Processes and Threads Implementation

Learning Outcomes

- A basic understanding of the MIPS R3000 assembly and compiler generated code.
- An understanding of the typical implementation strategies of processes and threads
  - Including an appreciation of the trade-offs between the implementation approaches
  - Kernel-threads versus user-level threads
- A detailed understanding of “context switching”

MIPS R3000

- Load/store architecture
  - No instructions that operate on memory except load and store
  - Simple load/stores to/from memory from/to registers
    - Store word: \texttt{sw r4, (r5)}
      - Store contents of r4 in memory using address contained in register r5
    - Load word: \texttt{lw r3, (r7)}
      - Load contents of memory into r3 using address contained in r7
      - Delay of one instruction after load before data available in destination register
      - Must always an instruction between a load from memory and the subsequent use of the register.
  - \texttt{lw, sw, lb, sb, lh, sh, ...}

- Arithmetic and logical operations are register to register operations
  - E.g., \texttt{add r3, r2, r1}
  - No arithmetic operations on memory

Example

- \texttt{add r3, r2, r1} ⇒ $r3 = r2 + r1$
- Some other instructions
  - \texttt{add, sub, and, or, xor, sll, srl}
  - \texttt{move r2, r1} ⇒ $r2 = r1$

Example code

Simple code example: $a = a + 1$

\begin{verbatim}
lw r4, 32(r29) // r29 = stack pointer
li r5, 1
add r0, r4, r5
sw r0, 32(r29)
\end{verbatim}
MIPS Registers

- User-mode accessible registers
  - 32 general purpose registers
    - r0 hardwired to zero
    - r31 the link register for jump-and-link (JAL) instruction
- HI/LO
  - 2 * 32-bits for multiply and divide
- PC
  - Not directly visible
  - Modified implicitly by jump and branch instructions

Branching and Jumping

- Branching and jumping have a branch delay slot
  - The instruction following a branch or jump is always executed prior to destination of jump

MIPS R3000

- RISC architecture – 5 stage pipeline
  - Instruction partially through pipeline prior to jmp having an effect

Jump and Link Instruction

- JAL is used to implement function calls
  - r31 = PC+8
- Return Address register (RA) is used to return from function call

Compiler Register Conventions

- Given 32 registers, which registers are used for
  - Local variables?
  - Argument passing?
  - Function call results?
  - Stack Pointer?

<table>
<thead>
<tr>
<th>Reg No</th>
<th>Name</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>zero</td>
<td>Always returns 0</td>
</tr>
<tr>
<td>1</td>
<td>sp</td>
<td>assembler (temporary) reserved for use by assembler</td>
</tr>
<tr>
<td>2-3</td>
<td>t0-t1</td>
<td>Value preserved, FP returned by subroutine</td>
</tr>
<tr>
<td>4-7</td>
<td>t2-t7</td>
<td>Temporary, FP (t0) and parameters for a subroutine</td>
</tr>
<tr>
<td>8-15</td>
<td>t8-t15</td>
<td>Temporary, subroutine may use without saving</td>
</tr>
<tr>
<td>16-25</td>
<td>t16-t25</td>
<td>Temporary, subroutine may use without saving</td>
</tr>
<tr>
<td>26-29</td>
<td>t26-t29</td>
<td>Subroutine “register variables” – a subroutine which will write one of these must save the old value and restore it before it ends, so the calling routine sets their values preserved</td>
</tr>
<tr>
<td>30</td>
<td>t30</td>
<td>Reserved for use by interrupts (not handle) — may change under certain circumstances</td>
</tr>
<tr>
<td>31</td>
<td>ra</td>
<td>Global pointer – some machine systems maintain this to give easy access to ordinary “static” or “extern” variables</td>
</tr>
<tr>
<td>32</td>
<td>sp</td>
<td>Link register – function return address for subroutine</td>
</tr>
<tr>
<td>33</td>
<td>t33</td>
<td>Subroutine “register variables” – a subroutine which need one can use this as a “hidden” pointer</td>
</tr>
<tr>
<td>34</td>
<td>t34</td>
<td>Return address for subroutine</td>
</tr>
</tbody>
</table>
Simple factorial

```c
int fact(int n)
{
    int r = 1;
    int i;
    for (i = 1; i <= n; i++) {
        r = r * i;
    }
    return r;
}
```

Function Stack Frames

- Each function call allocates a new stack frame for local variables, the return address, previous frame pointer etc.
  - Frame pointer: start of current stack frame
  - Stack pointer: end of current stack frame
- Example: assume f1() calls f2(), which calls f3().

Function Stack Frames

- Each function call allocates a new stack frame for local variables, the return address, previous frame pointer etc.
  - Frame pointer: start of current stack frame
  - Stack pointer: end of current stack frame
- Example: assume f1() calls f2(), which calls f3().

Stack Frame

- MIPS calling convention for gcc
  - Args 1-4 have space reserved for them

Example Code

```c
int sixargs(int a, int b, int c, int d, int e, int f)
{
    int i;
    return a + b + c + d + e + f;
}
```
0400011c <main>:
040011c: 27bdffd8 addiu sp,sp,-40
0400120: afbf0024 sw ra,36(sp)
0400124: afbe0020 sw s8,32(sp)
0400128: 03a0f021 move s8,sp
040012c: 24020005 li v0,5
0400130: afa20010 sw v0,16(sp)
0400134: 24020006 li v0,6
0400138: afa20014 sw v0,20(sp)
040013c: 24040001 li a0,1
0400140: 24050002 li a1,2
0400144: 24060003 li a2,3
0400148: 24070005 li a3,4
0400150: 26080018 sw v0,24(s8)
0400154: 2609001c sw a0,8(s8)
0400158: 260a0020 lw v1,34(sp)
040015c: 260b0020 lw a1,32(sp)
0400160: 0a000008 jr ra
0400164: 27bd9020 addiu sp,sp,40

040000b0 <sixargs>:
0400b0: 27badff8 addiu sp,sp,-8
0400b4: afbe0010 sw s8,0(sp)
0400b8: 03a0f021 move s8,sp
0400bc: afa20008 sw v0,8(sp)
0400c0: afc0000c sw a1,12(sp)
0400c4: afc00110 sw a2,16(sp)
0400c8: afc00114 sw a3,20(sp)
0400cc: afc00100 lw v1,8(sp)
0400da: afc0010c lw v9,12(sp)
0400ee: 00000000 nop
0400f0: 04211021 adds v6,v1,v0
0400f4: 06c00110 lw v1,16(sp)
0400f8: 00000000 nop
0400fc: 00431021 adds v6,v1,v0

004000b: 00431021 adds v6,v1,v0
000000: 8fc30010 lw v1,16(sp)
000004: 00000000 nop
000008: 0431021 adds v6,v1,v0
00000c: 06c00110 lw v1,20(sp)
000010: 00000000 nop
000014: 0431021 adds v6,v1,v0
000018: 06c00110 lw v1,24(sp)
00001c: 00000000 nop

Process

- Minimally consist of three segments
  - Text: contains the code (instructions)
  - Data: global variables
  - Stack: activation records of procedure/function/method
  - Local variables
- Note:
  - Data can dynamically grow up
  - E.g., malloc()-ing
  - The stack can dynamically grow down
  - E.g., increasing function call depth or recursion

Process Memory Layout

Stack
Gap
Data
Text

The Process Model

- Multiprogramming of four programs

Process

User Mode

Kernel Mode

Scheduler

Processes

Process’s user-level stack and execution state

Process’s in-kernel stack and execution state
Processes

- User-mode
  - Processes (programs) scheduled by the kernel
  - Isolated from each other
  - No concurrency issues between each other
- Kernel-mode
  - Nearly all activities still associated with a process
  - Kernel memory shared between all processes
  - Concurrency issues exist between processes concurrently executing in a system call

Threads

The Thread Model

(a) Three processes each with one thread
(b) One process with three threads

The Thread Model

- Items shared by all threads in a process
- Items that exist per thread

A Subset of POSIX threads API

```c
int pthread_create(pthread_t *t, const pthread_attr_t *attr, void *(*f)(void *), void *arg);
int pthread_exit(void *arg);
int pthread_mutex_init(pthread_mutex_t *lock, const pthread_mutexattr_t *attr);
int pthread_mutex_destroy(pthread_mutex_t *lock);
int pthread_mutex_lock(pthread_mutex_t *lock);
int pthread_mutex_unlock(pthread_mutex_t *lock);
int pthread_rwlock_init(pthread_rwlock_t *lock, const pthread_rwlockattr_t *attr);
int pthread_rwlock_destroy(pthread_rwlock_t *lock);
int pthread_rwlock_rdlock(pthread_rwlock_t *lock);
int pthread_rwlock_wrlock(pthread_rwlock_t *lock);
int pthread_rwlock_unlock(pthread_rwlock_t *lock);
```

Where to Implement Application

- User-level threads implemented in a library?
- Kernel-level threads implemented in the OS?
Implementing Threads in User Space

User-level Threads

• Implementation at user-level
  • User-level Thread Control Block (TCB), ready queue, blocked queue, and dispatcher
  • Kernel has no knowledge of the threads (it only sees a single process)
  • If a thread blocks waiting for a resource held by another thread inside the same process, its state is saved and the dispatcher switches to another ready thread
  • Thread management (create, exit, yield, wait) are implemented in a runtime support library

User-Level Threads

• Pros
  • Thread management and switching at user level is much faster than doing it in kernel level
  • No need to trap (take syscall exception) into kernel and back to switch
  • Dispatcher algorithm can be tuned to the application
    • E.g. use priorities
    • Can be implemented on any OS (thread or non-thread aware)
    • Can easily support massive numbers of threads on a per-application basis
    • Use normal application virtual memory
    • Kernel memory more constrained. Difficult to efficiently support wildly differing numbers of threads for different applications.

User-level Threads

• Cons
  • Threads have to yield() manually (no timer interrupt delivery to user-level)
  • Co-operative multithreading
    • A single poorly design/implemented thread can monopolise the available CPU time
    • There are work-arounds (e.g. a timer signal per second to enable pre-emptive multithreading), they are coarse grain and a kludge.
    • Does not take advantage of multiple CPUs (in reality, we still have a single threaded process as far as the kernel is concerned)

User-Level Threads

• Cons
  • If a thread makes a blocking system call (or takes a page fault), the process (and all the internal threads) blocks
  • Can't overlap I/O with computation
Implementing Threads in the Kernel

A threads package managed by the kernel

Kernel-provided Threads

- Also called kernel-level threads
- Even though they provide threads to applications
- Threads are implemented by the kernel
- TCBs are stored in the kernel
  - A subset of information in a traditional PCB
  - The subset related to execution context
  - TCBs have a PCB associated with them
  - Resources associated with the group of threads (the process)
- Thread management calls are implemented as system calls
  - E.g. create, wait, exit

Kernel-provided Threads

- Cons
  - Thread creation and destruction, and blocking and unblocking threads requires kernel entry and exit.
  - More expensive than user-level equivalent

Kernel-provided Threads

- Pros
  - Preemptive multithreading
  - Parallelism
  - Can overlap blocking I/O with computation
  - Can take advantage of a multiprocessor

Multiprogramming Implementation

Skeleton of what lowest level of OS does when an interrupt occurs – a context switch

1. Hardware stacks program counter, etc.
2. Hardware reads new program counter from interrupt vector.
3. Assembly language procedure saves registers.
4. Assembly language procedure sets up new stack.
5. C interrupt service runs (typically reads and buffers input).
6. Scheduler decides which process is to run next.
7. C procedure returns to the assembly code.
8. Assembly language procedure starts up new current process.
**Context Switch Terminology**

- A context switch can refer to:
  - A switch between threads
    - Involving saving and restoring of state associated with a thread
  - A switch between processes
    - Involving the above, plus extra state associated with a process.
    - E.g., memory maps

**Context Switch Occurrence**

- A switch between process/threads can happen any time the OS is invoked:
  - On a system call
    - Mandatory if system call blocks or on exit();
  - On an exception
    - Mandatory if offender is killed
  - On an interrupt
    - Triggering a dispatch is the main purpose of the timer interrupt

- A thread switch can happen between any two instructions

Note instructions do not equal program statements

**Context Switch**

- Context switch must be transparent for processes/threads
  - When dispatched again, process/thread should not notice that something else was running in the meantime (except for elapsed time)

  \[ \Rightarrow \text{OS must save all state that affects the thread} \]

- This state is called the process/thread context

- Switching between process/threads consequently results in a context switch.

**Simplified Explicit Thread Switch**

\[
\text{thread_switch}(a,b) \{
\text{thread_switch}(a,b) \{
\text{thread_switch}(b,a) \{
\}
\}
\}
\]

**Example Context Switch**

- Running in user mode, SP points to user-level stack (not shown on slide)

**Representation of Kernel Stack (Memory)**

SP
Example Context Switch

- Take an exception, syscall, or interrupt, and we switch to the kernel stack

```
SP
```

Example Context Switch

- We push a `trapframe` on the stack
  - Also called `exception frame`, `user-level context`...
  - Includes the user-level PC and SP

```
SP
```

Example Context Switch

- Call `C` code to process syscall, exception, or interrupt
  - Results in a `C` activation stack building up

```
C activation stack
trapframe
```

Example Context Switch

- The kernel decides to perform a context switch
  - It chooses a target thread (or process)
  - It pushes remaining kernel context onto the stack

```
Kemel State
C activation stack
trapframe
```

Example Context Switch

- Any other existing thread must
  - be in kernel mode (on a uni processor),
  - and have a similar stack layout to the stack we are currently using

```
Kernel State
C activation stack
trapframe
```

Example Context Switch

- We save the current SP in the PCB (or TCB), and load the SP of the target thread.
  - Thus we have switched contexts

```
Kemel State
C activation stack
trapframe
```

```
Kemel State
C activation stack
trapframe
```

```
Kemel State
C activation stack
trapframe
```

```
Kemel State
C activation stack
trapframe
```
Example Context Switch

• Load the target thread's previous context, and return to C

The Interesting Part of a Thread Switch

• What does the “push kernel state” part do???

Simplified OS/161 thread_switch

```c
static
void
thread_switch(threadstate_t newstate, struct wchan *wc)
{
    struct thread *cur, *next;
    cur = curthread;
    do {
        next = threadlist_remhead(&curcpu->c_runqueue);
        if (next == NULL) {
            cpu_idle();
        }
    } while (next == NULL);
    /* do the switch (in assembler in switch.S) */
    switchframe_switch(&cur->t_context, &next->t_context);
}
```
/* a0 contains the address of the switchframe pointer in the old thread.
 * a1 contains the address of the switchframe pointer in the new thread.
 * 
 * The switchframe pointer is really the stack pointer. The other
 * registers get saved on the stack, namely:
 * 
 *      s0-s6, s8
 *      gp, ra
 * 
 * The order must match <mips/switchframe.h>.
 * 
 * Note that while we'll explicitly need to save s7 too, because we
 * won't be able to hold both a thread saving it would interfere with the way
 * switchframe is managed by thread.c, so we'll just let thread.c
 * manage it.
 */

OS/161 switchframe_switch

    /* Allocate stack space for saving 10 registers. 10*4 = 40 */
    addi sp, sp, -40

    /* Save the registers */
    lw   ra, 36(sp)
    lw   gp, 32(sp)
    lw   s8, 28(sp)
    lw   s6, 24(sp)
    lw   s5, 20(sp)
    lw   s4, 16(sp)
    lw   s3, 12(sp)
    lw   s2,  8(sp)
    lw   s1,  4(sp)
    lw   s0,  0(sp)

    /* Store the old stack pointer in the old thread */
    sw   sp, 0(a0)

OS/161 switchframe_switch

    /* Get the new stack pointer from the new thread */
    lw   sp, 0(a1)
    nop           /* delay slot for load */

    /* Now, restore the registers */
    lw   s0, 0(sp)
    lw   s1,  4(sp)
    lw   s2,  8(sp)
    lw   s3, 12(sp)
    lw   s4, 16(sp)
    lw   s5, 20(sp)
    lw   s6, 24(sp)
    lw   s8, 28(sp)
    lw   gp, 32(sp)
    lw   ra, 36(sp)
    nop                  /* delay slot for load */

    /* and return. */
    j ra
    addi sp, sp, 40      /* in delay slot */

OS/161 switchframe_switch

    /* and return */
    j lr
    addi sp, sp, 48      /* in delay slot */

Revisiting Thread Switch