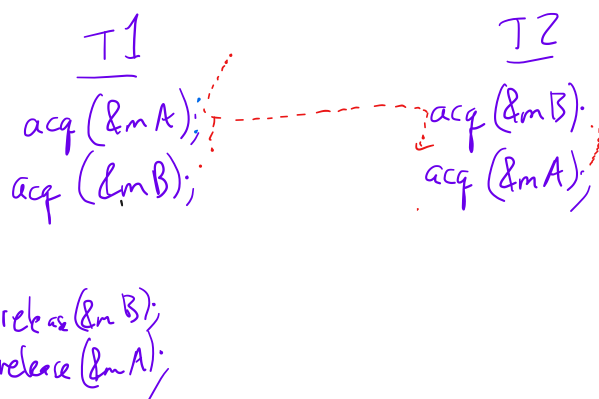


- 1. Last time
- 2. Deadlock, continued
- 3. Other progress issues
- 4. Performance issues
- 5. Programmability issues
- 6. Mutexes and interleaving

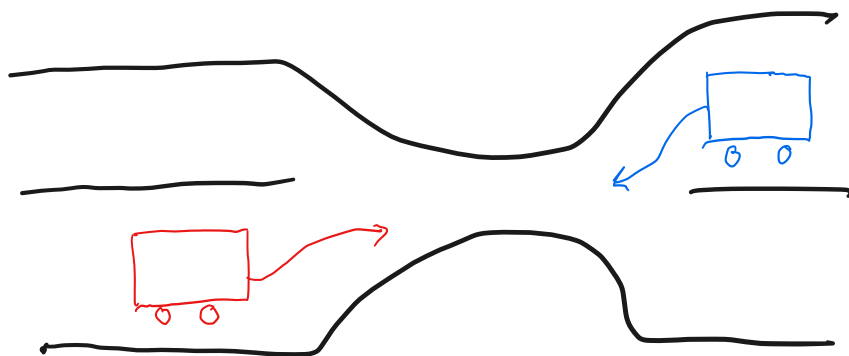
ONE HANDOUT

2. Deadlock



Happens when all four of these conditions are present:

- i. mutual exclusion
- ii. hold and wait
- iii. no pre-emption
- iv. circular wait



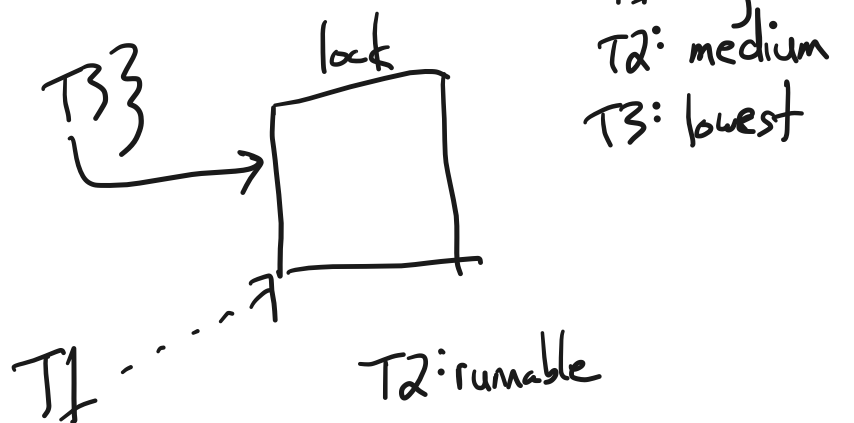
What can we do about deadlock?

- (a) ignore it
- (b) detect + recover
- (c) avoid algorithmically
- (d) negate one of the 4 conditions
- (e) static/dynamic detection tools

3. Other progress issues

- Starvation

- Priority inversion



Assume: highest-prio runnable thread runs.

4. Performance issues and tradeoffs

(a) Invoking spinlocks/mutexes can be expensive

MCS locks

futexes

(b) Coarse locks limit available parallelism...

(c) ... but fine-grained locking leads to complexity and hence bugs

5. Programmability issues

- loss of modularity

`printf("...");`

- what's the fundamental problem?

6. mutexes and interleavings

ops

mem barrier

ops

≡ A

release() < fence
B

"A stuff" happens before "B stuff"
although, within A and B, the "stuff"
can happen out of order from the perspective
of other cores, assuming they could view the
operations separately.

Feb 19, 25 9:01

handout07.txt

Page 1/3

```

1 CS 202
2 Handout 7 (Class 8)
3
4 1. Simple deadlock example
5
6     T1:
7         acquire(mutexA);
8         acquire(mutexB);
9
10        // do some stuff
11
12        release(mutexB);
13        release(mutexA);
14
15     T2:
16         acquire(mutexB);
17         acquire(mutexA);
18
19        // do some stuff
20
21        release(mutexA);
22        release(mutexB);
23

```

Feb 19, 25 9:01

handout07.txt

Page 2/3

```

24 2. More subtle deadlock example
25
26     Let M be a monitor (shared object with methods protected by mutex)
27     Let N be another monitor
28
29     class M {
30     private:
31         Mutex mutex_m;
32
33         // instance of monitor N
34         N another_monitor;
35
36         // Assumption: no other objects in the system hold a pointer
37         // to our "another_monitor"
38
39     public:
40         M();
41         ~M();
42         void methodA();
43         void methodB();
44     };
45
46     class N {
47     private:
48         Mutex mutex_n;
49         Cond cond_n;
50         int navailable;
51
52     public:
53         N();
54         ~N();
55         void* alloc(int nwanted);
56         void free(void*);
57     }
58
59     int
60     N::alloc(int nwanted) {
61         acquire(&mutex_n);
62         while (navailable < nwanted) {
63             wait(&cond_n, &mutex_n);
64         }
65
66         // peel off the memory
67
68         navailable -= nwanted;
69         release(&mutex_n);
70     }
71
72     void
73     N::free(void* returning_mem) {
74         acquire(&mutex_n);
75
76         // put the memory back
77
78         navailable += returning_mem;
79
80         broadcast(&cond_n, &mutex_n);
81
82         release(&mutex_n);
83     }
84
85

```

Feb 19, 25 9:01

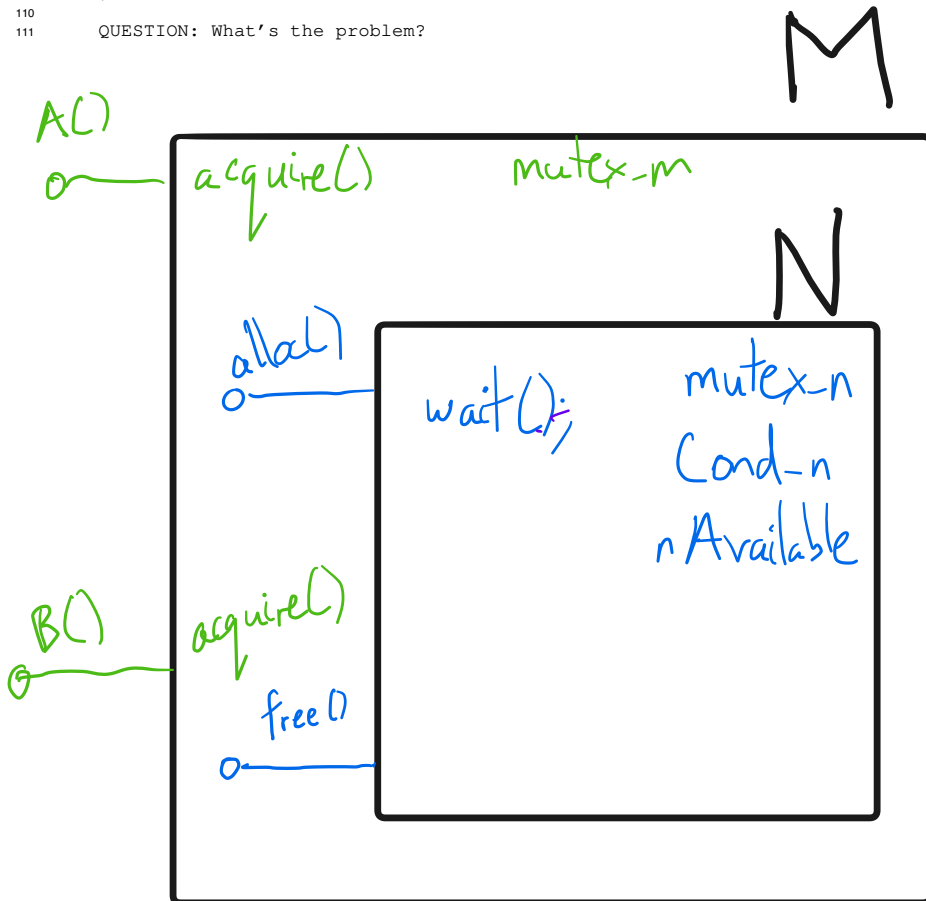
handout07.txt

Page 3/3

```

86 void
87 M::methodA() {
88
89     acquire(&mutex_m);
90
91     void* new_mem = another_monitor.alloc(int nbytes);
92
93     // do a bunch of stuff using this nice
94     // chunk of memory n allocated for us
95
96     release(&mutex_m);
97 }
98
99 void
100 M::methodB() {
101
102     acquire(&mutex_m);
103
104     // do a bunch of stuff
105
106     another_monitor.free(some_pointer);
107
108     release(&mutex_m);
109 }
110
111 QUESTION: What's the problem?

```



Feb 19, 25 9:07

filemap.txt

Page 1/2

```

1 2. Locking brings a performance vs. complexity trade-off
2
3 // SPDX-License-Identifier: GPL-2.0-only
4 /*
5  *      linux/mm/filemap.c
6  *
7  * Copyright (C) 1994-1999 Linus Torvalds
8  */
9
10 /*
11  * This file handles the generic file mmap semantics used by
12  * most "normal" filesystems (but you don't /have/ to use this:
13  * the NFS filesystem used to do this differently, for example)
14  */
15 #include <linux/export.h>
16 #include <linux/compiler.h>
17 #include <linux/dax.h>
18 #include <linux/fs.h>
19 #include <linux/sched/signal.h>
20 #include <linux/uaccess.h>
21 #include <linux/capability.h>
22 #include <linux/kernel_stat.h>
23 #include <linux/gfp.h>
24 #include <linux/mm.h>
25 #include <linux/swap.h>
26 #include <linux/swapops.h>
27 #include <linux/syscalls.h>
28 #include <linux/mman.h>
29 #include <linux/pagemap.h>
30 #include <linux/file.h>
31 #include <linux/uio.h>
32 #include <linux/error-injection.h>
33 #include <linux/hash.h>
34 #include <linux/writeback.h>
35 #include <linux/backing-dev.h>
36 #include <linux/pagevec.h>
37 #include <linux/security.h>
38 #include <linux/cpuset.h>
39 #include <linux/hugetlb.h>
40 #include <linux/memcontrol.h>
41 #include <linux/shmem_fs.h>
42 #include <linux/rmap.h>
43 #include <linux/delayacct.h>
44 #include <linux/psi.h>
45 #include <linux/ramfs.h>
46 #include <linux/page_idle.h>
47 #include <linux/migrate.h>
48 #include <linux/pipe_fs_i.h>
49 #include <linux/splice.h>
50 #include <linux/rcupdate_wait.h>
51 #include <linux/sched/mm.h>
52 #include <linux/fsnotify.h>
53 #include <asm/pgalloc.h>
54 #include <asm/tlbflush.h>
55 #include "internal.h"
56
57 #define CREATE_TRACE_POINTS
58 #include <trace/events/filemap.h>
59
60 /*
61  * FIXME: remove all knowledge of the buffer layer from the core VM
62  */
63 #include <linux/buffer_head.h> /* for try_to_free_buffers */
64
65 #include <asm/mman.h>
66
67 #include "swap.h"
68
69 /*
70  * Shared mappings implemented 30.11.1994. It's not fully working yet,
71  * though.
72  *
73  * Shared mappings now work. 15.8.1995 Bruno.

```

Feb 19, 25 9:07

filemap.txt

Page 2/2

```

74 *
75 * finished 'unifying' the page and buffer cache and SMP-threaded the
76 * page-cache, 21.05.1999, Ingo Molnar <mingo@redhat.com>
77 *
78 * SMP-threaded pagemap-LRU 1999, Andrea Arcangeli <andrea@suse.de>
79 */
80
81 /*
82 * Lock ordering:
83 *
84 * ->i_mmap_rwsem          (truncate_pagecache)
85 * ->private_lock          (__free_pte->block_dirty_folio)
86 * ->swap_lock             (exclusive_swap_page, others)
87 * ->i_pages lock
88 *
89 * ->i_rwsem
90 * ->invalidate_lock        (acquired by fs in truncate path)
91 * ->i_mmap_rwsem          (truncate->unmap_mapping_range)
92 *
93 * ->mmap_lock
94 * ->i_mmap_rwsem
95 * ->page_table_lock or pte_lock (various, mainly in memory.c)
96 * ->i_pages lock          (arch-dependent flush_dcache_mmap_lock)
97 *
98 * ->mmap_lock
99 * ->invalidate_lock        (filemap_fault)
100 * ->lock_page             (filemap_fault, access_process_vm)
101 *
102 * ->i_rwsem                (generic_perform_write)
103 * ->mmap_lock              (fault_in_readable->do_page_fault)
104 *
105 * bdi->wb.list_lock
106 * sb_lock                  (fs/fs-writeback.c)
107 * ->i_pages lock           (__sync_single_inode)
108 *
109 * ->i_mmap_rwsem
110 * ->anon_vma.lock          (vma_merge)
111 *
112 * ->anon_vma.lock
113 * ->page_table_lock or pte_lock (anon_vma_prepare and various)
114 *
115 * ->page_table_lock or pte_lock
116 * ->swap_lock              (try_to_unmap_one)
117 * ->private_lock           (try_to_unmap_one)
118 * ->i_pages lock           (try_to_unmap_one)
119 * ->lrvec->lru_lock        (follow_page_mask->mark_page_accessed)
120 * ->lrvec->lru_lock        (check_pte_range->folio_isolate_lru)
121 * ->private_lock           (folio_remove_rmap_pte->set_page_dirty)
122 * ->i_pages lock           (folio_remove_rmap_pte->set_page_dirty)
123 * bdi.wb->list_lock        (folio_remove_rmap_pte->set_page_dirty)
124 * ->inode->i_lock           (folio_remove_rmap_pte->set_page_dirty)
125 * bdi.wb->list_lock        (zap_pte_range->set_page_dirty)
126 * ->inode->i_lock           (zap_pte_range->set_page_dirty)
127 * ->private_lock           (zap_pte_range->block_dirty_folio)
128 */
129
130 static void page_cache_delete(struct address_space *mapping,
131                             struct folio *folio, void *shadow)
132 {
133     XA_STATE(xas, &mapping->i_pages, folio->index);
134     long nr = 1;
135
136 [the point is: fine-grained locking leads to complexity.]

```

Feb 19, 25 8:51

dclp.txt

Page 1/2

```

1  3. Cautionary tale
2
3  Consider the code below:
4
5      struct foo {
6          int abc;
7          int def;
8      };
9      static int ready = 0;
10     static mutex_t mutex;
11     static struct foo* ptr = 0;
12
13     void
14     doublecheck_alloc()
15     {
16         if (!ready) { /* <-- accesses shared variable w/out holding mutex */
17             mutex_acquire(&mutex);
18             if (!ready) {
19                 ptr = alloc_foo(); /* <-- sets ptr to be non-zero */
20                 ready = 1;
21             }
22             mutex_release(&mutex);
23
24             return;
25         }
26     }
27
28
29 This is an example of the so-called "double-checked locking pattern."
30 The programmer's intent is to avoid a mutex acquisition in the common
31 case that 'ptr' is already initialized. So the programmer checks a flag
32 called 'ready' before deciding whether to acquire the mutex and
33 initialize 'ptr'. The intended use of doublecheck_alloc() is something
34 like this:
35
36     void f() {
37         doublecheck_alloc();
38         ptr->abc = 5;
39     }
40
41     void g() {
42         doublecheck_alloc();
43         ptr->def = 6;
44     }
45
46 We assume here that mutex_acquire() and mutex_release() are implemented
47 correctly (each contains memory barriers internally, etc.). Furthermore,
48 we assume that the compiler does not reorder instructions.
49
50 NEVERTHELESS, on multi-CPU machines that do not offer sequential
51 consistency, doublecheck_alloc() is broken. What is the bug?
52
53 -----
54
55 Unfortunately, double-checked initialization (or double-checked locking
56 as it's sometimes known) is a common coding pattern. Even some
57 references on threads suggest it! Still, it's broken.
58
59 While you can fix it (in C) by adding barriers (exercise:
60 where?), this is not recommended, as the code is tricky to reason about.
61 One of the points of this example is to show you why it's so important
62 to protect global data with a mutex, even if "all" one is doing is
63 reading memory, and even if the shortcut looks harmless.
64
65

```

Feb 19, 25 8:51

dclp.txt

Page 2/2

```
66 Finally, here are some references on this topic:
67
68 --http://www.aristeia.com/Papers/DDJ_Jul_Aug_2004_revised.pdf
69 explores issues with this pattern in C++
70
71 --The "Double-Checked Locking is Broken" Declaration:
72 http://www.cs.umd.edu/~pugh/java/memoryModel/DoubleCheckedLocking.html
73
74 --C++11 provides a way to implement the pattern correctly and
75 portably (again, using memory barriers):
76 https://preshing.com/20130930/double-checked-locking-is-fixed-in-cpp11/
```