COSC 458-647
Application Software Security

Assembly Language

Basic Instructions

IA32 Instruction Format

- General format:

```
[prefix] instruction operands
```

- Prefix used only in String Functions

- Operands represent the direction of operands
  - Single operand instruction: XXX `<src>`
  - Two operand instruction: XXX `<dest>` `<src>`
    - XXX represents the instruction opcode
    - src & dest represent the source and destination operands respectively

Instruction Format

- General format:

  `[prefix] instruction operands`

- Prefix used only in String Functions

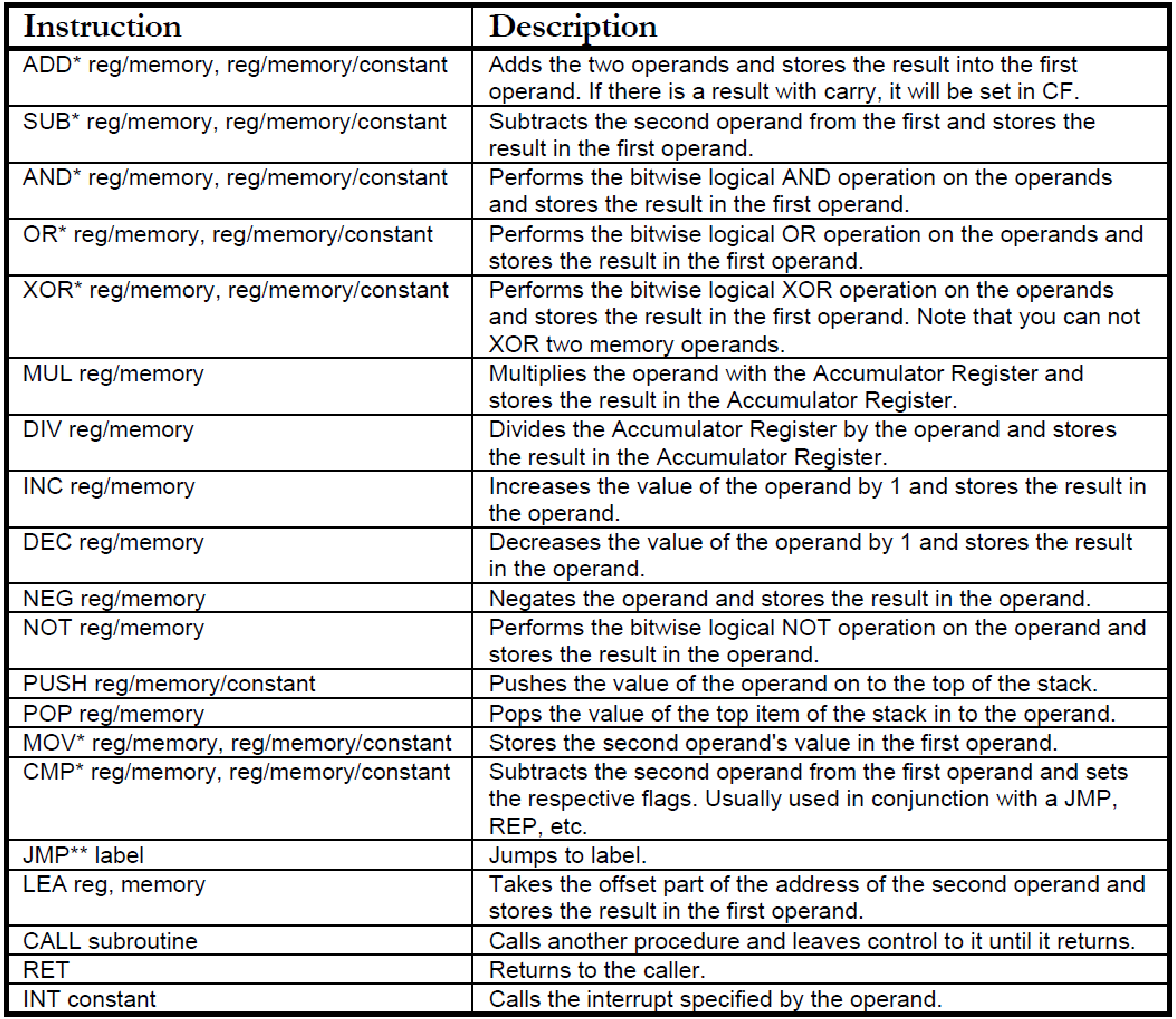
- Operands represent the direction of operands
  - Single operand instruction: **`XXX <src>`**
  - Two operand instruction: **`XXX <dest> <src>`**
    - `XXX` represents the instruction opcode
    - **`src`** & **`dest`** represent the source and destination operands respectively

IA32 Instruction Format (cont’d)

- Source operands can be:
  - Register/Memory reference/Immediate value

- Destination operands can be:
  - Register/Memory reference

- Note:
  - The Intel CPU does NOT allow both source and destination operands to be memory references



Instruction	Description
ADD* reg/memory, reg/memory/constant	Adds the two operands and stores the result into the first operand. If there is a result with carry, it will be set in CF.
SUB* reg/memory, reg/memory/constant	Subtracts the second operand from the first and stores the result in the first operand.
AND* reg/memory, reg/memory/constant	Performs the bitwise logical AND operation on the operands and stores the result in the first operand.
OR* reg/memory, reg/memory/constant	Performs the bitwise logical OR operation on the operands and stores the result in the first operand.
XOR* reg/memory, reg/memory/constant	Performs the bitwise logical XOR operation on the operands and stores the result in the first operand. Note that you can not XOR two memory operands.
MUL reg/memory	Multiplies the operand with the Accumulator Register and stores the result in the Accumulator Register.
DIV reg/memory	Divides the Accumulator Register by the operand and stores the result in the Accumulator Register.
INC reg/memory	Increases the value of the operand by 1 and stores the result in the operand.
DEC reg/memory	Decreases the value of the operand by 1 and stores the result in the operand.
NEG reg/memory	Negates the operand and stores the result in the operand.
NOT reg/memory	Performs the bitwise logical NOT operation on the operand and stores the result in the operand.
PUSH reg/memory/constant	Pushes the value of the operand on to the top of the stack.
POP reg/memory	Pops the value of the top item of the stack in to the operand.
MOV* reg/memory, reg/memory/constant	Stores the second operand's value in the first operand.
CMP* reg/memory, reg/memory/constant	Subtracts the second operand from the first operand and sets the respective flags. Usually used in conjunction with a JMP, REP, etc.
JMP** label	Jumps to label.
LEA reg, memory	Takes the offset part of the address of the second operand and stores the result in the first operand.
CALL subroutine	Calls another procedure and leaves control to it until it returns.
RET	Returns to the caller.
INT constant	Calls the interrupt specified by the operand.

Memory References

- Same as a C/C++ pointer access

- Pointer operands appear within square brackets e.g.
  - **`MOV EAX, [0x00040222h]`**

Contents of memory location 0x00040222

  - Can also have register names instead of hex addresses
    - e.g. **`MOV EAX, [EBX]`**

Refers to the contents of register EBX

Memory References (cont’d)

- Control the size of memory accessed by preceding the memory reference with a size:
  - BYTE PTR: byte access
  - WORD PTR: two byte access
  - DWORD PTR: four byte access

```
mov BYTE PTR [ebx], 2       ; Move 2 into the single byte at the address stored in EBX.

mov WORD PTR [ebx], 2       ; Move the 16-bit integer representation of 2 into the 2
                            bytes starting at the address in EBX.

mov DWORD PTR [ebx], 2      ; Move the 32-bit integer representation of 2 into the 4
                            bytes starting at the address in EBX.
```

Memory References (cont’d)

- Invalid Memory Accesses
    - Accessing illegal memory
        - CPU generates a general protection fault: (GPF)
    - Access a memory location that does not exist
        - CPU generates a page fault

NOP

- NOP – No operation
  - Takes no arguments
  - Commonly used by the compiler as padding INSIDE functions so as keep them properly aligned

```
var1:  dd    0FFh
str1:  db    “my dog has fleas”,10
var2:  dd    0

; Here are some simple instructions
             mov     eax, [var1] ; notice the brackets
             mov     edx, str1   ; notice the not brackets
             call    dspmsg
             jmp     done
             mov     ebx, [var2] ; this will never happen
             cmp     ecx, 0x8 ; this also will never happen
done:  nop
```

Stack Manipulation Instructions

- PUSH <argument>
  - Pushes a word/double word on the stack
  - Argument can be a register/memory location/immediate value (hardcoded number)
  - E.g. `push eax;   push msg;   push dword 0x9`

- POP <argument>
  - Pop a word/double word from the stack
  - E.g. `pop ecx; pop ebx`

- Note:
  - PUSH decrements the ESP while POP increments the ESP

Stack Manipulation Instructions (cont’d)

- PUSHAD
  - Push (save) all general purpose registers

- POPAD
  - Pop (restore) all general purpose registers

- Avoids long sequence of PUSH/POP instructions to save/restore the registers
  - Used mainly in system code

Example

```
//Swap the EAX and EBX values. The sequence
//gives you an idea of how it could be done.

            PUSH EAX
            PUSH EBX

            POP EAX
            POP EBX
```

Arithmetic Instructions

- ADD <dest> <src>: <dest> ← <dest> + <src>

```
add <reg>,<reg>
add <reg>,<mem>
add <mem>,<reg>
add <reg>,<con>
add <mem>,<con>
```

E.g.

```
add eax, 10                ;;— EAX ← EAX + 10
add BYTE PTR [var], 10     ;; add 10 to the single byte stored at memory address var
```

- SUB <dest> <src>: <dest> ← <dest> - <src>

```
sub <reg>,  <reg>
sub <reg>,  <mem>
sub <mem>,  <reg>
sub <reg>,  <con>
sub <mem>,  <con>
```

E.g.

```
sub al, ah         ;; AL ← AL - AH
sub eax, 216       ;; subtract 216 from the value stored in EAX
```

INC, DEC - Increment, Decrement

- The inc instruction increments the contents of its operand by one. The dec instruction decrements the contents of its operand by one.
- Syntax

```
inc <reg>
inc <mem>
dec <reg>
dec <mem>
```

- Examples

```
dec eax                ;;subtract one from the contents of EAX.
inc DWORD PTR [var] ;; add one to the 32-bit integer stored at location var
```

Arithmetic Instructions (cont’d)

- DIV/MUL: Unsigned Division/Multiplication
  - Uses the EDX register to store the high bytes of double-word and higher (64 bit) results.
  - EAX stores the low bytes

- IDIV/IMUL: Signed Division/Multiplication
  - IMUL sometimes has 3 operands:
    - IMUL <dest> <src1> <src2>

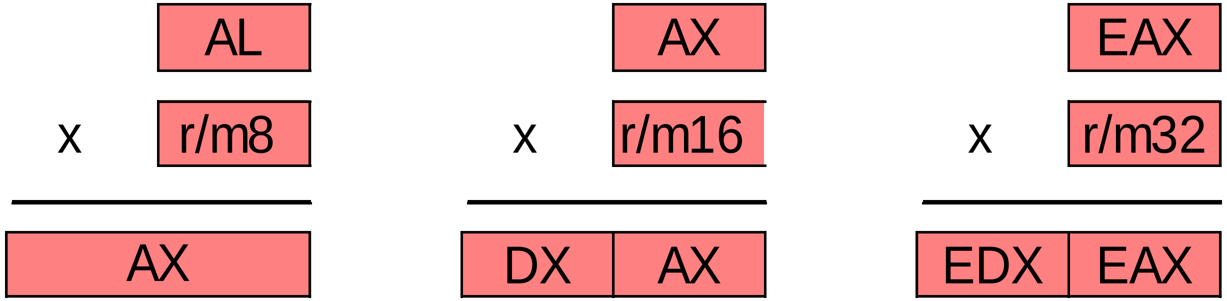
MUL Instruction - (Unsigned Multiply)

- Multiplies an 8-, 16-, or 32-bit operand by either AL, AX or EAX.

```
MUL r/m8
MUL r/m16
MUL r/m32
```

MUL Instruction

- Note that the product is stored in a register (or group of registers) twice the size of the operands.
- The operand can be a register or a memory operand



AL
x r/m8
AX

AX
x r/m16
DX AX

EAX
x r/m32
EDX EAX

MUL Examples

```
mov   AL, 5h
mov   BL, 10h
mul   BL
```

AX stores AL*BL ; AX = 0050h, CF = 0

(no overflow - the Carry flag is 0 because the upper half of AX is zero)

MUL Examples

```
.data
  val1  WORD 2000h
  val2  WORD 0100h
.code
  mov  AX, val1
  mul  val2                    ; DX:AX = 00200000h, CF = 1 (CF is 1 because DX is not zero)
```

12345h * 1000h, using 32-bit operands:

```
mov    EAX,    12345h
mov    EBX,    1000h
mul    EBX ;                   ; EDX:EAX = 0000000012345000h, CF=0
```

Your turn

- What will be the hexadecimal values of DX, AX, and the Carry flag after the following instructions execute?

```
mov   AX,    1234h
mov   BX,    100h
mul   BX
```

Pointer Manipulation Instructions

- LEA: Load Effective Address
  - LEA <dest> <src>
  - Loads the destination register with the address of the source operand
  - Used to emulate pointer access

Example

```
//...
int * pInt ;
int iVal ;

// The following instruction sequence is identical to the
// C code: pInt = &iVal ;
  LEA EAX , iVal
  MOV [pInt] , EAX

//..
```

(Example from Debugging Applications by John Robbins)

Another example

```
//....
char szBuff [ MAX_PATH ] ;

// Another example of accessing a pointer through LEA.
// This is identical to the C code:
// GetWindowsDirectory ( szBuff , MAX_PATH ) ;

   PUSH 104h             // Push MAX_PATH as the second parameter.
   LEA  ECX, szBuff      // Get the address of szBuff.
   PUSH ECX              // Push the address of szBuff as the first parameter.
   CALL DWORD PTR            [GetWindowsDirectory]
```

Function Call Instruction

- CALL <argument>
  - argument can be a register \ memory reference \ parameter \ global offset
  - Automatically pushes the return address on the stack and decrements ESP

Function Return Instruction

- RET <optional argument>
  - Argument says how many bytes to pop off the stack (to account for parameters passed to the function)
  - Pops the callers return address off the top of stack and put it in the instruction pointer
    - Return address validity is NOT checked!!!: potential security hazard

Data Manipulation Instructions

- AND <dest> <src> : logical AND
- OR <dest> <src> : logical OR
- NOT <arg>: logical NOT
  - One’s complement negation (Bit Flipping)
- NEG <arg>:
  - Two’s complement negation

Data Manipulation Instructions (cont’d)

- XOR <dest> <src>: logical XOR
  - Fastest way to zero out a register!!!

- INC/DEC <arg> : increment/decrement
  - Often used in speed optimized code (executes in single clock cycle)
  - Directly maps to the C++ operators:
    - ++ : INC
    - -- : DEC

Data Manipulation Instructions (cont’d)

- SHL/SHR <arg> : shift left and Shift right
  - SHL: fastest way to multiply by 2 (C++: <<)
  - SHR: fastest way to divide by 2 (C++: >>)
- MOVZX <dest> <src>: move with zero extend
- MOVSX <dest> <src>: Move with sign extend

Compare Instruction

- CMP <arg1> <arg2>
  - compare arg1 and arg2 and set the appropriate conditional flags in the EFLAGS register

Test Instruction

- TEST <arg1> <arg2> : Bitwise AND of both arguments and sets the appropriate conditional flags
  - PL (SF)
  - ZR (ZF)
  - PE (PF)

Jump Instructions

- JE <label> : Jump if equal
- JL <label> : Jump if less than
- JG <label> : Jump if greater than
- JNE <label> : Jump if not equal to
- JGE <label> : Jump if greater than or equal to
- JLE <label> : Jump if Less than or equal to

Jump Instructions (Cont’d)

- Always follow a CMP/TEST instruction
- JMP condition is always the opposite of the original conditional

Loop Instruction

- Loop <label>
  - Decrement ECX and if ECX isn’t 0, go and re-execute the code sequence marked by <label>

- Rarely used by the VS.NET compiler

Function Calling Conventions

- Specifies how parameters are passed to a function
  - Passed in through stack/registers

- Specifies how stack cleanup occurs upon function return
  - Who performs the cleanup, the caller or the callee?

(Supplied handout has table summarizing the various calling conventions)

Instruction usage examples

- Discuss usage of the previously mentioned instructions
- Generated by VS.NET during compilation
- Examples discussed:
    - Function Entry and Exit
    - Global variable, Local variable and Function parameter access

Function Entry (Prolog)

- Compiler generated at the beginning of a function (before the actual processing code of the function)

- This code sets up the stack for access to the function’s local variables and parameters (the Stack Frame)

Prolog Example

```
// Standard prolog setup
  PUSH EBP            // Save the stack frame register.
  MOV  EBP, ESP       // Set the local function stack frame to ESP.
  SUB  ESP , 20h      // Make room on the stack for 0x20 bytes of local
                      // variables. The SUB instruction appears only if the
                      // function has local variables.
```

Function Exit (Epilog)

- Compiler generated (after the end of the processing code of the function)
- Undoes the operations of the prolog
    - Stack cleanup can be performed here

```
// Standard epilog teardown
MOV ESP , EBP   // Restore the stack value.
POP EBP       // Restore the saved stack frame register.
```

Backup slides

Read a byte from stdin

```
; read a byte from stdin
      mov    eax, 3 ; sys_read system call
      mov    ebx, 0 ; read from standard input
      mov    ecx, variable ; address to pass to
      mov    edx, 1 ; input length
      int    0x80 ; call the kernel

; write a byte to stdout
      mov    eax, 4 ; sys_write system call
      mov    ebx, 1 ; write to standard output
      mov    ecx, variable ; memory address
      mov    edx, 1 ; output length
      int    0x80 ; call the kernel

; quit the program
      mov    eax, 1 ; sys_exit system call
      mov    ebx, 0 ; sys_exit system call
      int    0x80 ; call the kernel
```