CMSC 313 Lecture 11

- Project 3 Questions
- Recap Last Lecture
- Separate Compilation
- Stack Instructions: PUSH, POP
- Subroutines (a.k.a. Functions) in Assembly

Project 3: An Error-Correcting Code Due: Thursday October 14, 2004

Objective

The objectives of this programming project are 1) for you to gain familiarity with data manipulation at the bit level and 2) for you to write more complex assembly language programs.

Background

In Project 2, we saw that checksums can be used detect corrupted files. However, there is not much we can do after we have detected the corruption. An error-correcting code is able to fix errors, not just detect them.

In this project, we will use a 31-bit Hamming code that can correct a 1-bit error in each 32-bit codeword. Each 32-bit codeword encodes 3 bytes of the original data. The format of the codeword is shown on the next page.

Assignment

Write an assembly language program that encodes the input file using the codeword format described below. As in Project 2, use Unix input and output redirection:

```
./a.out <ifile >ifile.ham
```

Some details:

- Your program must read a block of bytes from the input. You should not read from the input one byte at a time or three bytes at a time. (That would be terribly inefficient.)
- You may assume that when the operating system returns with 0 bytes read that the end of the input file has been reached. On the other hand, you may not assume that the end of the file has been reached when the operating system gives you fewer bytes than your block size. Similarly, you may not assume that the operating system will comply with your request for a number of input bytes that is divisible by 3.
- The 32-bit codewords must be written out in little-endian format.

The C source code for two programs `decode.c` and `corrupt.c` are provided in the GL file system in the directory: `/afs/umbc.edu/users/c/h/chang/pub/cs313`. These two programs can be used to decode an encoded file and to corrupt an encoded file. You can use these programs to check if your program is working correctly. Both programs use I/O redirection.

Record some sample runs of your program using the Unix `script` command. You should show that you can encode a file using your program, then decode it and obtain a file that is identical to the original. Use the Unix `diff` command to compare the original file with the decoded file. You should also show that this works when the file is corrupted.

Implementation Notes

- The parity flag PF is set to 1 if the result of an instruction contains an even number of 1's. Unfortunately, PF only looks at the lowest 8 bits of the result. For this project, you will need to compute 32-bit parities. Here's a simple way to comput the parity of the EAX register.

```
mov     ebx,eax
shr     eax,16
xor     ax,bx
xor     al,ah
jp      even_label
```

  Note that the EAX and EBX registers are modified in this process, so you may need to use different registers.

- A main issue in this project is handling the "extra characters" at the end of a block of input after you have processed all the 3-byte "groups". E.g., if your block size 128, then you will have 2 characters left over after processing 42 three-byte groups (42 x 3 = 126). These 2 extra characters must be grouped with the first character of the next block (if there is a next block). Think about this situation *before* you begin coding.

- Another main issue is the last 32-bit word output by your program. Note that the bits m1 and m0 must be set *before* you compute the parity bits p4, p3, p2, p1 and p0.

Turning in your program

Use the UNIX `submit` command on the GL system to turn in your project. You should submit two files: 1) the assembly language program and 2) the typescript file of sample runs of your program. The class name for submit is `cs313_0101`. The name of the assignment name is `proj3`. The UNIX command to do this should look something like:

```
submit cs313_0101 proj3 encode.asm typescript
```

Codeword format

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
a7	a6	a5	a4	a3	a2	a1	a0	b7	b6	b5	b4	b3	b2	b1	p4	b0	c7	c6	c5	c4	c3	c2	p3	c1	c0	m1	p2	m0	p1	p0	0

bit 0 is not used and always holds a 0.

1st byte of data = a7 a6 a5 a4 a3 a2 a1 a0

2nd byte of data = b7 b6 b5 b4 b3 b2 b1 b0

3rd byte of data = c7 c6 c5 c4 c3 c2 c1 c0

p4, p3, p2, p1 and p0 are used to ensure that these bit positions have an even number of 1's:

```
p0:  1  3  5  7  9 11 13 15 17 19 21 23 25 27 29 31
p1:  2  3  6  7 10 11 14 15 18 19 22 23 26 27 30 31
p2:  4  5  6  7 12 13 14 15 20 21 22 23 28 29 30 31
p3:  8  9 10 11 12 13 14 15 24 25 26 27 28 29 30 31
p4: 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
```

m1 and m0 are only used in the last word of the encoded file. They depend on the original file size (in number of bytes).

m1 m0 = 00 if the file size mod 3 is 0

m1 m0 = 01 if the file size mod 3 is 1

m1 m0 = 10 if the file size mod 3 is 2

Last Time

- **Compilers: mechanical process to turn high-level languages to assembly language**
  - ◇ You can even write a compiler yourself with the help of lex & yacc
  - ◇ Anything you can do in C you can also do in assembly language
- **Assembler**
  - ◇ Needs two passes to back patch forward references
  - ◇ Converts assembly language mnemonics to machine code
  - ◇ You can even do this yourself by hand with the help of some tables from the Intel manual :-P
- **Linking & Loading**
  - ◇ Linker resolves external references
  - ◇ Loader assigns addresses to code and data sections.
  - ◇ The loader must also patch instructions with real addresses

```
; File: sep1.asm
;
; File 1 for separate compilation example

global gvar1, _start
extern gvar2, add_these

        section .data
foo:    db      12h
gvar1:  dd      17h
lvar1:  dd      42h

        section .text
_start: mov     eax, [gvar1]
        mov     ebx, [gvar2]
        mov     ecx, [lvar1]

        call    add_these       ; gvar1 := eax+ebx+ecx
        mov     ebx, [gvar1]    ; store in return code
        mov     eax, 1          ; sycall number for exit
        int     080h            ; bye-bye
```

```
; File: sep2.asm
;
; File 2 for separate compilation example

global gvar2, add_these
extern gvar1

        section .data
bar:    dw      07h
gvar2:  dd      03h
lvar1:  dd      02h     ; same name as other lvar1, OK

        section .text
add_these:                              ; no regs altered!
        mov     [gvar1], dword 0        ; clear destination
        add     [gvar1], eax
        add     [gvar1], ebx
        add     [gvar1], ecx
        ret
```

```
 1                          ; File: sep1.asm
 2                          ;
 3                          ; File 1 for separate compilation example
 4
 5                          global gvar1, _start
 6                          extern gvar2, add_these
 7
 8                                  section .data
 9
10 00000000 12             foo:     db      12h
11 00000001 17000000       gvar1:   dd      17h
12 00000005 42000000       lvar1:   dd      42h
13
14                                  section .text
15 00000000 A1[01000000]   _start:  mov     eax, [gvar1]
16 00000005 8B1D[00000000]          mov     ebx, [gvar2]
17 0000000B 8B0D[05000000]          mov     ecx, [lvar1]
18
19 00000011 E8(00000000)            call    add_these       ; gvar1 := eax+ebx+ecx
20 00000016 8B1D[01000000]          mov     ebx, [gvar1]    ; store in return code
21 0000001C B801000000              mov     eax, 1          ; sycall number for exit
22 00000021 CD80                    int     080h            ; bye-bye
```

```
 1                                      ; File: sep2.asm
 2                                      ;
 3                                      ; File 2 for separate compilation example
 4
 5                                      global gvar2, add_these
 6                                      extern gvar1
 7
 8                                       section .data
 9
10 00000000 0700                       bar:  dw   07h
11 00000002 03000000                   gvar2:    dd   03h
12 00000006 02000000                   lvar1:    dd   02h  ; same name as other lvar1, OK
13
14                                       section .text
15                                     add_these:                  ; no regs altered!
16 00000000 C705[00000000]0000-         mov [gvar1], dword 0 ; clear destination
17 00000008 0000
18 0000000A 0105[00000000]              add [gvar1], eax
19 00000010 011D[00000000]              add [gvar1], ebx
20 00000016 010D[00000000]              add [gvar1], ecx
21 0000001C C3                          ret
```

```
linux3% nasm -f elf -l sep1.lst sep1.asm
linux3% nasm -f elf -l spe2.lst sep2.asm
linux3% ld sep1.o sep2.o
linux3% a.out
linux3% echo $?
92
linux3%


linux3% objdump -h sep1.o

sep1.o:     file format elf32-i386

Sections:
Idx Name          Size      VMA       LMA       File off  Algn
  0 .data         00000009  00000000  00000000  00000180  2**2
                  CONTENTS, ALLOC, LOAD, DATA
  1 .text         00000023  00000000  00000000  00000190  2**4
                  CONTENTS, ALLOC, LOAD, RELOC, READONLY, CODE
  2 .comment      0000001c  00000000  00000000  000001c0  2**0
                  CONTENTS, READONLY
linux3%


linux3% objdump -t sep1.o

sep1.o:     file format elf32-i386

SYMBOL TABLE:
00000000 l    df *ABS*  00000000 sep1.asm
00000000 l    d  *ABS*  00000000
00000000 l    d  .data  00000000
00000000 l    d  .text  00000000
00000000 l       .data  00000000 foo
00000005 l       .data  00000000 lvar1
00000000         *UND*  00000000 gvar2
00000000         *UND*  00000000 add_these
00000001 g       .data  00000000 gvar1
00000000 g       .text  00000000 _start
```

```
linux3% objdump -h sep2.o

sep2.o:     file format elf32-i386

Sections:
Idx Name          Size      VMA       LMA       File off  Algn
  0 .data         0000000a  00000000  00000000  00000180  2**2
                  CONTENTS, ALLOC, LOAD, DATA
  1 .text         0000001d  00000000  00000000  00000190  2**4
                  CONTENTS, ALLOC, LOAD, RELOC, READONLY, CODE
  2 .comment      0000001c  00000000  00000000  000001b0  2**0
                  CONTENTS, READONLY
linux3%
linux3% objdump -t sep2.o

sep2.o:     file format elf32-i386

SYMBOL TABLE:
00000000 l    df *ABS*  00000000 sep2.asm
00000000 l    d  *ABS*  00000000
00000000 l    d  .data  00000000
00000000 l    d  .text  00000000
00000000 l       .data  00000000 bar
00000006 l       .data  00000000 lvar1
00000000         *UND*  00000000 gvar1
00000002 g       .data  00000000 gvar2
00000000 g       .text  00000000 add_these
```

```
linux3% objdump -h a.out

a.out:     file format elf32-i386

Sections:
Idx Name          Size      VMA       LMA       File off  Algn
  0 .text         0000004d  08048080  08048080  00000080  2**4
                  CONTENTS, ALLOC, LOAD, READONLY, CODE
  1 .data         00000016  080490d0  080490d0  000000d0  2**2
                  CONTENTS, ALLOC, LOAD, DATA
  2 .bss          00000002  080490e6  080490e6  000000e6  2**0
                  CONTENTS
  3 .comment      00000038  00000000  00000000  000000e8  2**0
                  CONTENTS, READONLY
linux3% objdump -t a.out

a.out:     file format elf32-i386

SYMBOL TABLE:
08048080 l    d  .text  00000000
080490d0 l    d  .data  00000000
080490e6 l    d  .bss   00000000
00000000 l    d  .comment       00000000
00000000 l    d  *ABS*  00000000
00000000 l    d  *ABS*  00000000
00000000 l    d  *ABS*  00000000
00000000 l    df *ABS*  00000000 sep1.asm
080490d0 l       .data  00000000 foo
080490d5 l       .data  00000000 lvar1
00000000 l    df *ABS*  00000000 sep2.asm
080490dc l       .data  00000000 bar
080490e2 l       .data  00000000 lvar1
080480cd g     O *ABS*  00000000 _etext
080480b0 g       .text  00000000 add_these
08048080 g       .text  00000000 _start
080490de g       .data  00000000 gvar2
080490e6 g     O *ABS*  00000000 __bss_start
080490d1 g       .data  00000000 gvar1
080490e6 g     O *ABS*  00000000 _edata
080490e8 g     O *ABS*  00000000 _end
```

```
linux3% objdump -d a.out

a.out:     file format elf32-i386

Disassembly of section .text:

08048080 <_start>:
 8048080:       a1 d1 90 04 08          mov    0x80490d1,%eax
 8048085:       8b 1d de 90 04 08       mov    0x80490de,%ebx
 804808b:       8b 0d d5 90 04 08       mov    0x80490d5,%ecx
 8048091:       e8 1a 00 00 00          call   80480b0 <add_these>
 8048096:       8b 1d d1 90 04 08       mov    0x80490d1,%ebx
 804809c:       b8 01 00 00 00          mov    $0x1,%eax
 80480a1:       cd 80                   int    $0x80
 80480a3:       90                      nop
 80480a4:       90                      nop
 80480a5:       90                      nop
 80480a6:       90                      nop
 80480a7:       90                      nop
 80480a8:       90                      nop
 80480a9:       90                      nop
 80480aa:       90                      nop
 80480ab:       90                      nop
 80480ac:       90                      nop
 80480ad:       90                      nop
 80480ae:       90                      nop
 80480af:       90                      nop

080480b0 <add_these>:
 80480b0:       c7 05 d1 90 04 08 00    movl   $0x0,0x80490d1
 80480b7:       00 00 00
 80480ba:       01 05 d1 90 04 08       add    %eax,0x80490d1
 80480c0:       01 1d d1 90 04 08       add    %ebx,0x80490d1
 80480c6:       01 0d d1 90 04 08       add    %ecx,0x80490d1
 80480cc:       c3                      ret
linux3%
```

```
linux3% objdump -s a.out

a.out:     file format elf32-i386

Contents of section .text:
 8048080 a1d19004 088b1dde 9004088b 0dd59004  ................
 8048090 08e81a00 00008b1d d1900408 b8010000  ................
 80480a0 00cd8090 90909090 90909090 90909090  ................
 80480b0 c705d190 04080000 00000105 d1900408  ................
 80480c0 011dd190 0408010d d1900408 c3        ............
Contents of section .data:
 80490d0 12170000 00420000 00000000 07000300  .....B..........
 80490e0 00000200 0000                        ......
Contents of section .bss:
 80490e6 0000                                 ..
Contents of section .comment:
 0000 00546865 204e6574 77696465 20417373  .The Netwide Ass
 0010 656d626c 65722030 2e393800 00546865  embler 0.98..The
 0020 204e6574 77696465 20417373 656d626c   Netwide Assembl
 0030 65722030 2e393800                    er 0.98.
linux3% exit
```

```
; File: sep3.asm
;
; File 3 for separate compilation example

extern _start, add_these

        section .data

lvar1:  dd      03h     ; same name as other lvar1, OK

        section .text
test3:                          ; no regs altered!
        cmp     [lvar1], dword 7
        jne     _start
        jmp     add_these
```

```
linuxserver1% nasm -f elf sep3.asm

linuxserver1% objdump -t sep3.o

sep3.o:     file format elf32-i386

SYMBOL TABLE:
00000000 l    df *ABS*  00000000 sep3.asm
00000000 l    d  *ABS*  00000000
00000000 l    d  .data  00000000
00000000 l    d  .text  00000000
00000000 l       .data  00000000 lvar1
00000000 l       .text  00000000 test3
00000000         *UND*  00000000 _start
00000000         *UND*  00000000 add_these


linuxserver1% ld sep1.o sep2.o sep3.o
```

```
linuxserver1% objdump -t a.out

a.out:     file format elf32-i386

SYMBOL TABLE:
08048080 l    d  .text  00000000
080490e8 l    d  .data  00000000
080490fc l    d  .bss   00000000
00000000 l    d  .comment       00000000
00000000 l    d  *ABS*  00000000
00000000 l    d  *ABS*  00000000
00000000 l    d  *ABS*  00000000
00000000 l    df *ABS*  00000000 sep1.asm
080490ec l       .data  00000000 lvar1
00000000 l    df *ABS*  00000000 sep2.asm
080490f4 l       .data  00000000 lvar1
00000000 l    df *ABS*  00000000 sep3.asm
080490f8 l       .data  00000000 lvar1
080480d0 l       .text  00000000 test3
080480b0 g       .text  00000000 add_these
08048080 g       .text  00000000 _start
080490f0 g       .data  00000000 gvar2
080490fc g       *ABS*  00000000 __bss_start
080490e8 g       .data  00000000 gvar1
080490fc g       *ABS*  00000000 _edata
080490fc g       *ABS*  00000000 _end
```

```
linuxserver1% objdump -d a.out
a.out:     file format elf32-i386

Disassembly of section .text:

08048080 <_start>:
 8048080:       a1 e8 90 04 08          mov    0x80490e8,%eax
 8048085:       8b 1d f0 90 04 08       mov    0x80490f0,%ebx
 804808b:       8b 0d ec 90 04 08       mov    0x80490ec,%ecx
 8048091:       e8 1a 00 00 00          call   80480b0 <add_these>
 8048096:       8b 1d e8 90 04 08       mov    0x80490e8,%ebx
 804809c:       b8 01 00 00 00          mov    $0x1,%eax
 80480a1:       cd 80                   int    $0x80

080480b0 <add_these>:
 80480b0:       c7 05 e8 90 04 08 00    movl   $0x0,0x80490e8
 80480b7:       00 00 00
 80480ba:       01 05 e8 90 04 08       add    %eax,0x80490e8
 80480c0:       01 1d e8 90 04 08       add    %ebx,0x80490e8
 80480c6:       01 0d e8 90 04 08       add    %ecx,0x80490e8
 80480cc:       c3                      ret

080480d0 <test3>:
 80480d0:       81 3d f8 90 04 08 07    cmpl   $0x7,0x80490f8
 80480d7:       00 00 00
 80480da:       0f 85 a0 ff ff ff       jne    8048080 <_start>
 80480e0:       e9 cb ff ff ff          jmp    80480b0 <add_these>
```

Stack Instructions

- **PUSH *op***
  - the stack pointer ESP is decremented by the size of the operand
  - the operand is copied to [ESP]
- **POP *op***
  - the reverse of PUSH
  - [ESP] is copied to the destination operand
  - ESP is incremented by the size of the operand
- **Where is the stack?**
  - The stack has its own section
  - Linux processes wake up with ESP initialized properly
  - The stack grows "upward" – toward smaller addresses
  - Memory available to the stack set using 'limit'

PUSH—Push Word or Doubleword Onto the Stack

Opcode	Instruction	Description
FF /6	PUSH *r/m16*	Push *r/m16*
FF /6	PUSH *r/m32*	Push *r/m32*
50+*rw*	PUSH *r16*	Push *r16*
50+*rd*	PUSH *r32*	Push *r32*
6A	PUSH *imm8*	Push *imm8*
68	PUSH *imm16*	Push *imm16*
68	PUSH *imm32*	Push *imm32*
0E	PUSH CS	Push CS
16	PUSH SS	Push SS
1E	PUSH DS	Push DS
06	PUSH ES	Push ES
0F A0	PUSH FS	Push FS
0F A8	PUSH GS	Push GS

Description

Decrements the stack pointer and then stores the source operand on the top of the stack. The address-size attribute of the stack segment determines the stack pointer size (16 bits or 32 bits), and the operand-size attribute of the current code segment determines the amount the stack pointer is decremented (2 bytes or 4 bytes). For example, if these address- and operand-size attributes are 32, the 32-bit ESP register (stack pointer) is decremented by 4 and, if they are 16, the 16-bit SP register is decremented by 2. (The B flag in the stack segment's segment descriptor determines the stack's address-size attribute, and the D flag in the current code segment's segment descriptor, along with prefixes, determines the operand-size attribute and also the address-size attribute of the source operand.) Pushing a 16-bit operand when the stack address-size attribute is 32 can result in a misaligned the stack pointer (that is, the stack pointer is not aligned on a doubleword boundary).

The PUSH ESP instruction pushes the value of the ESP register as it existed before the instruction was executed. Thus, if a PUSH instruction uses a memory operand in which the ESP register is used as a base register for computing the operand address, the effective address of the operand is computed before the ESP register is decremented.

In the real-address mode, if the ESP or SP register is 1 when the PUSH instruction is executed, the processor shuts down due to a lack of stack space. No exception is generated to indicate this condition.

IA-32 Architecture Compatibility

For IA-32 processors from the Intel 286 on, the PUSH ESP instruction pushes the value of the ESP register as it existed before the instruction was executed. (This is also true in the real-address and virtual-8086 modes.) For the Intel 8086 processor, the PUSH SP instruction pushes the new value of the SP register (that is the value after it has been decremented by 2).

PUSH—Push Word or Doubleword Onto the Stack (Continued)

Operation

```
IF StackAddrSize     32
THEN
    IF OperandSize     32
        THEN
            ESP     ESP  4;
            SS:ESP     SRC; (* push doubleword *)
        ELSE (* OperandSize     16*)
            ESP     ESP  2;
            SS:ESP     SRC; (* push word *)
    FI;
ELSE (* StackAddrSize     16*)
    IF OperandSize     16
        THEN
            SP     SP  2;
             SS:SP     SRC; (* push word *)
        ELSE (* OperandSize     32*)
            SP     SP  4;
            SS:SP     SRC; (* push doubleword *)
    FI;
FI;
```

Flags Affected

None.

Protected Mode Exceptions

#GP(0)	If a memory operand effective address is outside the CS, DS, ES, FS, or GS segment limit.
	If the DS, ES, FS, or GS register is used to access memory and it contains a null segment selector.
#SS(0)	If a memory operand effective address is outside the SS segment limit.
#PF(fault-code)	If a page fault occurs.
#AC(0)	If alignment checking is enabled and an unaligned memory reference is made while the current privilege level is 3.

Real-Address Mode Exceptions

#GP	If a memory operand effective address is outside the CS, DS, ES, FS, or GS segment limit.

POP—Pop a Value from the Stack

Opcode	Instruction	Description
8F /0	POP *m16*	Pop top of stack into *m16*; increment stack pointer
8F /0	POP *m32*	Pop top of stack into *m32*; increment stack pointer
58+ *rw*	POP *r16*	Pop top of stack into *r16*; increment stack pointer
58+ *rd*	POP *r32*	Pop top of stack into *r32*; increment stack pointer
1F	POP DS	Pop top of stack into DS; increment stack pointer
07	POP ES	Pop top of stack into ES; increment stack pointer
17	POP SS	Pop top of stack into SS; increment stack pointer
0F A1	POP FS	Pop top of stack into FS; increment stack pointer
0F A9	POP GS	Pop top of stack into GS; increment stack pointer

Description

Loads the value from the top of the stack to the location specified with the destination operand and then increments the stack pointer. The destination operand can be a general-purpose register, memory location, or segment register.

The address-size attribute of the stack segment determines the stack pointer size (16 bits or 32 bits—the source address size), and the operand-size attribute of the current code segment determines the amount the stack pointer is incremented (2 bytes or 4 bytes). For example, if these address- and operand-size attributes are 32, the 32-bit ESP register (stack pointer) is incremented by 4 and, if they are 16, the 16-bit SP register is incremented by 2. (The B flag in the stack segment's segment descriptor determines the stack's address-size attribute, and the D flag in the current code segment's segment descriptor, along with prefixes, determines the operand-size attribute and also the address-size attribute of the destination operand.)

If the destination operand is one of the segment registers DS, ES, FS, GS, or SS, the value loaded into the register must be a valid segment selector. In protected mode, popping a segment selector into a segment register automatically causes the descriptor information associated with that segment selector to be loaded into the hidden (shadow) part of the segment register and causes the selector and the descriptor information to be validated (see the "Operation" section below).

A null value (0000-0003) may be popped into the DS, ES, FS, or GS register without causing a general protection fault. However, any subsequent attempt to reference a segment whose corresponding segment register is loaded with a null value causes a general protection exception (#GP). In this situation, no memory reference occurs and the saved value of the segment register is null.

The POP instruction cannot pop a value into the CS register. To load the CS register from the stack, use the RET instruction.

If the ESP register is used as a base register for addressing a destination operand in memory, the POP instruction computes the effective address of the operand after it increments the ESP register. For the case of a 16-bit stack where ESP wraps to 0h as a result of the POP instruction, the resulting location of the memory write is processor-family-specific.

POP—Pop a Value from the Stack (Continued)

The POP ESP instruction increments the stack pointer (ESP) before data at the old top of stack is written into the destination.

A POP SS instruction inhibits all interrupts, including the NMI interrupt, until after execution of the next instruction. This action allows sequential execution of POP SS and MOV ESP, EBP instructions without the danger of having an invalid stack during an interrupt[1]. However, use of the LSS instruction is the preferred method of loading the SS and ESP registers.

Operation

```
IF StackAddrSize     32
    THEN
        IF OperandSize     32
            THEN
                DEST     SS:ESP; (* copy a doubleword *)
                ESP     ESP + 4;
            ELSE (* OperandSize     16*)
                DEST     SS:ESP; (* copy a word *)
            ESP     ESP + 2;
        FI;
    ELSE (* StackAddrSize     16* )
        IF OperandSize     16
            THEN
                DEST     SS:SP; (* copy a word *)
                SP     SP + 2;
            ELSE (* OperandSize     32 *)
                DEST     SS:SP; (* copy a doubleword *)
                SP     SP + 4;
        FI;
FI;
```

Loading a segment register while in protected mode results in special checks and actions, as described in the following listing. These checks are performed on the segment selector and the segment descriptor it points to.

```
IF SS is loaded;
    THEN
        IF segment selector is null
            THEN #GP(0);
```

1. Note that in a sequence of instructions that individually delay interrupts past the following instruction, only the first instruction in the sequence is guaranteed to delay the interrupt, but subsequent interrupt-delaying instructions may not delay the interrupt. Thus, in the following instruction sequence:
   ```
   STI
   POP SS
   POP ESP
   ```
   interrupts may be recognized before the POP ESP executes, because STI also delays interrupts for one instruction.

POP—Pop a Value from the Stack (Continued)

```
        FI;
        IF segment selector index is outside descriptor table limits
            OR segment selector's RPL ≠ CPL
            OR segment is not a writable data segment
            OR DPL ≠ CPL
                THEN #GP(selector);
        FI;
        IF segment not marked present
            THEN #SS(selector);
    ELSE
        SS    segment selector;
        SS    segment descriptor;
    FI;
FI;
IF DS, ES, FS, or GS is loaded with non-null selector;
THEN
    IF segment selector index is outside descriptor table limits
        OR segment is not a data or readable code segment
        OR ((segment is a data or nonconforming code segment)
            AND (both RPL and CPL > DPL))
                THEN #GP(selector);
        IF segment not marked present
            THEN #NP(selector);
    ELSE
        SegmentRegister    segment selector;
        SegmentRegister    segment descriptor;
    FI;
FI;
IF DS, ES, FS, or GS is loaded with a null selector;
    THEN
        SegmentRegister    segment selector;
        SegmentRegister    segment descriptor;
FI;
```

Flags Affected

None.

Protected Mode Exceptions

#GP(0)	If attempt is made to load SS register with null segment selector.
	If the destination operand is in a nonwritable segment.
	If a memory operand effective address is outside the CS, DS, ES, FS, or GS segment limit.

Subroutine Instructions

- **CALL *label***
  - Used to call a subroutine
  - PUSHes the instruction pointer (EIP) on the stack
  - jump to the label
  - does NOTHING else
- **RET**
  - reverse of CALL
  - POPs the instruction pointer (EIP) off the stack
  - execution proceeds from the instruction after the CALL instruction
- **Parameters?**

CALL—Call Procedure

Opcode	Instruction	Description
E8 *cw*	CALL *rel16*	Call near, relative, displacement relative to next instruction
E8 *cd*	CALL *rel32*	Call near, relative, displacement relative to next instruction
FF /2	CALL *r/m16*	Call near, absolute indirect, address given in *r/m16*
FF /2	CALL *r/m32*	Call near, absolute indirect, address given in *r/m32*
9A *cd*	CALL *ptr16:16*	Call far, absolute, address given in operand
9A *cp*	CALL *ptr16:32*	Call far, absolute, address given in operand
FF /3	CALL *m16:16*	Call far, absolute indirect, address given in *m16:16*
FF /3	CALL *m16:32*	Call far, absolute indirect, address given in *m16:32*

Description

Saves procedure linking information on the stack and branches to the procedure (called procedure) specified with the destination (target) operand. The target operand specifies the address of the first instruction in the called procedure. This operand can be an immediate value, a general-purpose register, or a memory location.

This instruction can be used to execute four different types of calls:

- Near call—A call to a procedure within the current code segment (the segment currently pointed to by the CS register), sometimes referred to as an intrasegment call.
- Far call—A call to a procedure located in a different segment than the current code segment, sometimes referred to as an intersegment call.
- Inter-privilege-level far call—A far call to a procedure in a segment at a different privilege level than that of the currently executing program or procedure.
- Task switch—A call to a procedure located in a different task.

The latter two call types (inter-privilege-level call and task switch) can only be executed in protected mode. See the section titled “Calling Procedures Using Call and RET” in Chapter 6 of the *IA-32 Intel Architecture Software Developer’s Manual, Volume 1*, for additional information on near, far, and inter-privilege-level calls. See Chapter 6, *Task Management*, in the *IA-32 Intel Architecture Software Developer’s Manual, Volume 3*, for information on performing task switches with the CALL instruction.

Near Call. When executing a near call, the processor pushes the value of the EIP register (which contains the offset of the instruction following the CALL instruction) onto the stack (for use later as a return-instruction pointer). The processor then branches to the address in the current code segment specified with the target operand. The target operand specifies either an absolute offset in the code segment (that is an offset from the base of the code segment) or a relative offset (a signed displacement relative to the current value of the instruction pointer in the EIP register, which points to the instruction following the CALL instruction). The CS register is not changed on near calls.

CALL—Call Procedure (Continued)

For a near call, an absolute offset is specified indirectly in a general-purpose register or a memory location (*r/m16* or *r/m32*). The operand-size attribute determines the size of the target operand (16 or 32 bits). Absolute offsets are loaded directly into the EIP register. If the operand-size attribute is 16, the upper two bytes of the EIP register are cleared to 0s, resulting in a maximum instruction pointer size of 16 bits. (When accessing an absolute offset indirectly using the stack pointer [ESP] as a base register, the base value used is the value of the ESP before the instruction executes.)

A relative offset (*rel16* or *rel32*) is generally specified as a label in assembly code, but at the machine code level, it is encoded as a signed, 16- or 32-bit immediate value. This value is added to the value in the EIP register. As with absolute offsets, the operand-size attribute determines the size of the target operand (16 or 32 bits).

Far Calls in Real-Address or Virtual-8086 Mode. When executing a far call in real-address or virtual-8086 mode, the processor pushes the current value of both the CS and EIP registers onto the stack for use as a return-instruction pointer. The processor then performs a "far branch" to the code segment and offset specified with the target operand for the called procedure. Here the target operand specifies an absolute far address either directly with a pointer (*ptr16:16* or *ptr16:32*) or indirectly with a memory location (*m16:16* or *m16:32*). With the pointer method, the segment and offset of the called procedure is encoded in the instruction, using a 4-byte (16-bit operand size) or 6-byte (32-bit operand size) far address immediate. With the indirect method, the target operand specifies a memory location that contains a 4-byte (16-bit operand size) or 6-byte (32-bit operand size) far address. The operand-size attribute determines the size of the offset (16 or 32 bits) in the far address. The far address is loaded directly into the CS and EIP registers. If the operand-size attribute is 16, the upper two bytes of the EIP register are cleared to 0s.

Far Calls in Protected Mode. When the processor is operating in protected mode, the CALL instruction can be used to perform the following three types of far calls:

- Far call to the same privilege level.
- Far call to a different privilege level (inter-privilege level call).
- Task switch (far call to another task).

In protected mode, the processor always uses the segment selector part of the far address to access the corresponding descriptor in the GDT or LDT. The descriptor type (code segment, call gate, task gate, or TSS) and access rights determine the type of call operation to be performed.

If the selected descriptor is for a code segment, a far call to a code segment at the same privilege level is performed. (If the selected code segment is at a different privilege level and the code segment is non-conforming, a general-protection exception is generated.) A far call to the same privilege level in protected mode is very similar to one carried out in real-address or virtual-8086 mode. The target operand specifies an absolute far address either directly with a pointer (*ptr16:16* or *ptr16:32*) or indirectly with a memory location (*m16:16* or *m16:32*). The operand-size attribute determines the size of the offset (16 or 32 bits) in the far address. The new code segment selector and its descriptor are loaded into CS register, and the offset from the instruction is loaded into the EIP register.

CALL—Call Procedure (Continued)

```
TASK-GATE:
    IF task gate DPL < CPL or RPL
        THEN #GP(task gate selector);
    FI;
    IF task gate not present
        THEN #NP(task gate selector);
    FI;
    Read the TSS segment selector in the task-gate descriptor;
    IF TSS segment selector local/global bit is set to local
        OR index not within GDT limits
            THEN #GP(TSS selector);
    FI;
    Access TSS descriptor in GDT;

    IF TSS descriptor specifies that the TSS is busy (low-order 5 bits set to 00001)
            THEN #GP(TSS selector);
    FI;
    IF TSS not present
        THEN #NP(TSS selector);
    FI;
    SWITCH-TASKS (with nesting) to TSS;
    IF EIP not within code segment limit
        THEN #GP(0);
    FI;
END;

TASK-STATE-SEGMENT:
    IF TSS DPL < CPL or RPL
    OR TSS descriptor indicates TSS not available
        THEN #GP(TSS selector);
    FI;
    IF TSS is not present
        THEN #NP(TSS selector);
    FI;
    SWITCH-TASKS (with nesting) to TSS
    IF EIP not within code segment limit
        THEN #GP(0);
    FI;
END;
```

Flags Affected

All flags are affected if a task switch occurs; no flags are affected if a task switch does not occur.

RET—Return from Procedure

Opcode	Instruction	Description
C3	RET	Near return to calling procedure
CB	RET	Far return to calling procedure
C2 *iw*	RET *imm16*	Near return to calling procedure and pop *imm16* bytes from stack
CA *iw*	RET *imm16*	Far return to calling procedure and pop *imm16* bytes from stack

Description

Transfers program control to a return address located on the top of the stack. The address is usually placed on the stack by a CALL instruction, and the return is made to the instruction that follows the CALL instruction.

The optional source operand specifies the number of stack bytes to be released after the return address is popped; the default is none. This operand can be used to release parameters from the stack that were passed to the called procedure and are no longer needed. It must be used when the CALL instruction used to switch to a new procedure uses a call gate with a non-zero word count to access the new procedure. Here, the source operand for the RET instruction must specify the same number of bytes as is specified in the word count field of the call gate.

The RET instruction can be used to execute three different types of returns:

- Near return—A return to a calling procedure within the current code segment (the segment currently pointed to by the CS register), sometimes referred to as an intrasegment return.
- Far return—A return to a calling procedure located in a different segment than the current code segment, sometimes referred to as an intersegment return.
- Inter-privilege-level far return—A far return to a different privilege level than that of the currently executing program or procedure.

The inter-privilege-level return type can only be executed in protected mode. See the section titled “Calling Procedures Using Call and RET” in Chapter 6 of the *IA-32 Intel Architecture Software Developer’s Manual, Volume 1*, for detailed information on near, far, and inter-privilege-level returns.

When executing a near return, the processor pops the return instruction pointer (offset) from the top of the stack into the EIP register and begins program execution at the new instruction pointer. The CS register is unchanged.

When executing a far return, the processor pops the return instruction pointer from the top of the stack into the EIP register, then pops the segment selector from the top of the stack into the CS register. The processor then begins program execution in the new code segment at the new instruction pointer.

RET—Return from Procedure (Continued)

The mechanics of an inter-privilege-level far return are similar to an intersegment return, except that the processor examines the privilege levels and access rights of the code and stack segments being returned to determine if the control transfer is allowed to be made. The DS, ES, FS, and GS segment registers are cleared by the RET instruction during an inter-privilege-level return if they refer to segments that are not allowed to be accessed at the new privilege level. Since a stack switch also occurs on an inter-privilege level return, the ESP and SS registers are loaded from the stack.

If parameters are passed to the called procedure during an inter-privilege level call, the optional source operand must be used with the RET instruction to release the parameters on the return. Here, the parameters are released both from the called procedure's stack and the calling procedure's stack (that is, the stack being returned to).

Operation

```
(* Near return *)
IF instruction     near return
    THEN;
        IF OperandSize     32
            THEN
                IF top 12 bytes of stack not within stack limits THEN #SS(0); FI;
                EIP     Pop();
            ELSE (* OperandSize     16 *)
                IF top 6 bytes of stack not within stack limits
                    THEN #SS(0)
                FI;
                tempEIP     Pop();
                tempEIP     tempEIP AND 0000FFFFH;
                IF tempEIP not within code segment limits THEN #GP(0); FI;
                EIP     tempEIP;
        FI;
    IF instruction has immediate operand
        THEN IF StackAddressSize=32
            THEN
                ESP     ESP + SRC; (* release parameters from stack *)
            ELSE (* StackAddressSize=16 *)
                SP     SP + SRC; (* release parameters from stack *)
        FI;
    FI;

(* Real-address mode or virtual-8086 mode *)
IF ((PE     0) OR (PE     1 AND VM     1)) AND instruction     far return
    THEN;
```

RET—Return from Procedure (Continued)

```
        ELSE (* OperandSize=16 *)
            EIP     Pop();
            EIP     EIP AND 0000FFFFH;
            CS      Pop(); (* 16-bit pop; segment descriptor information also loaded *)
            CS(RPL)     CPL;
            ESP     ESP + SRC; (* release parameters from called procedure's stack *)
            tempESP     Pop();
            tempSS      Pop(); (* 16-bit pop; segment descriptor information also loaded *)
             (* segment descriptor information also loaded *)
            ESP     tempESP;
            SS      tempSS;
    FI;
    FOR each of segment register (ES, FS, GS, and DS)
        DO;
            IF segment register points to data or non-conforming code segment
            AND CPL > segment descriptor DPL; (* DPL in hidden part of segment register *)
                THEN (* segment register invalid *)
                    SegmentSelector     0; (* null segment selector *)
            FI;
        OD;
    For each of ES, FS, GS, and DS
    DO
        IF segment selector index is not within descriptor table limits
            OR segment descriptor indicates the segment is not a data or
                readable code segment
            OR if the segment is a data or non-conforming code segment and the segment
                descriptor's DPL < CPL or RPL of code segment's segment selector
                THEN
                    segment selector register     null selector;
    OD;
    ESP     ESP + SRC; (* release parameters from calling procedure's stack *)
```

Flags Affected

None.

Protected Mode Exceptions

#GP(0)	If the return code or stack segment selector null.
	If the return instruction pointer is not within the return code segment limit
#GP(selector)	If the RPL of the return code segment selector is less then the CPL.
	If the return code or stack segment selector index is not within its descriptor table limits.
	If the return code segment descriptor does not indicate a code segment.

```
; File: subroutine.asm
;
; example of subroutines in assembly language.

%define STDOUT 1
%define SYSCALL_EXIT  1
%define SYSCALL_WRITE 4

        SECTION .data                   ; initialized data section

msg1:   db "Hello World", 10, 0         ; C-style \0 term. string

msg2:   db "Good-bye, blue sky", 10, 0


        SECTION .text                   ; Code section.
        global  _start                  ; let loader see entry point

_start: nop                             ; Entry point.
pstart:                                 ; address for gdb

        mov     eax, msg1               ; print first string
        call    print

        mov     eax, msg2               ; print second string
        call    print


        ; final exit
        ;
pexit:  mov     eax, SYSCALL_EXIT       ; exit function
        mov     ebx, 0                  ; exit code, 0=normal
        int     080h                    ; ask kernel to take over
```

```
; Subroutine print
; writes null-terminated string with address in eax
;
print:
        ; find \0 character and count length of string
        ;
        mov     edi, eax                ; use edi as index
        mov     edx, 0                  ; initialize count

count:  cmp     [edi], byte 0           ; null char?
        je      end_count
        inc     edx                     ; update index & count
        inc     edi
        jmp     short count
end_count:

        ; make syscall to write
        ; edx already has length of string
        ;
        mov     ecx, eax                ; Arg3: addr of message
        mov     eax, SYSCALL_WRITE      ; write function
        mov     ebx, STDOUT             ; Arg1: file descriptor
        int     080h                    ; ask kernel to write
        ret

; end of subroutine
```

```
linux3% gdb a.out
GNU gdb 19991004
Copyright 1998 Free Software Foundation, Inc.

(gdb) disas *pstart
Dump of assembler code for function pstart:
0x8048081 <pstart>:     mov    %eax,0x80490c0
0x8048086 <pstart+5>:   call   0x80480a1 <print>
0x804808b <pstart+10>:  mov    %eax,0x80490cd
0x8048090 <pstart+15>:  call   0x80480a1 <print>
0x8048095 <pexit>:      mov    %eax,0x1
0x804809a <pexit+5>:    mov    %ebx,0x0
0x804809f <pexit+10>:   int    0x80
End of assembler dump.

(gdb) break *pstart
Breakpoint 1 at 0x8048081
(gdb) break *print
Breakpoint 2 at 0x80480a1

(gdb) run
Starting program: /afs/umbc.edu/users/c/h/chang/home/asm/sub/a.out

Breakpoint 1, 0x8048081 in pstart ()
(gdb) print/x $esp
$1 = 0x7ffffb90
(gdb) cont
Continuing.

Breakpoint 2, 0x80480a1 in print ()
(gdb) print/x $esp
$2 = 0x7ffffb8c
(gdb) x/1wx $esp
0x7ffffb8c:     0x0804808b
```

```
(gdb) cont
Continuing.
Hello World

Breakpoint 2, 0x80480a1 in print ()
(gdb) print/x $eax
$3 = 0x80490cd
(gdb) x/20cb &msg2
0x80490cd <msg2>:       71 'G'  111 'o' 111 'o' 100 'd' 45 '-'  98
'b'  121 'y' 101 'e'
0x80490d5 <msg2+8>:     44 ','  32 ' '  98 'b'  108 'l' 117 'u' 101
'e' 32 ' '  115 's'
0x80490dd <msg2+16>:    107 'k' 121 'y' 10 '\n' 0 '\000'
(gdb) x/1wx $esp
0x7ffffb8c:     0x08048095

(gdb) cont
Continuing.
Good-bye, blue sky

Program exited normally.
(gdb) quit
linux3% exit
```

```
; File: recursive.asm
;
; example of subroutines in assembly language.

%define STDOUT 1
%define SYSCALL_EXIT  1
%define SYSCALL_WRITE 4

        SECTION .data                   ; initialized data section

msg1:   db "Hello World", 10, 0         ; C-style \0 terminated
string

msg2:   db 10, "Good-bye, blue sky", 10, 0

char:   db 0, 0                         ; single char followed by \(


        SECTION .text                   ; Code section.
        global  _start                  ; let loader see entry point

_start: nop                             ; Entry point.
pstart:                                 ; address for gdb

        mov     eax, msg1               ; print first string
        call    print

        mov     al, '5'
        call    recurse

        mov     eax, msg2               ; print second string
        call    print


        ; final exit
        ;
pexit:  mov     eax, SYSCALL_EXIT       ; exit function
        mov     ebx, 0                  ; exit code, 0=normal
        int     080h                    ; ask kernel to take over
```

```
; A recursive subroutine
; counts down to '0'
; parameter stored in register al

recurse:
        cmp     al, '0'                 ; don't go below '0'
        jae     rcont
        ret                             ; go back

rcont:  push    ax                      ; save al
        dec     al                      ; param for recursive call
        call    recurse                 ; recursively count down
        pop     ax                      ; restore count
        mov     [char], al              ; prepare string for printi
        mov     eax, char               ; param for print subrout.
        call    print
        ret


; Subroutine print
; writes null-terminated string with address in eax
;
print:
        ; find \0 character and count length of string
        ;
        mov     edi, eax                ; use edi as index
        mov     edx, 0                  ; initialize count

count:  cmp     [edi], byte 0           ; null char?
        je      end_count
        inc     edx                     ; update index & count
        inc     edi
        jmp     short count
end_count:

        ; make syscall to write
        ; edx already has length of string
        ;
        mov     ecx, eax                ; Arg3: addr of message
        mov     eax, SYSCALL_WRITE      ; write function
        mov     ebx, STDOUT             ; Arg1: file descriptor
        int     080h                    ; ask kernel to write
        ret

; end of subroutine
```

```
linux3% nasm -f elf recurse.asm
linux3% ld recurse.o
linux3%

linux3% a.out
Hello World
012345
Good-bye, blue sky
linux3%
```

Next Time

- Linux/gcc/i386 function call convention