CSE 351 – Final Exam – Spring 2016
June 8, 2016

Solution

Name: ____________________________
UWNetID: ____________________________

Please do not turn the page until 2:30.

Instructions

• The exam is closed book, closed notes, no calculators, no mobile phones, no laptops.
• You will have 1 hour and 50 minutes. Please stop promptly at 4:20 (or when asked).
• There are 160 points total, divided unevenly among 8 problems (each with multiple parts).
• The exam is printed double-sided. If you separate any pages, be sure to print your name at the top of each separated page so we can match them up.
• Useful reference material can be found on the last few pages of the exam. Feel free to tear it off.

Advice

• Read questions carefully before starting. If you do not understand a question or are unsure of what is expected for an answer, please ask!
• Write down thoughts and intermediate steps so you can get partial credit. But clearly indicate what is your final answer.
• Questions are not necessarily in order of difficulty; some can be done relatively quickly if you know the answer, while others take more time. So make sure you get to all the questions.
• Relax. Take your time, think critically, and use the provided reference material to your advantage.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Structs &amp; Data Layout</td>
<td>30</td>
</tr>
<tr>
<td>2. Optimizing for Big Data</td>
<td>15</td>
</tr>
<tr>
<td>3. Address Translation</td>
<td>25</td>
</tr>
<tr>
<td>4. Operation: Cache Is King</td>
<td>20</td>
</tr>
<tr>
<td>5. Java of the C</td>
<td>10</td>
</tr>
<tr>
<td>6. Programs, processes, and processors (oh my!)</td>
<td>25</td>
</tr>
<tr>
<td>7. Bug Zapper</td>
<td>15</td>
</tr>
<tr>
<td>8. Miscellaneous</td>
<td>20</td>
</tr>
</tbody>
</table>
1. Structs & Data Layout (30 pts)

Use the following snippet of C code to answer the questions for this problem.

```c
typedef struct Item {
    char * name;
    int quantity;
    double weight;
} Item;

Item * inventory[100];
size_t inventory_size = 0;

void add_to_inventory(Item * item) {
    inventory[inventory_size] = item;
    inventory_size++;
}

int main(int argc, char const *argv[]) {
    int num_groceries = 2;
    Item * groceries = (Item*)malloc(2*sizeof(Item));
    groceries[0].name = "lemonade";
    /* ... */

    Item * favorite = inventory + 1;
    /* ... */
}
```

(a) Each of the following C expressions computes an address. Fill in the table with the C type (some sort of pointer, for example, int*), and the memory region in which the address would be, using the letters from the table above (for example, if the address is in kernel memory, then you would answer "A").

<table>
<thead>
<tr>
<th>C expression</th>
<th>C Type</th>
<th>Memory region (A-G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;inventory_size</td>
<td>size_t*</td>
<td>E</td>
</tr>
<tr>
<td>&amp;groceries[1]</td>
<td>Item*</td>
<td>D</td>
</tr>
<tr>
<td>groceries[0].name</td>
<td>char*</td>
<td>F</td>
</tr>
<tr>
<td>&amp;num_groceries</td>
<td>int*</td>
<td>B</td>
</tr>
<tr>
<td>add_to_inventory (address of function)</td>
<td>void (<em>)(Item</em>)</td>
<td>F</td>
</tr>
</tbody>
</table>

(b) Does Item have any internal or external fragmentation? If so, where is it?

It has internal fragmentation because weight has to be 8-byte aligned, so there are 4 bytes of wasted space between quantity and weight.
(c) Is it possible to reduce the size of Item by reordering its members? If so, what reordering would minimize the size? If not, why not? Explain in about 1 sentence.

No, reordering "int quantity" last would only make the fragmentation be external instead. The overall size of the struct still has to be a multiple of 8.

(d) Fill in the following table with the value of each C expression, using the given initial values. Write "?" if the value cannot be determined. Note: these addresses are fake, they will not help you solve part (a) above, sorry. It may help you to draw a picture of memory (and may help us give partial credit).

<table>
<thead>
<tr>
<th>C expression</th>
<th>Value (hex or decimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;lemonade&quot;</td>
<td>0x2000</td>
</tr>
<tr>
<td>inventory</td>
<td>0x6000</td>
</tr>
<tr>
<td>groceries</td>
<td>0x4000</td>
</tr>
<tr>
<td>sizeof(Item)</td>
<td>24</td>
</tr>
<tr>
<td>&amp;groceries[0].quantity</td>
<td>0x4008                 ← groceries+8</td>
</tr>
<tr>
<td>&amp;groceries[1].weight</td>
<td>0x4028                 ← groceries+24+16</td>
</tr>
<tr>
<td>groceries[0].name + 3</td>
<td>0x2003</td>
</tr>
<tr>
<td>&amp;favorite-&gt;name</td>
<td>0x6018                 ← inventory + 24</td>
</tr>
</tbody>
</table>

(e) Write a function called add_avocado(), with return type void, that adds avocados to the inventory (new item with name: "avocado", quantity: 4, weight: 0.24). Your implementation should use the provided add_to_inventory function. Don’t worry about missing semicolons. See reference sheet for C functions.

```c
void add_avocado() {
    Item * avocado = (Item*)malloc(sizeof(Item));
    avocado->name = "avocado";
    avocado->quantity = 4;
    avocado->weight = 0.24;
    add_to_inventory(avocado);
}
```
2. Optimizing for Big Data (15 pts)

Imagine you are working on a new deep learning system (a "big data" application) that must process a vast number of images in order to learn patterns (such as cute cat pics). However, it is not running as fast as you need it to, so you are trying to improve its performance. The following questions will have you try to optimize the application by changing different aspects of the system.

(a) The application usually operates with batches of many large images at a time. When running a single instance, performance is acceptable, but when running two instances in different processes, memory accesses frequently take several thousand times longer than a typical DRAM (main memory) access should. What is most likely causing this slowdown? What could you change about the hardware to fix this?

The working set is too large to fit in memory, so the different processes are thrashing as they keep trying to page in memory from disk. We could fix this by adding more RAM to the machine. (we know that it can't be a cache problem because it's slower than DRAM, context switching could be a problem, but it wouldn't result in slowdowns this extreme mostly because the OS isn't dumb)

(b) The application allocates and frees lots of 32-byte blocks from the heap. What could you do to make the memory allocator work better for this use case? Say whether your optimization would improve throughput or utilization (or both). Answer does not need to be complete sentences.

The allocator could keep a special free list for 32-byte blocks. This is called a segmented free list. It would improve throughput because we don't have to search to find a block of the right size. (There's also some argument to be made that this will improve utilization by helping get a "best fit" for 32-byte allocations.

(c) When processing images, the application sequentially accesses each element in order with a stride of 1. What single change to the cache geometry would most improve (reduce) the miss rate? (circle one)

✓ A. Increase cache block size. ← Only way to improve spatial locality is to increase block size.
   B. Increase number of sets (and decrease associativity).
   C. Increase associativity (and decrease number of sets).
   D. Switch to write-through and write-no-allocate.

(d) Registers and caches typically use the same technology (SRAM) to store data, whereas main memory is typically implemented out of a slower technology called DRAM.

(i) Why, then, are registers faster than L1 cache hits? Give one (short) reason.

Cache lookups still require address translation. The L1 is also larger and slightly further away, so it takes longer to access.

(ii) If registers and L1 are so fast, why don't we build all of memory this way? Give one (short) reason.

Registers and L1 are only fast because they're small and close (on-chip). Larger SRAM would be slower, far too costly in terms of money and power, and wouldn't fit on the chip.
3. Address Translation (25 pts)

Imagine we have a machine with 16-bit virtual addresses, 12-bit physical addresses, and:
- Page size of 256 bytes.
- Translation lookaside buffer (TLB) with 8 ways and 4 sets.
- One-level cache with capacity of 256 bytes, 16-byte cache block size, and 2-way associativity.

(a) For the virtual address below, label the bits used for each component, either by labelling boxes or with arrows to indicate ranges of bits. *Hint:* there may be more than one label for some bits.

(i) Virtual page offset ("VPO")
(ii) Virtual page number ("VPN")
(iii) TLB index ("index")
(iv) TLB tag ("tag")

(b) How many total page-table-entries are there per process in this system?

256-byte pages = $2^{8}$, $2^{16} / 2^{8} = 2^{8} \text{ VPNs} - 256 \text{ page table entries per process}$

(c) How many *sets* are there in the cache?

256 bytes total / 16 bytes/block = 16 cache blocks total
16 blocks / 2 ways = 8 sets

(d) Assume that the virtual address above has been translated to a physical address in memory.
*Fill in the known bits* of the physical address below, and *label the bits for each component* as you did in part (a) — again, some bits may have more than one label.

(i) Physical page number ("PPN")
(ii) Physical page offset ("PPO")
(iii) Cache index ("index")
(iv) Cache tag ("tag")
(v) Cache offset ("offset")
Choose your own adventure. For this problem, there are two different scenarios. Your job is to number the events according to the order in which they would occur for the given scenario. Leave blank any events that would not occur for that scenario. When you get to a "STOP HERE" in a scenario, then you'll know that adventure has come to an end.

(i) **Page fault (first column).** Memory access resulting in a page fault for an allocated page (stop when the page fault has been handled, but data has not been returned).

(ii) **Page hit (second column).** Memory access from before re-executes, resulting in a page hit now.

<table>
<thead>
<tr>
<th>Order for page fault:</th>
<th>Order for page hit:</th>
<th>Events:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>CPU issues virtual memory access.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CPU translates virtual address to physical address.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Page fault detected.</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>MMU fetches page table entry from memory.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMU constructs physical address using page table entry.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>MMU checks TLB for page table entry and <em>finds it</em>.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>MMU checks TLB for page table entry and <em>does not find it</em>.</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Page fault handler chooses a page to evict and does so.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Load requested page into memory and update page table entry in TLB.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Load requested block into cache.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Lookup physical address in main memory.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Lookup physical address in cache, not found.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lookup physical address on disk.</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Return control to program before original memory access. STOP HERE.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Return control to program after original memory access. STOP HERE.</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Return data to the program. STOP HERE.</td>
</tr>
</tbody>
</table>
4. Operation: Cache Is King (20 pts)

Your mission, should you choose to accept it (please do), is to simulate executing memory accesses on a system with a cache. Assume that the contents of physical memory (in hex) are as shown below:

<table>
<thead>
<tr>
<th>Address</th>
<th>0x00</th>
<th>0x08</th>
<th>0x10</th>
<th>0x18</th>
<th>0x20</th>
<th>0x28</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td>+4</td>
<td>+5</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>01</td>
<td>02</td>
<td>03</td>
<td>04</td>
<td>05</td>
</tr>
<tr>
<td></td>
<td>ca</td>
<td>fe</td>
<td>be</td>
<td>ef</td>
<td>fe</td>
<td>ed</td>
</tr>
<tr>
<td></td>
<td>a0</td>
<td>b0</td>
<td>c0</td>
<td>d0</td>
<td>e0</td>
<td>f0</td>
</tr>
<tr>
<td></td>
<td>0a</td>
<td>0b</td>
<td>0c</td>
<td>0d</td>
<td>0e</td>
<td>0f</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>11</td>
<td>22</td>
<td>33</td>
<td>44</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>ca</td>
<td>fe</td>
<td>be</td>
<td>ef</td>
<td>ed</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>00</td>
<td>ca</td>
<td>fe</td>
<td>be</td>
<td>ef</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>00</td>
<td>ca</td>
<td>fe</td>
<td>be</td>
<td>ef</td>
</tr>
</tbody>
</table>

Now, simulate executing the following memory accesses on a 2-way set-associative cache, with 4-byte cache blocks and 4 sets and an LRU (least recently used) replacement policy; treat each access as a single-byte read (e.g. char read). Fill in the contents of the cache as you go, then fill in the value read and whether or not the access was a hit or a miss. Assume the cache starts empty. If the contents of a cache cell change, then cross out the old value (or erase) and write the new value. You may leave any unmodified cells empty.

<table>
<thead>
<tr>
<th>Access (1-byte):</th>
<th>Value</th>
<th>Hit (H) or Miss (M)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) 0x08</td>
<td>ca</td>
<td>M</td>
</tr>
<tr>
<td>(b) 0x13</td>
<td>d0</td>
<td>M</td>
</tr>
<tr>
<td>(c) 0x19</td>
<td>0b</td>
<td>M</td>
</tr>
<tr>
<td>(d) 0x0b</td>
<td>ef</td>
<td>H</td>
</tr>
<tr>
<td>(e) 0x28</td>
<td>88</td>
<td>M</td>
</tr>
<tr>
<td>(f) 0x29</td>
<td>99</td>
<td>H</td>
</tr>
</tbody>
</table>

**Grading Rubric**

**Accesses (9 pts)**
- Values: ½ pt each
- Hit/Miss: 1 pt each

**Contents of cache (11 pts)**
- For each cache line:
  - 1 pt for tag
  - ½ pt for valid bit
  - 1 pt for data
- 2 pts for correct sets
- 1 pt for conflict/replacement
5. Java of the C (10 pts)

Imagine you prefer programming in Java, but your curmudgeonly boss (or instructor) insists that you use C. You miss Java’s convenience and safety features, so to make your life easier, you start to implement some of them yourself in C. You start by implementing a function, `error`, to serve like an “exception” in Java: it takes a string message as an argument, prints the error message, then exits the program with an error code. Now you are trying to implement an array data structure that behaves similar to Java’s, using the struct declared below. In your implementations, you need not worry about semicolons.

```c
void error(char * message) {
    printf("error: %s", message);
    exit(1);
}
```

```c
typedef struct Array {
    int length;   // number of elements in the array
    int * data;    // pointer to storage for array data
} Array;
```

(a) **Array Constructor.** Create a Java-like constructor in C that creates a new array of integers (using the `Array` struct from above) and initializes it. *Hint:* Java arrays do not have a vtable, but think carefully about everything else that Java does when creating new objects. *See reference sheet for C library functions.*

```c
Array* new_Array(int n) {
    Array* a = (Array*)malloc(sizeof(Array));
    a->length = n;
    a->data = (int*)malloc(n * sizeof(int));
    for (int i = 0; i < n; i++) {
        a->data[i] = 0;
    }
    return a;
}
```

(b) **Array Out Of Bounds.** Implement a safe array access function in C, using the `error` function to exit if there was an attempted out of bounds access.

```c
int get(Array* a, int i) {
    if (i < 0 || i >= a->length) {
        error("out of bounds");
    }
    return a->data[i];
}
```

---

**Grading Rubric**

**Constructor (6 pts)**
- (1) Allocate array struct
- (2) Malloc data (correct size)
- (1) Zero contents
- (1) Set length field
- (1) Return pointer to array (must have been malloc’ed)

**Array Out Of Bounds (4 pts)**
- (1) Lower bound
- (1) Upper bound
- (1) Call error
- (1) Return correct data
6. Programs, processes, and processors (oh my!) (25 pts)

(a) Consider the following C code on the left (running on Linux), then give one possible output of running it. Assume that printf flushes its output immediately.

```c
void oz() {
    char * name = "toto\n";
    printf("dorothy\n");
    if (fork() == 0) {
        name = "wizard\n";
        printf("scarecrow\n");
        fork();
        printf("tinman\n");
        exit(0);
        printf("witch\n");
    } else {
        printf("lion\n");
    }
    printf(name);
}
```

Possible output:

<table>
<thead>
<tr>
<th>dorothy</th>
<th>dorothy</th>
</tr>
</thead>
<tbody>
<tr>
<td>scarecrow</td>
<td>lion</td>
</tr>
<tr>
<td>tinman</td>
<td>toto</td>
</tr>
<tr>
<td>tinman</td>
<td>scarecrow</td>
</tr>
<tr>
<td>lion</td>
<td>tinman</td>
</tr>
<tr>
<td>toto</td>
<td>tinman</td>
</tr>
</tbody>
</table>

(b) "Pay no attention to the man behind the curtain." We have seen several different mechanisms used to create illusions or abstractions for running programs:

A. Context switch  
B. Virtual memory  
C. Virtual method tables (vtables)  
D. Caches  
E. Timer interrupt  
F. Stack discipline  
G. None of the above, or impossible.

For each of the following, indicate which mechanism above (A-F) enables the behavior, or G if the behavior is impossible or untrue.

(i) ____  Allows operating system kernel to run to make scheduling decisions.

(ii) _____  Prevents buffer overflow exploits.

(iii) _____  Allows multiple instances of the same program to run concurrently.

(iv) _____  Lets programs use more memory than the machine has.

(v) _____  Makes recently accessed memory faster.

(vi) _____  Multiple processes appear to run concurrently on a single processor.

(vii) _____  Enables programs to run different code depending on an object's type.

(viii)_____  Allows an x86-64 machine to execute code for a different ISA.
(c) Give an example of a **synchronous** exception, what could trigger it, and where the exception handler would return control to in the original program.

**Page fault**: triggered by access to virtual address not in memory, returns to the instruction that caused the fault.

**Trap**: used to for syscalls to do something protected by the kernel, returns to after the calling instruction.

(d) In what way does address translation (virtual memory) help make `exec` fast? Explain in less than 2 sentences. *Hint*: it may help to write down what happens during `exec`.

*Address translation is a form of *indirection*, it allows us to implement `fork` without copying the whole process's memory, and `exec` without loading the whole program into memory at once.*

(e) Which of the following **can** a running process determine, assuming it does **not** have access to a timer? *(check all that apply)*

- [X] Its own process ID
- [ ] Size of physical memory
- [X] Size of the virtual address space
- [ ] L1 cache associativity
- [ ] When context switches happen

(f) For each of the following, fill in what is responsible for making the decision: hardware ("HW"), operating system ("OS"), or program ("P").

(i) **OS**  Which physical page a virtual page is mapped to.

(ii) **HW**  Which cache line is evicted for a conflict in a set-associative cache.

(iii) **OS**  Which page is evicted from physical memory during a page fault.

(iv) **HW**  Translation from virtual address to physical address.

(v) **P**  Whether data is stored in the stack or the heap.

(vi) **P**  Data layout optimized for spatial locality
7. Bug Zapper (15 pts)

In this problem, it will be your job to hunt down the bug or performance problem in each snippet of x86 or C code. Indicate the line that causes the error by circling it, then give a brief explanation (less than a sentence); something like: "wrong size" or "this should only be called once"). If there is no bug, then write "No problems."

(a) void shell_exec(char* command, char* args[]) {
    if (fork() == 0) {
        execv(command, args);
        int status;
        wait(&status); ← Wait should be called from the parent. This will never get called.
    }
}

(b) void work() {
    char * buf = (char*)malloc(sizeof(char) * 8);
    gets(buf); ← Gets should never be used because it is unsafe.
    do_something(buf);
    free(buf);
}

(c) void iterate() {
    int * a = (int*)malloc(sizeof(int) * 4);
    for (int i=0; i < 4; i++) {
        *a = 0;
        a += 1;
    }
    free(a); ← Not a pointer to the beginning of a malloc'd block.
}

(d) convert():
    movsbsq (%rdi), %ax ← Wrong size register.
    ret

(e) Indicate where the problem is in the x86-64 translation of this C function.

```
long sum(long * array, long n) {
    long total = 0;
    for (long i=0; i < n; i++) {
        total += array[i];
    }
    return total;
}
```

```
sum(long*, long):
    movl    $0, %edx
    movl    $0, %eax
    .L3:
    cmpq    %rsi, %rdx
    jl     .L2
    addq    (%rdi,%rdx,8), %rax
    addq    $1, %rdx
    jmp     .L3
    .L2:
    rep ret
```
(f) Indicate where the problem is in the x86-64 translation of this C function.

```c
void foo(int * x) {
    return *x + 1;
}
```

```c
foo(int*):
    Should be movl → leal (%rdi), %eax
    addl $1, %eax
    ret
```

(g) Performance debugging. For this one, rewrite the row_sums function to improve performance.

```c
int N = 8192;
int big_data[N][N];

void row_sums(int * sums) {
    for (int i=0; i < N; i++) {
        sums[i] = 0;
    }
    for (int i=0; i < N; i++) {
        for (int j=0; j < N; j++) {
            sums[j] += big_data[j][i];
        }
    }
}
```

```c
void row_sums(int * sums) {
    for (int i=0; i < N; i++) {
        // eliminates risk of conflicts by using register!
        int total = 0;
        for (int j=0; j < N; j++) {
            total += big_data[i][j];
        }
        sums[i] = total;
    }
}
```
8. Miscellaneous (20 pts)

(a) Which of the following are valid reasons why virtual memory pages should be larger than cache blocks? (check all that apply)

☐ Physical memory is always smaller than virtual memory.
☐ Otherwise page tables would not fit in memory.
☒ It takes much longer to access disk than memory.
☒ The TLB typically holds fewer entries than the cache.

(b) Which of the following can only happen for a mis-aligned memory access (such as a movq to an address that is not a multiple of 8)? (check all that apply)

☒ Load 2 cache lines for one memory access.
☐ Stack smashing.
☐ Misaligned address exception leading to an abort.
☒ A TLB hit and a page fault for a single access.

(c) Which of the following buffer overflow attacks are possible on x86-64 in Linux? (check all that apply)

☒ Forcing the program to execute an arbitrary function in the binary.
☐ Executing new code inserted on the stack.
☐ Preventing the operating system from running.
☐ Reading another process’ memory.

(d) How large are x86-64 instructions (in bytes)?

1-16 bytes (also would accept variable-length)

(e) When would write-non-allocate be a good choice for a memory access? (circle one)

A. When initializing large data structures.
B. When data fits in the L1 cache.
C. When writing a location after reading it.
D. When writing a location without reading it. ← No temporal locality
E. Never.

(f) What does the translation lookaside buffer hold? (circle one)

A. Physical page offsets.
B. Physical to virtual page mappings.
C. Virtual to physical page mappings. ← Page table entries map virtual page numbers to physical
D. Recently accessed cache lines.
E. None of the above.
(g) Java arrays can hold objects of different sizes. How do they accomplish this? Answer in one sentence.

Java arrays hold references to objects on the heap rather than the objects themselves.

(h) In a first-fit, explicit free list allocator, coalescing is important because it: (check all that are true)

- Reduces fragmentation.
- Increases memory utilization.
- Reduces search time when allocating.
- Prevents double-freeing.

(i) Garbage collectors work by: (circle one)

- Compiler checking that all allocations are matched with a corresponding free.
- Cleaning up objects that have not been accessed in a given amount of time.
- Freeing variables implicitly when the function returns.
- Iterating over all objects and freeing any that are no longer reachable.

(j) Which of the following is dictated by the x86-64 instruction set architecture?

- Size of registers.
- Size of physical addresses.
- Which register to use for return values.
- Which register holds the address of the instruction about to be executed.
- How many levels of cache there are.

(k) The JVM and x86 are both forms of "instruction set architectures". For each of the following, label if it is true for JVM, x86, or Both:

- JVM  Values are typed (e.g. you can tell if it is signed or unsigned)
- Both  Instructions are encoded in a compact binary representation.
- Both  Source language must be compiled down to lower-level instructions.
- x86   Has a fixed number of named registers.
- Both  Execution involves repeatedly interpreting and running the next instruction.
- Both  Programs are portable among multiple different implementations of the ISA.
- JVM  Has special instructions for getting fields of objects or structs.
References

Powers of 2

| $2^0 = 1$ | $2^{-1} = 0.5$ |
| $2^2 = 4$ | $2^{-2} = 0.25$ |
| $2^3 = 8$ | $2^{-3} = 0.125$ |
| $2^4 = 16$ | $2^{-4} = 0.0625$ |
| $2^6 = 64$ | $2^{-5} = 0.03125$ |
| $2^8 = 256$ |
| $2^{10} = 1024$ |

Hex Conversions

<table>
<thead>
<tr>
<th>Hex</th>
<th>Binary</th>
<th>Decimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0xA$</td>
<td>1010</td>
<td>10</td>
</tr>
<tr>
<td>$0xa$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0xF$</td>
<td>1111</td>
<td>15</td>
</tr>
<tr>
<td>$0x10$</td>
<td>0010 0000</td>
<td>16</td>
</tr>
<tr>
<td>$0x20$</td>
<td>0010 0000</td>
<td>32</td>
</tr>
</tbody>
</table>

Sizes

<table>
<thead>
<tr>
<th>C type</th>
<th>x86-64 suffix</th>
<th>Size (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>char</td>
<td>b</td>
<td>1</td>
</tr>
<tr>
<td>short</td>
<td>w</td>
<td>2</td>
</tr>
<tr>
<td>int</td>
<td>l</td>
<td>4</td>
</tr>
<tr>
<td>long</td>
<td>q</td>
<td>8</td>
</tr>
</tbody>
</table>

Assembly Instructions

- `mov a,b`: Copy from `a` to `b`
- `movs a,b`: Copy from `a` to `b` with sign extension.
- `movz a,b`: Copy from `a` to `b` with zero extension.
- `lea a,b`: Compute address and store in `b`. *Note: the scaling parameter of memory operands can only be 1, 2, 4, or 8.*
- `push src`: Push `src` onto the stack and decrement stack pointer.
- `pop dst`: Pop from the stack into `dst` and increment stack pointer.
- `call <func>`: Push return address onto stack and jump to a procedure.
- `ret`: Pop return address and jump there.
- `add a,b`: Add `a` to `b` and store in `b` (and sets flags)
- `imul a,b`: Multiply `a` by `b` and store in `b` (and sets flags)
- `and a,b`: Bitwise AND of `a` and `b`, store in `b` (and sets flags)
- `sar a,b`: Shift value of `b right (arithmetic)` by `a` bits, store in `b` (and sets flags)
- `shr a,b`: Shift value of `b right (logical)` by `a` bits, store in `b` (and sets flags)
- `shl a,b`: Shift value of `b left` by `a` bits, store in `b` (and sets flags)
- `cmp a,b`: Compare `b` with `a` (compute `b−a` and set condition codes based on result).
- `test a,b`: Bitwise AND `a` and `b` and set condition codes based on result.
- `jmp <label>`: Jump to address
- `j_ <label>`: Conditional jump based on condition codes (*more on next page*)
- `set_ a`: Set byte based on condition codes.
Conditionals  

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Description</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>je</td>
<td>“Equal”</td>
<td>a == b</td>
</tr>
<tr>
<td>jne</td>
<td>“Not equal”</td>
<td>a != b</td>
</tr>
<tr>
<td>js</td>
<td>“Sign” (negative)</td>
<td>a &amp; b &lt; 0</td>
</tr>
<tr>
<td>jns</td>
<td>(non-negative)</td>
<td>a &amp; b &gt;= 0</td>
</tr>
<tr>
<td>jg</td>
<td>“Greater”</td>
<td>a &gt; b</td>
</tr>
<tr>
<td>jge</td>
<td>“Greater or equal”</td>
<td>a &amp; b &gt;= 0</td>
</tr>
<tr>
<td>jl</td>
<td>“Less”</td>
<td>a &lt; b</td>
</tr>
<tr>
<td>jle</td>
<td>“Less or equal”</td>
<td>a &amp; b &lt;= 0</td>
</tr>
<tr>
<td>ja</td>
<td>“Above” (unsigned &gt;)</td>
<td>a &gt; b</td>
</tr>
<tr>
<td>jb</td>
<td>“Below” (unsigned &lt;)</td>
<td>a &lt; b</td>
</tr>
</tbody>
</table>

C Functions

`void* malloc(size_t size):`  
Allocate size bytes from the heap.

`void* calloc(size_t n, size_t size):`  
Allocate n * size bytes and initialize to 0.

`void free(void* ptr):`  
Free the memory space pointed to by ptr.

`size_t sizeof(type):`  
Returns the size of a given type (in bytes).

`char* gets(char* s):`  
Reads a line from stdin into the buffer.

`pid_t fork():`  
Create a new process by duplicating calling process.

`pid_t wait(int * status):`  
Blocks calling process until any child process exits.

`int execv(char* path, char * argv[]):`  
Replace current process image with new image.

### Registers

<table>
<thead>
<tr>
<th>Name</th>
<th>Convention</th>
<th>Name of &quot;virtual&quot; register</th>
</tr>
</thead>
<tbody>
<tr>
<td>%rax</td>
<td>Return value – Caller saved</td>
<td>%eax %ax %al</td>
</tr>
<tr>
<td>%rbx</td>
<td>Callee saved</td>
<td>%ebx %bx %bl</td>
</tr>
<tr>
<td>%rcx</td>
<td>Argument #4 – Caller saved</td>
<td>%ecx %cx %cl</td>
</tr>
<tr>
<td>%rdx</td>
<td>Argument #3 – Caller saved</td>
<td>%edx %dx %dl</td>
</tr>
<tr>
<td>%rsi</td>
<td>Argument #2 – Caller saved</td>
<td>%esi %si %sil</td>
</tr>
<tr>
<td>%rdi</td>
<td>Argument #1 – Caller saved</td>
<td>%edi %di %dl</td>
</tr>
<tr>
<td>%rsp</td>
<td>Stack pointer</td>
<td>%esp %sp %spl</td>
</tr>
<tr>
<td>%rbp</td>
<td>Callee saved</td>
<td>%ebp %bp %bpl</td>
</tr>
<tr>
<td>%r8</td>
<td>Argument #5 – Caller saved</td>
<td>%r8d %r8w %r8b</td>
</tr>
<tr>
<td>%r9</td>
<td>Argument #6 – Caller saved</td>
<td>%r9d %r9w %r9b</td>
</tr>
<tr>
<td>%r10</td>
<td>Caller saved</td>
<td>%r10d %r10w %r10b</td>
</tr>
<tr>
<td>%r11</td>
<td>Caller saved</td>
<td>%r11d %r11w %r11b</td>
</tr>
<tr>
<td>%r12</td>
<td>Callee saved</td>
<td>%r12d %r12w %r12b</td>
</tr>
<tr>
<td>%r13</td>
<td>Callee saved</td>
<td>%r13d %r13w %r13b</td>
</tr>
<tr>
<td>%r14</td>
<td>Callee saved</td>
<td>%r14d %r14w %r14b</td>
</tr>
<tr>
<td>%r15</td>
<td>Callee saved</td>
<td>%r15d %r15w %r15b</td>
</tr>
</tbody>
</table>