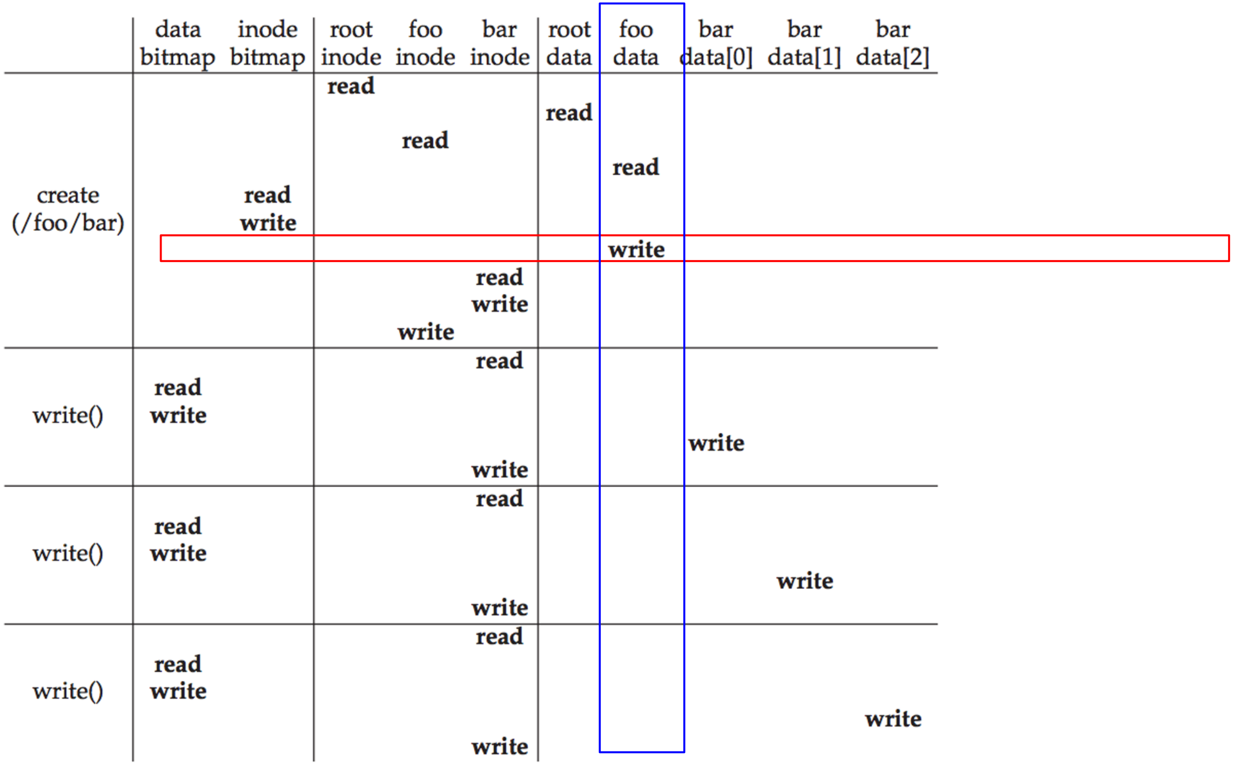
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



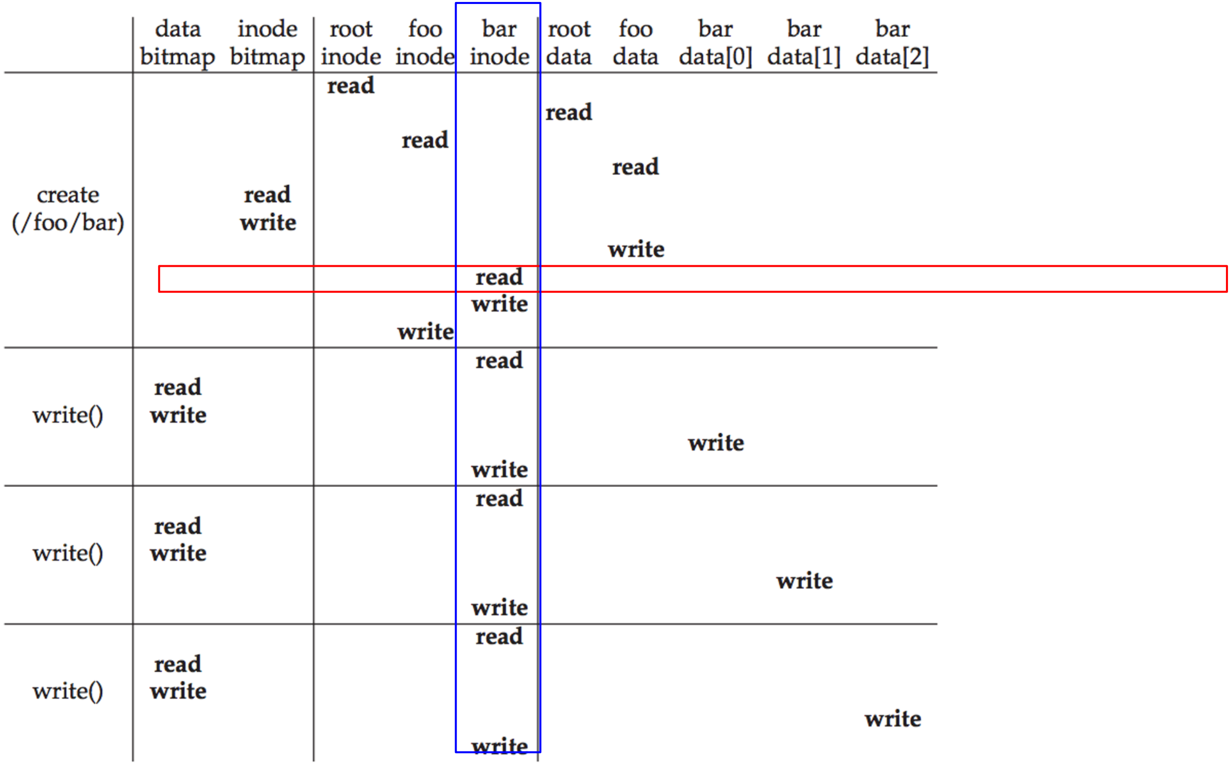
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



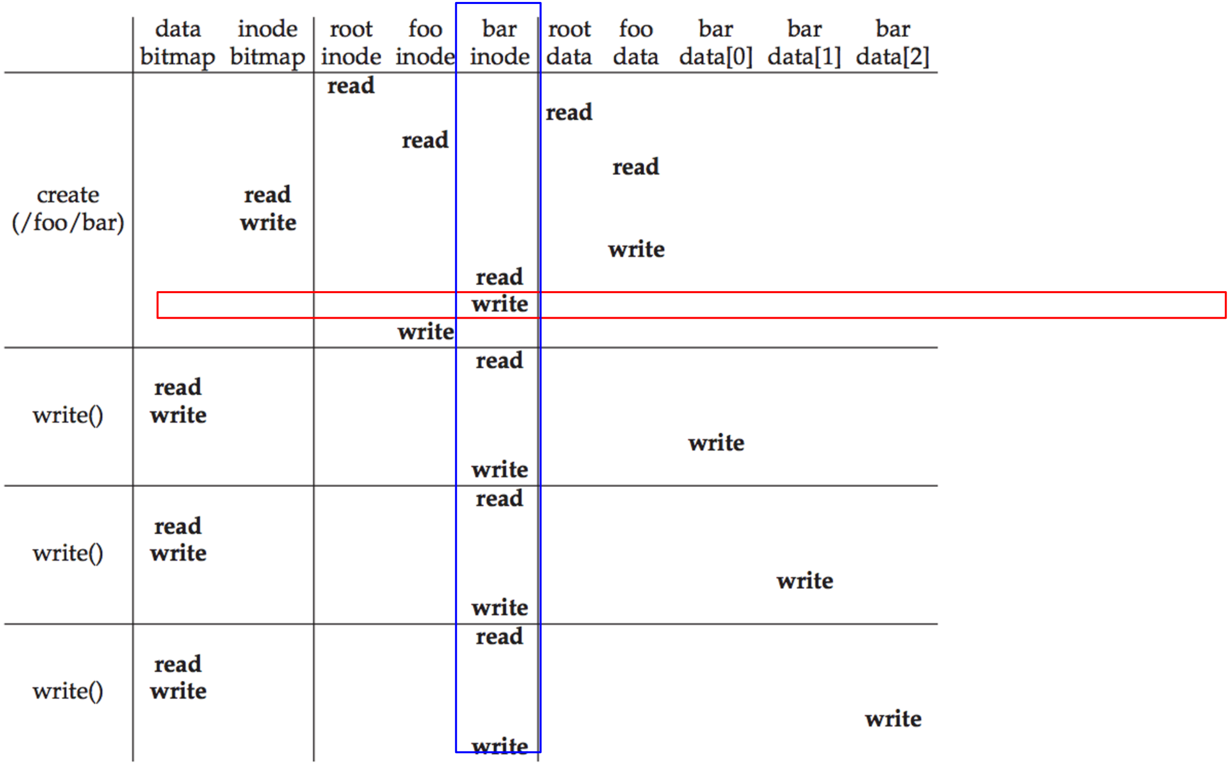
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



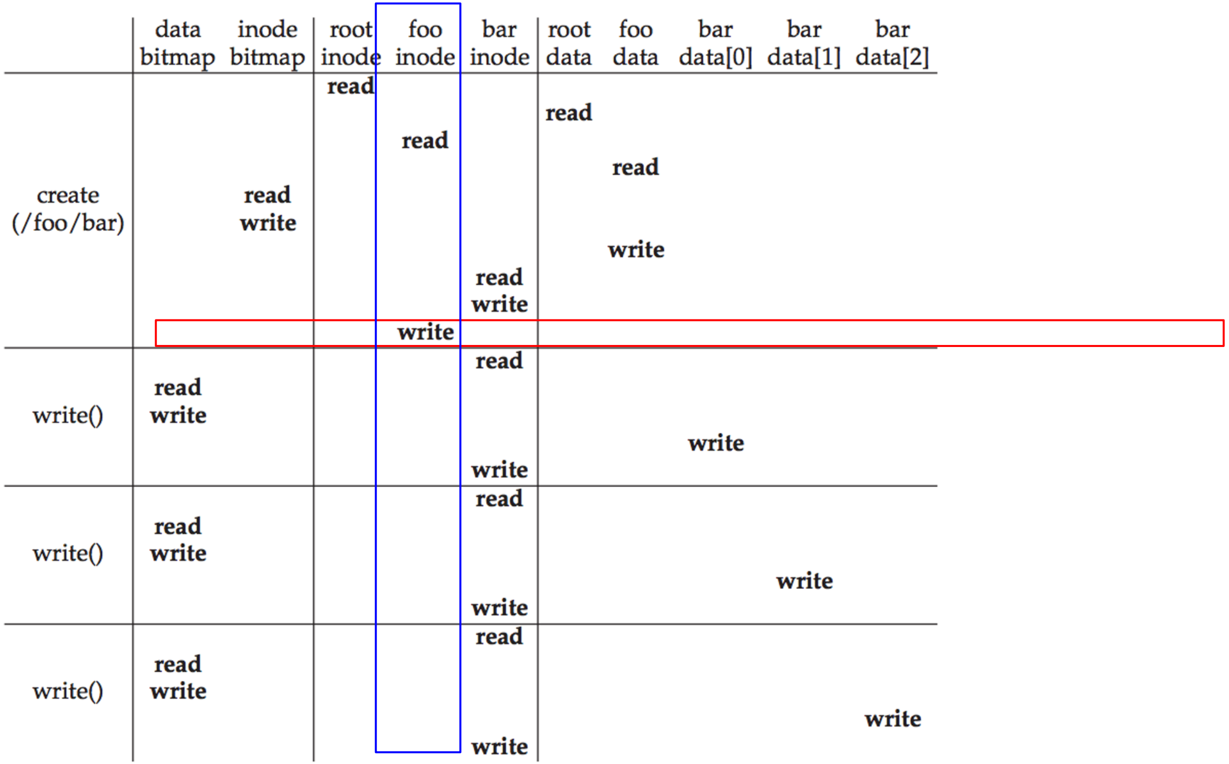
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



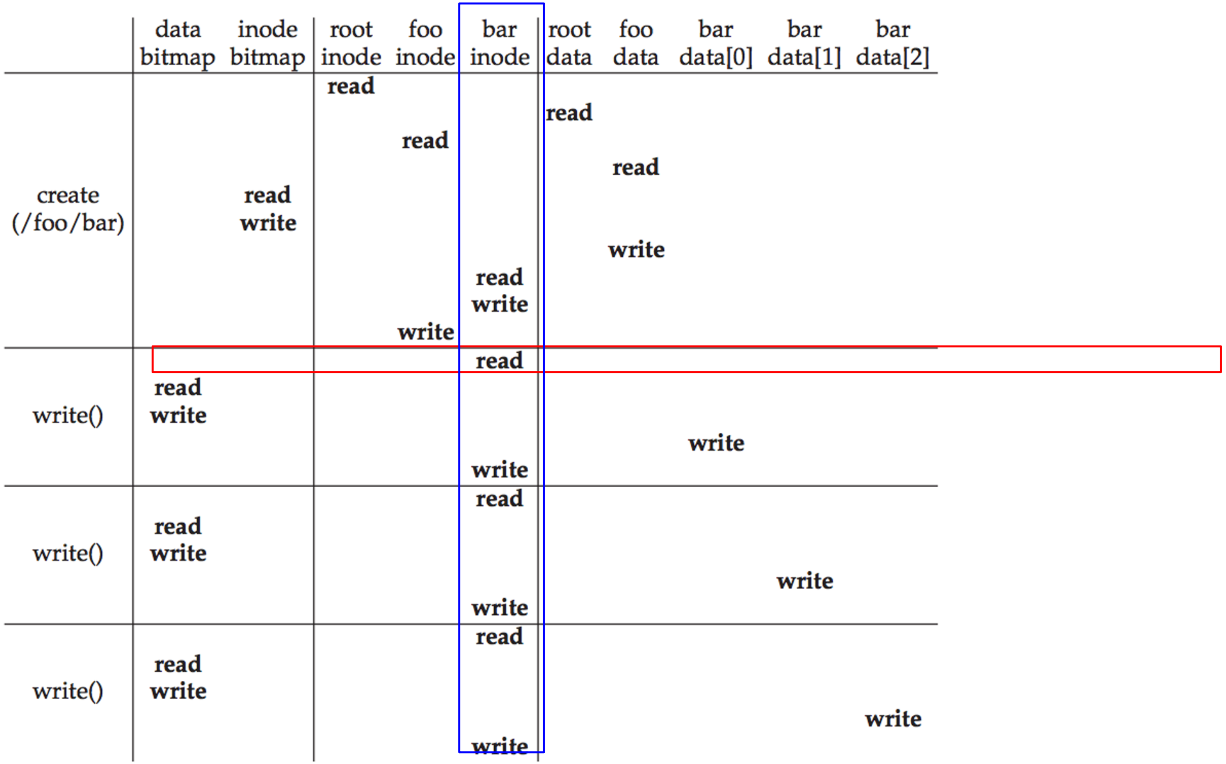
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



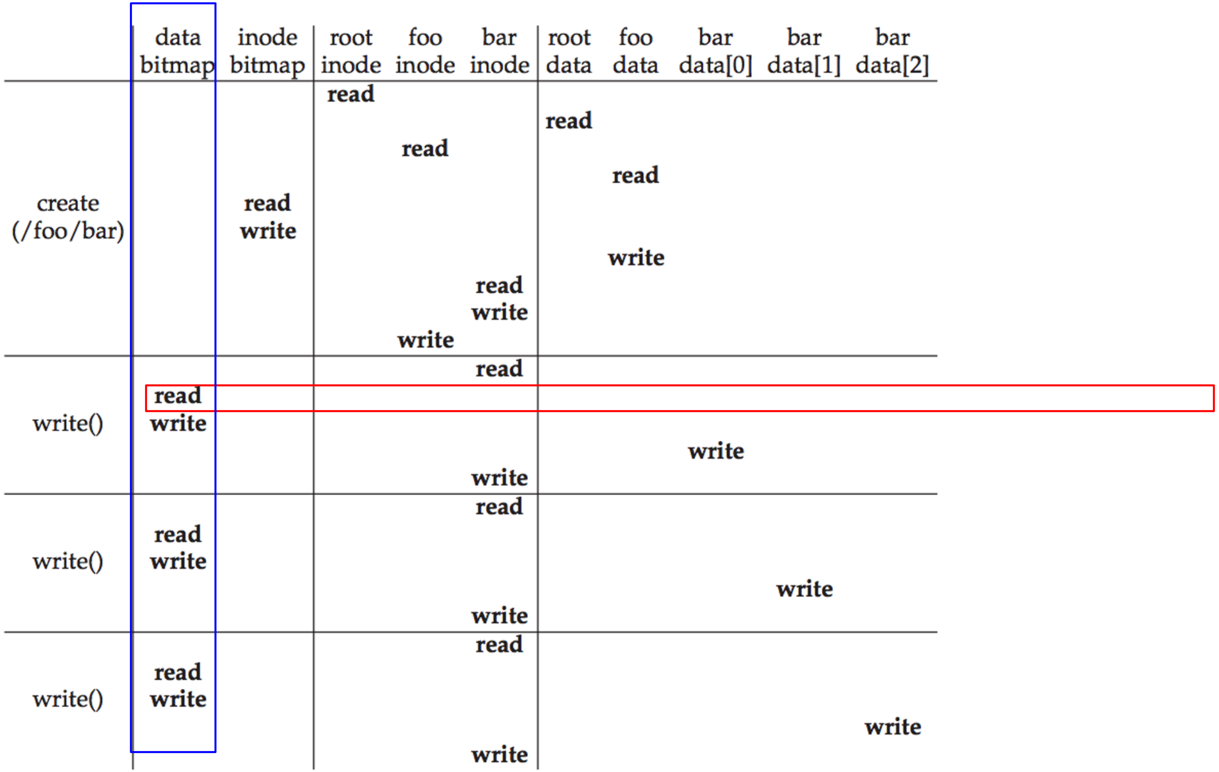
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



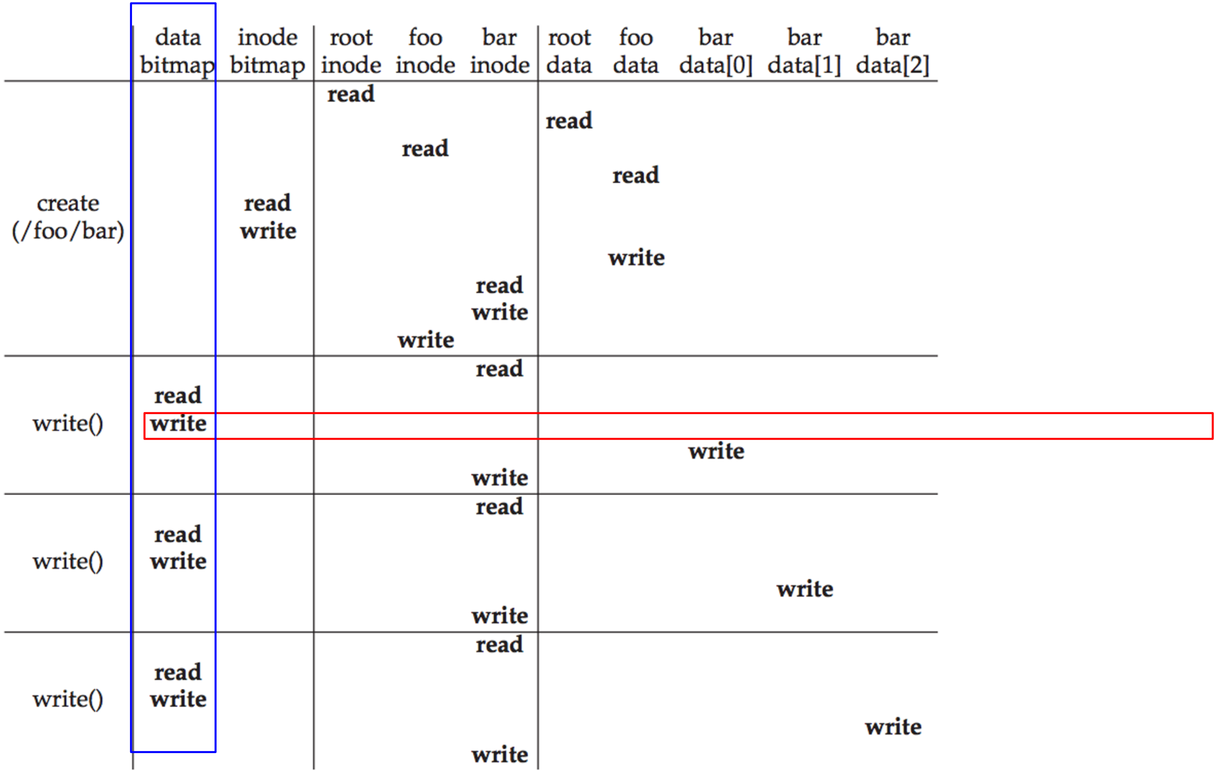
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



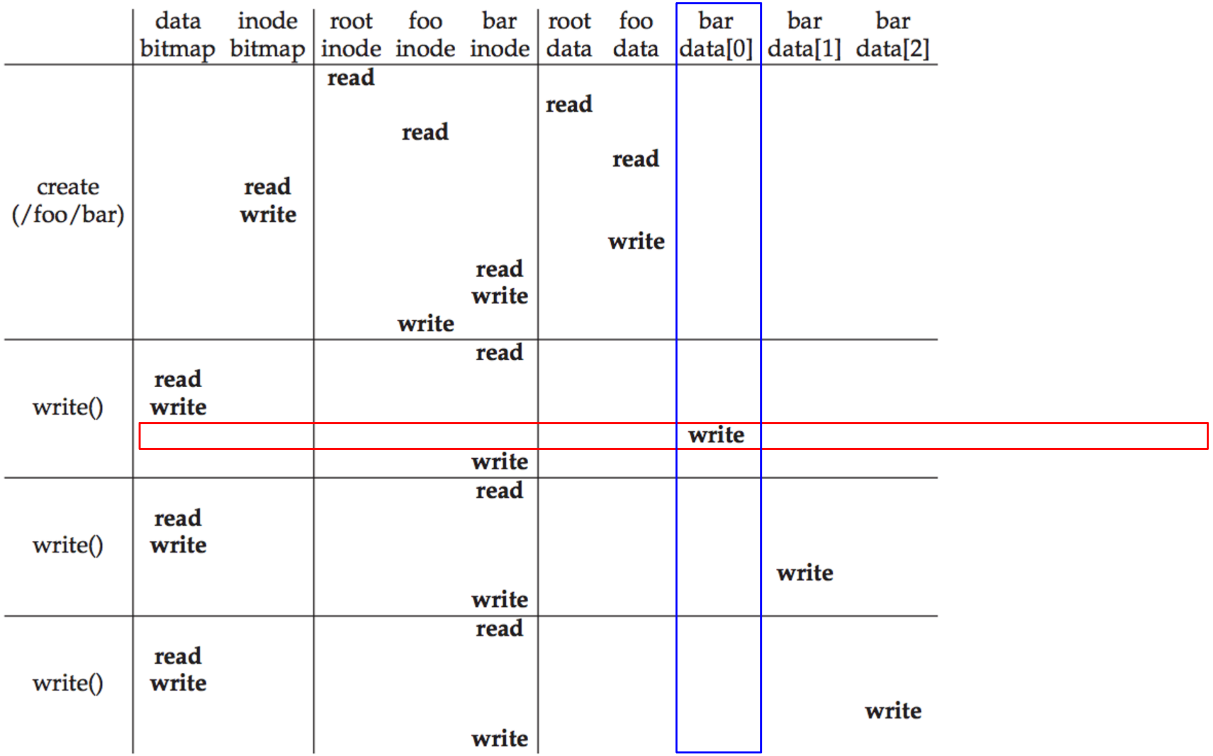
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



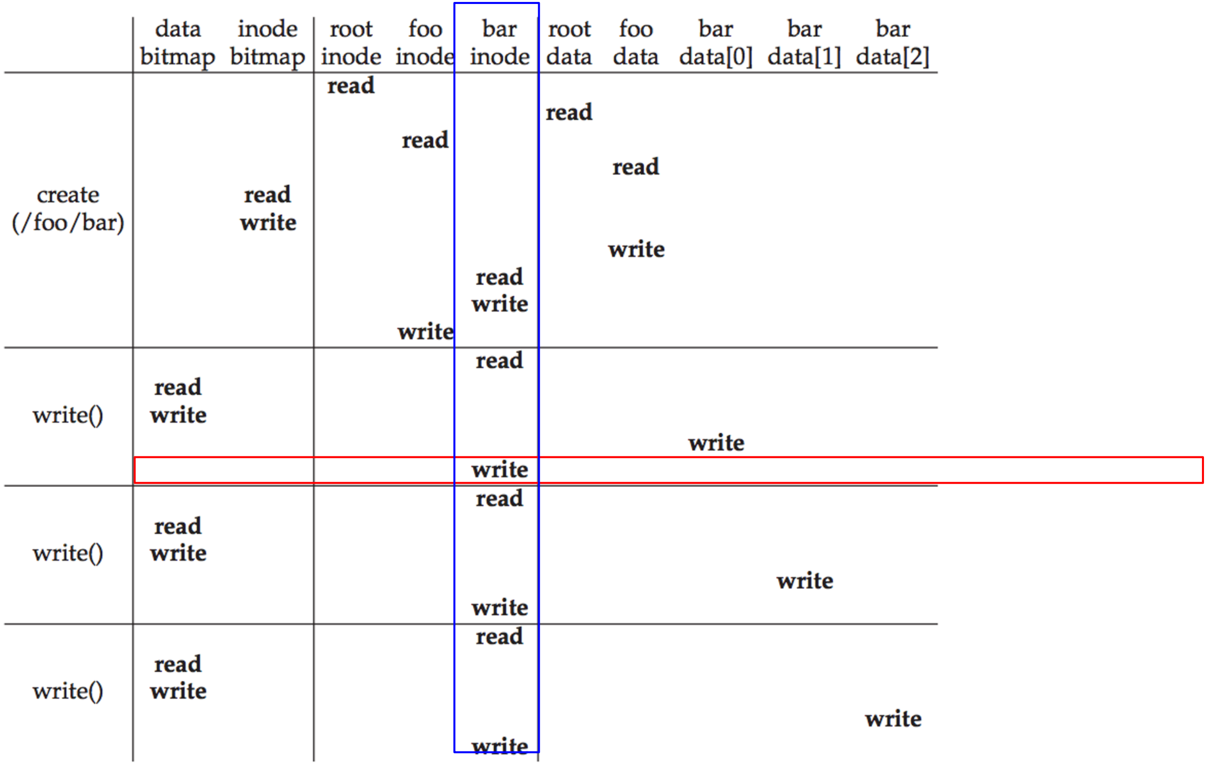
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



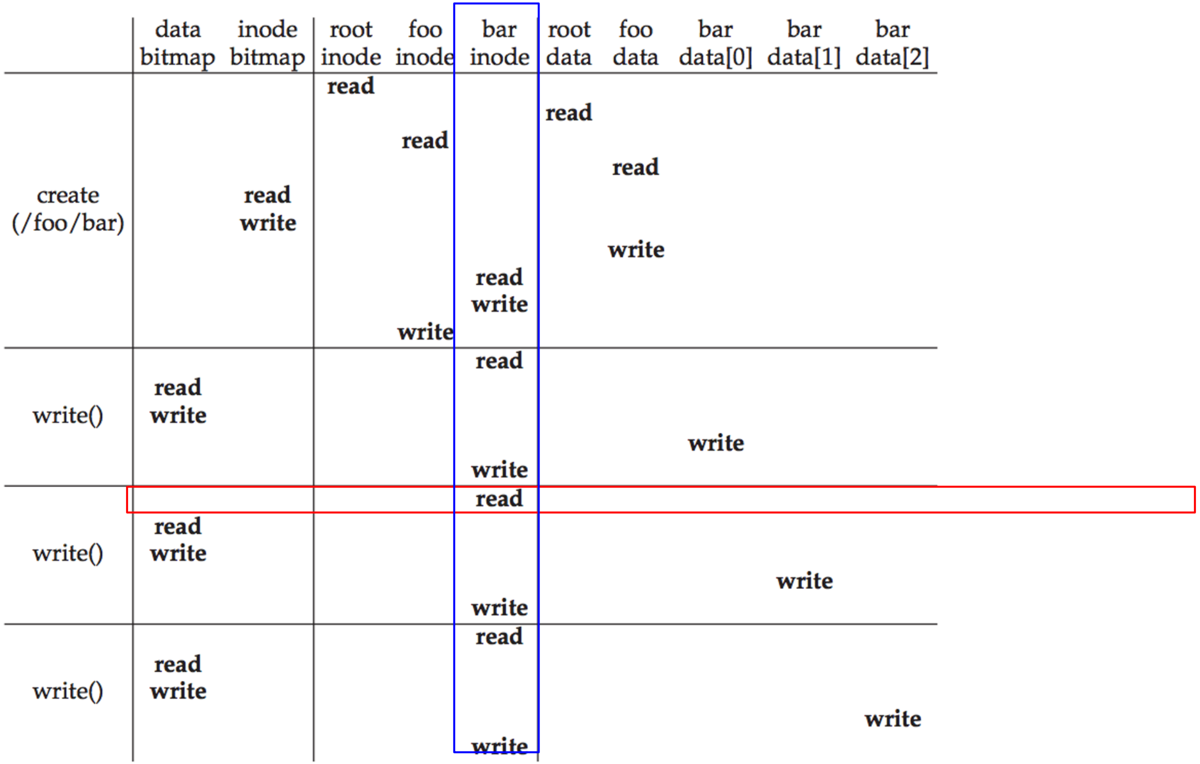
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



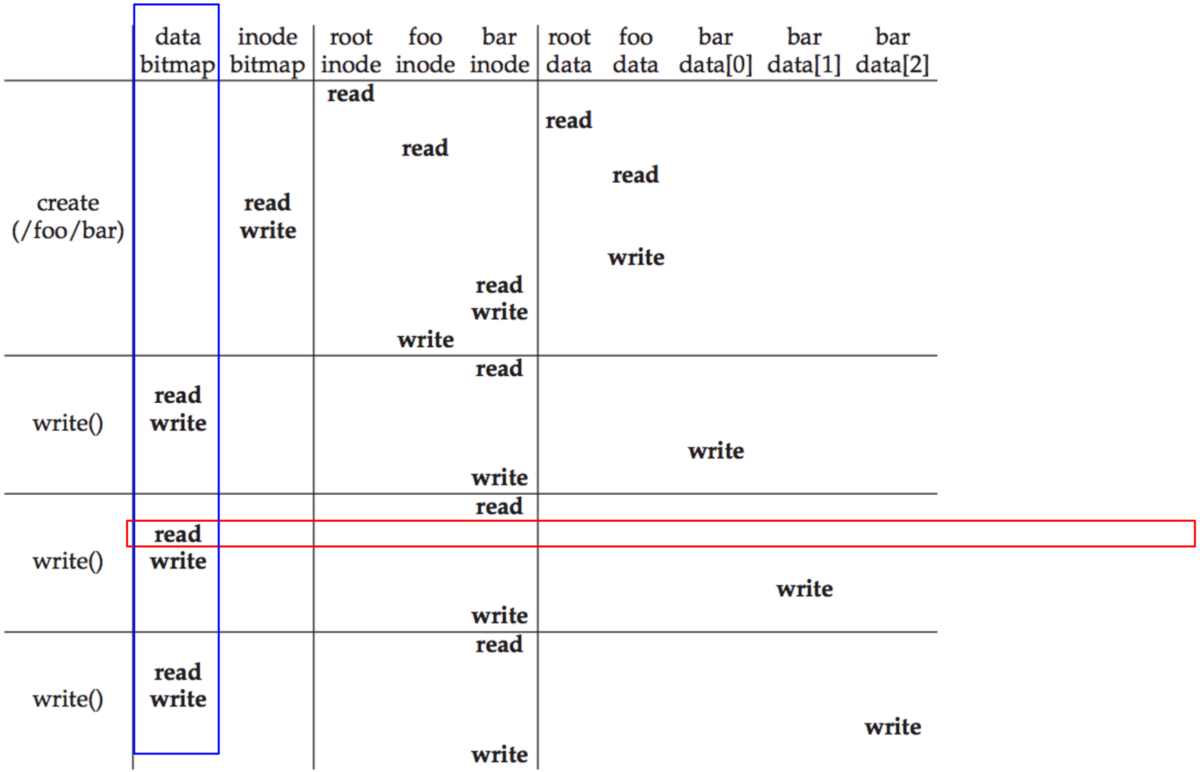
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



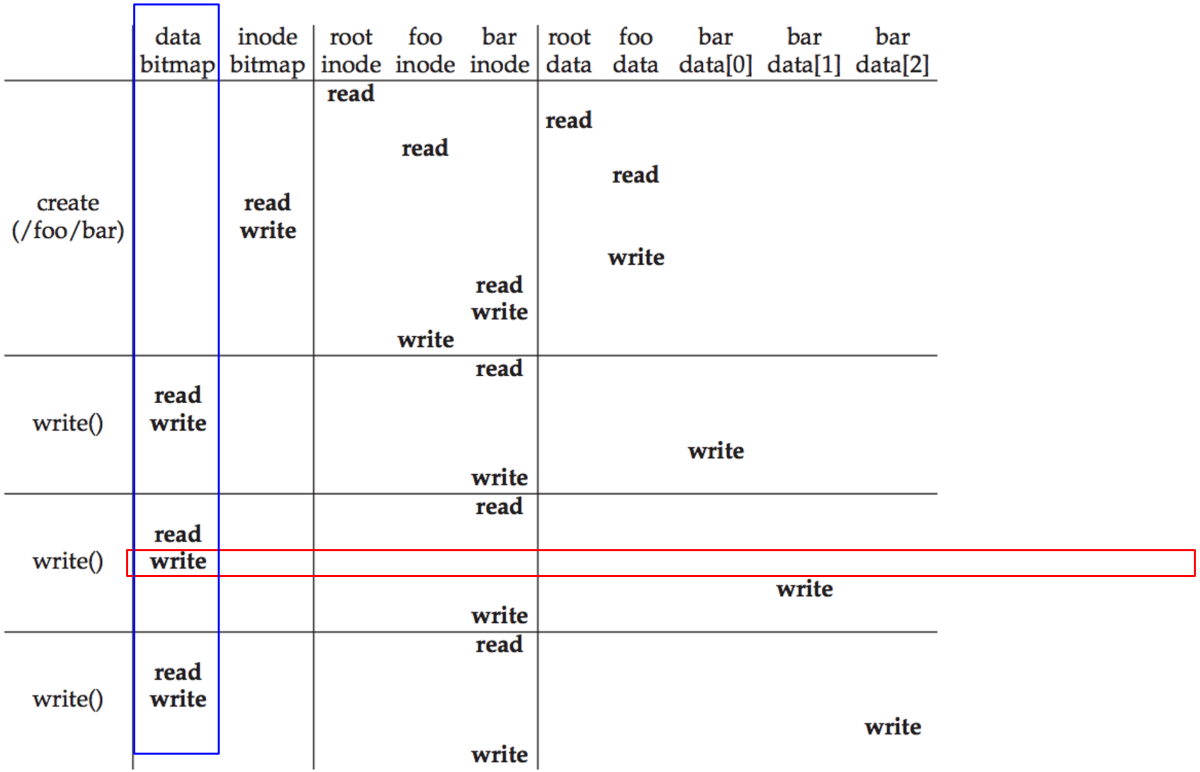
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



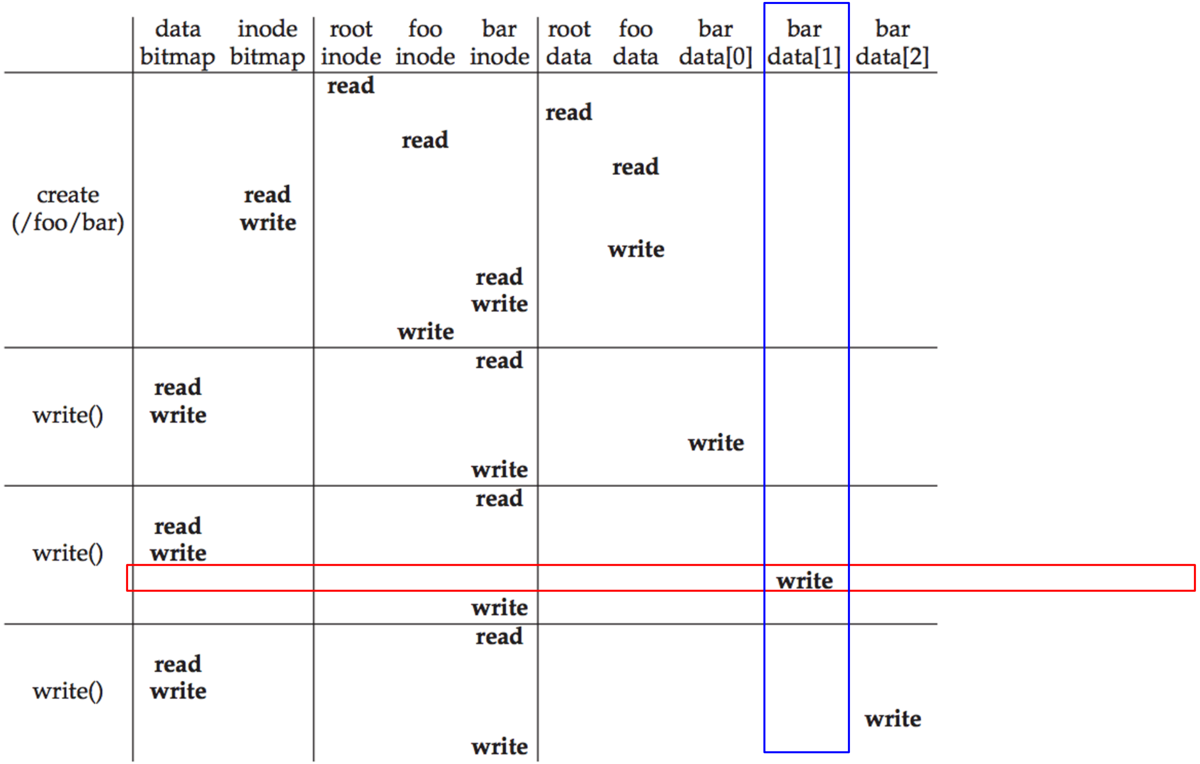
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



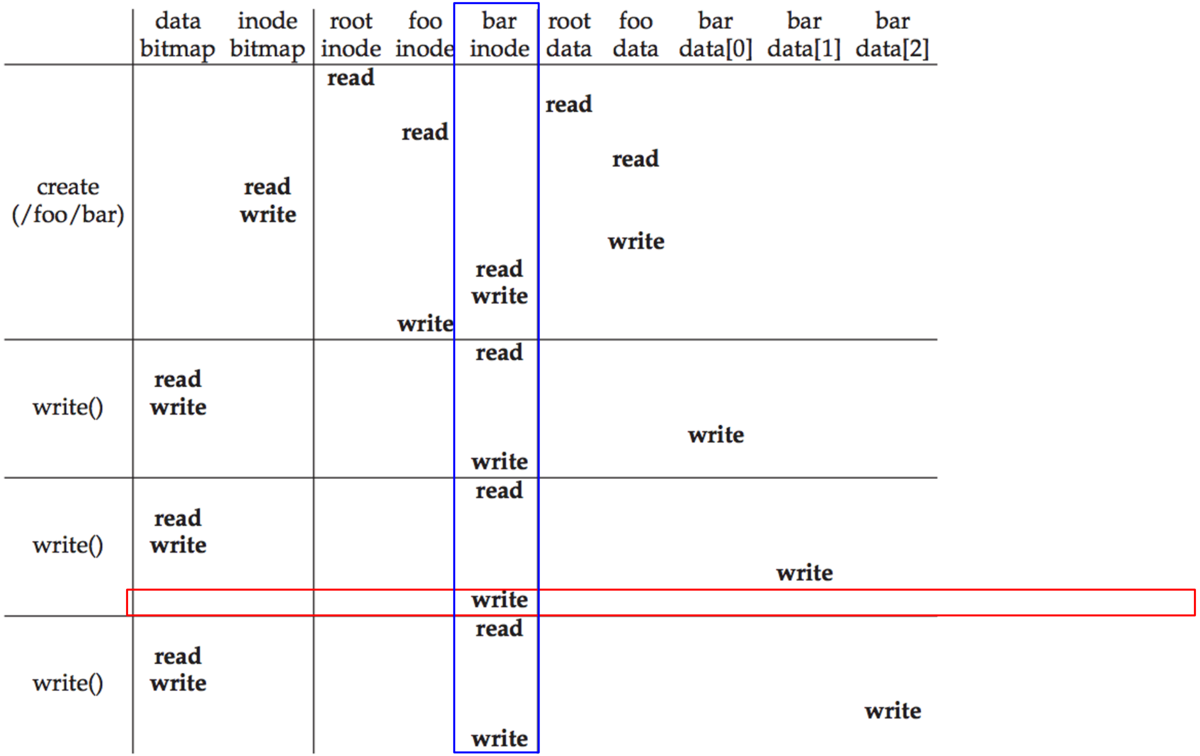
	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



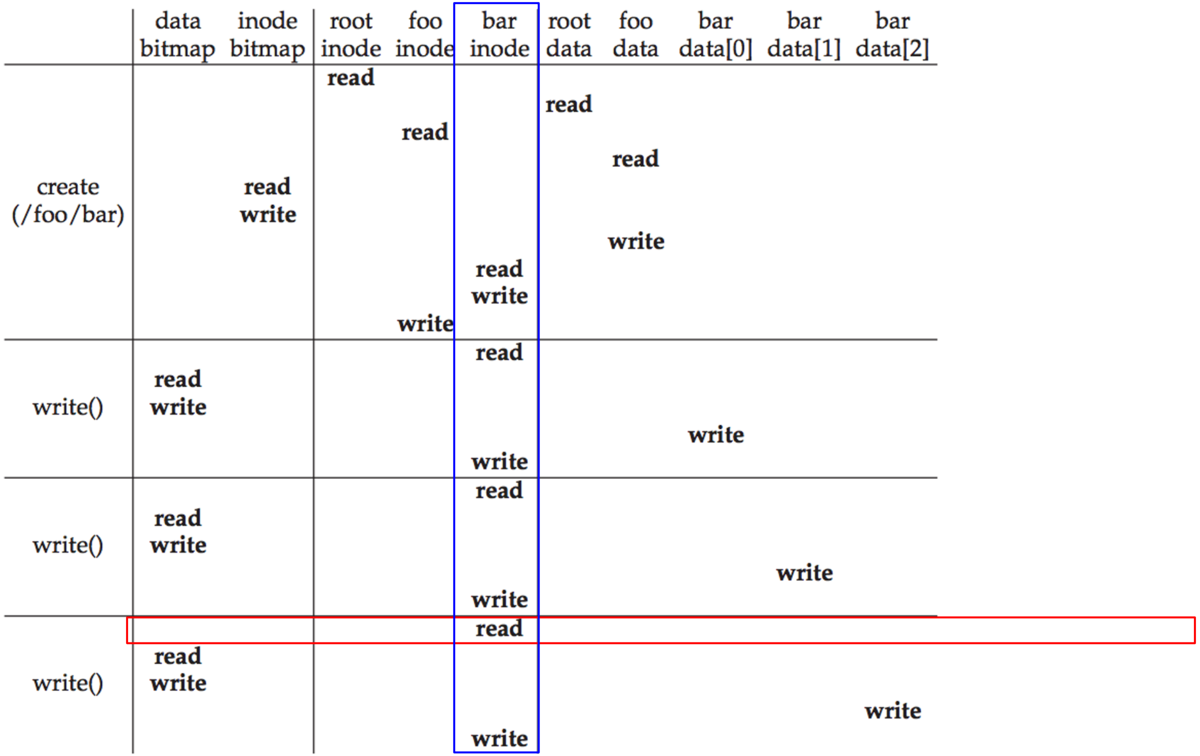
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



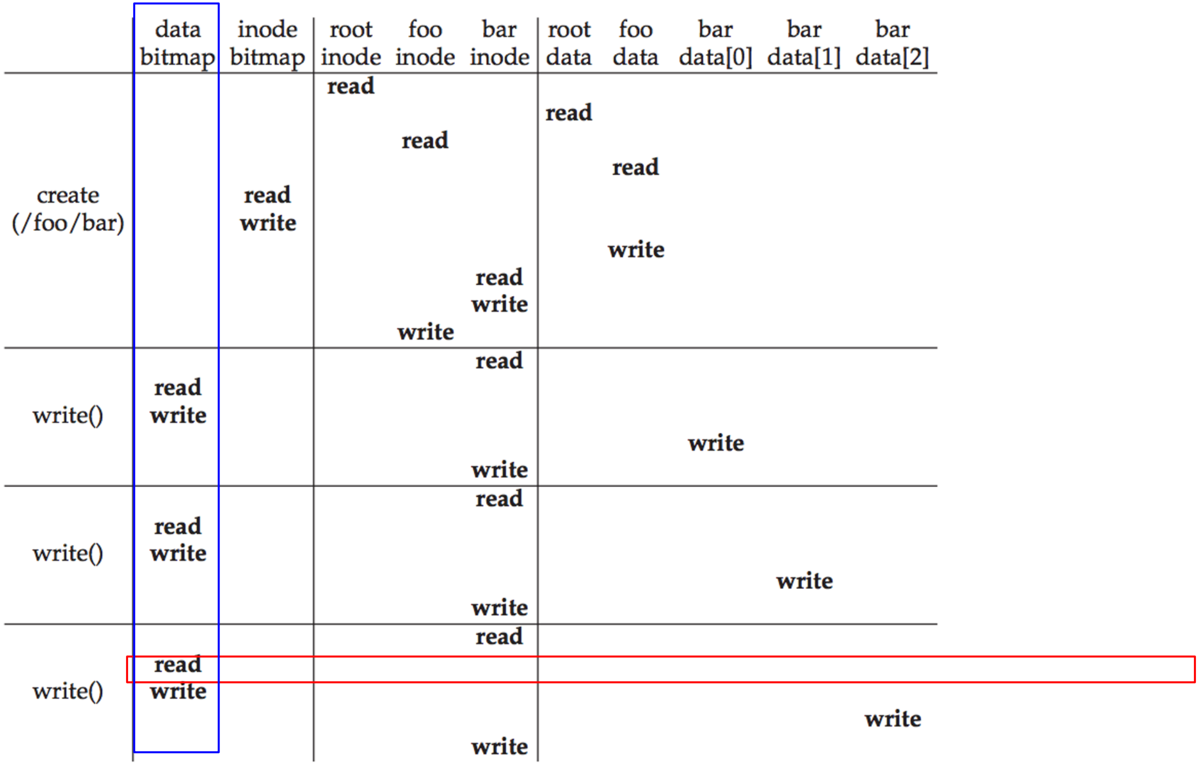
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



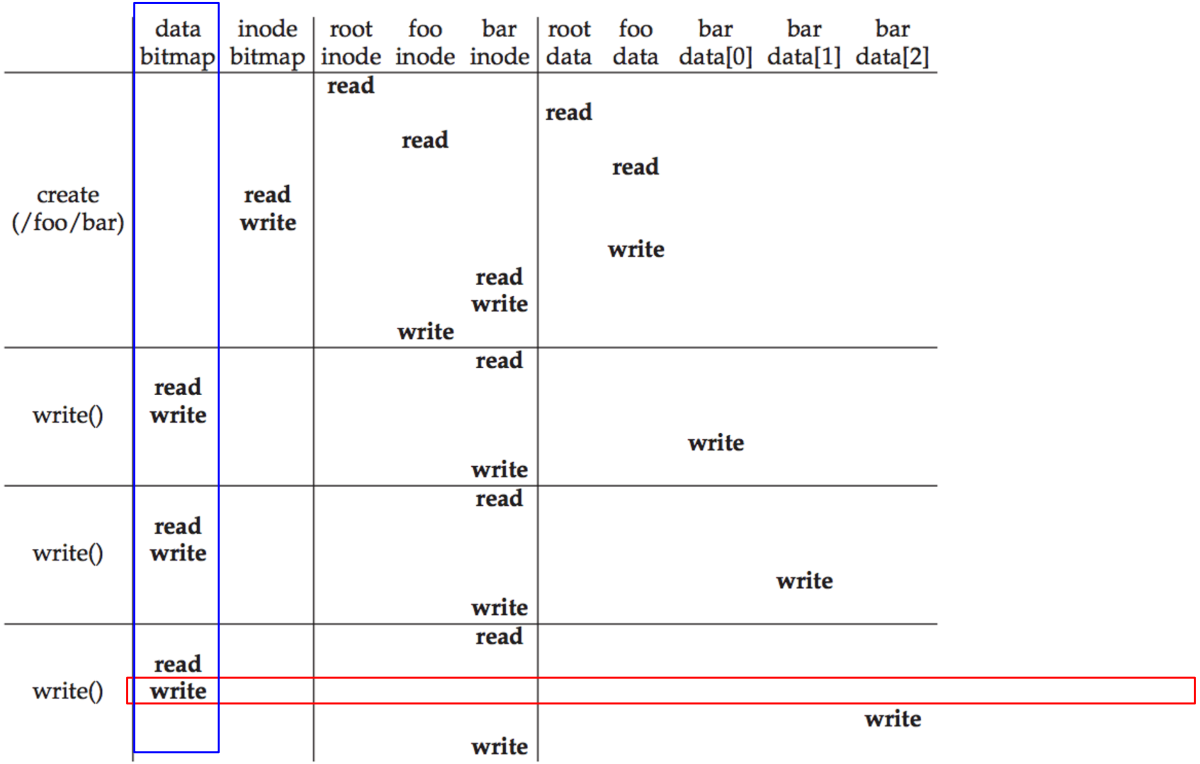
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



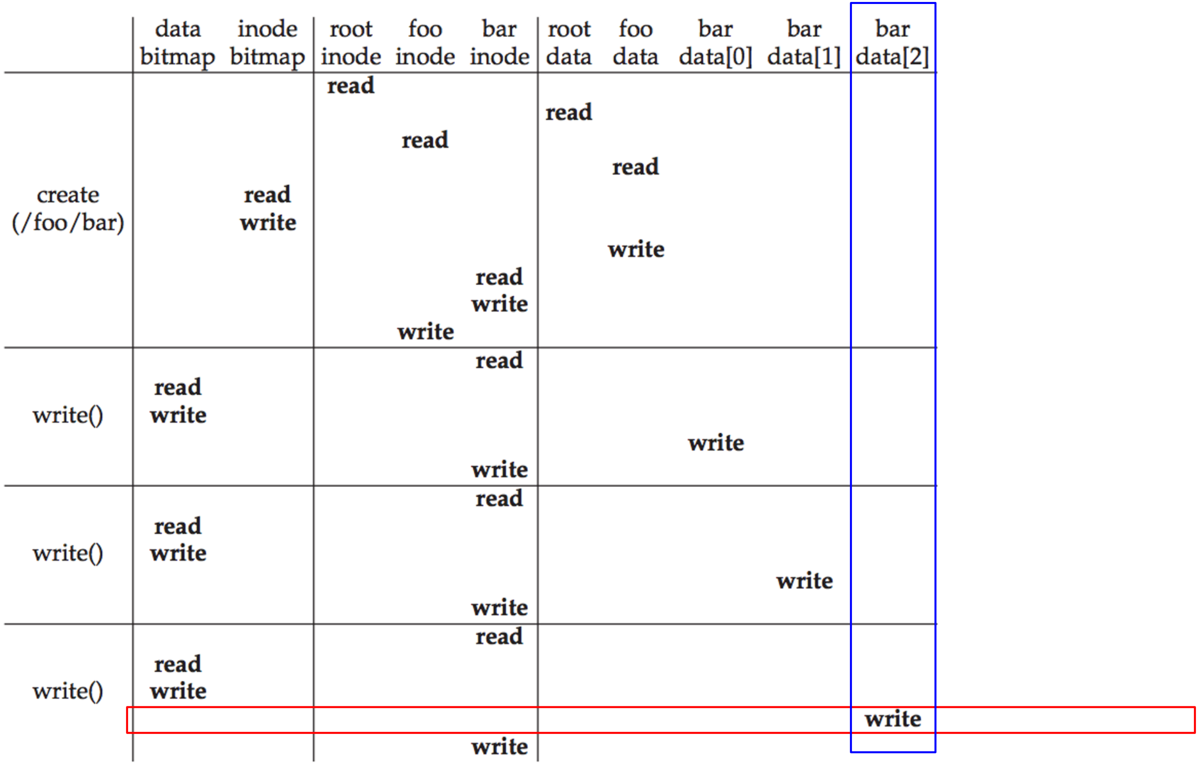
	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



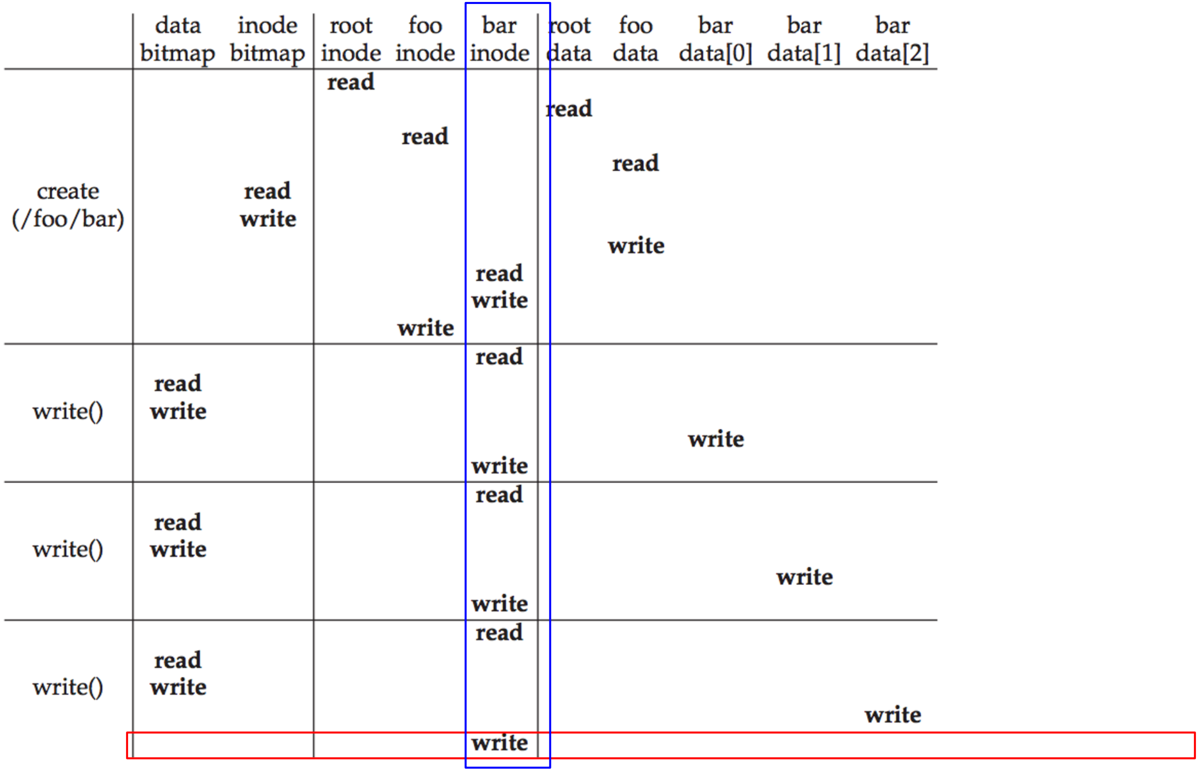
	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



	data	inode	root	foo	bar	root	foo	bar	bar	bar
	bitmap	bitmap	inode	inode	inode	data	data	data[0]	data[1]	data[2]
create (/foo/bar)			**read**							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					**read**					
	read									
	write									
								write		
					write					
write()					**read**					
	read									
	write									
									write	
					write					
write()					**read**					
	read									
	write									
										write
					write					

Example: Creating/Writing a File



	data bitmap	inode bitmap	root inode	foo inode	bar inode	root data	foo data	bar data[0]	bar data[1]	bar data[2]
create (/foo/bar)			read							
						read				
				read						
							read			
		read								
		write								
							write			
					read					
					write					
				write						
write()					read					
	read									
	write									
								write		
					write					
write()					read					
	read									
	write									
									write	
					write					
write()					read					
	read									
	write									
										write
					write					

🤪

That’s crazy.

🤪

That’s crazy.

Please tell me you know the fix.

🛠️

Caching

- Most performance problems can be solved with caching….
- Modern OSs have a “unified page cache” that caches blocks from the file system in a cache with memory pages
- So, the first read of a directory may be slow, but subsequent ones are *really* fast.

Buffering

- Delay work in the hopes it goes away.
- So, if we hold all the writes in memory, we can consolidate them (also good for disk scheduling)
- Example: last accessed in the inode + modifying the data blocks all becomes one write
- Tradeoff: your writes may not be on disk after a crash

The End