Section Notes 1 (Solutions)

Setup

Resources

printf, C variable types/declarations, etc: http://cs61.seas.harvard.edu/wiki/2015/Resources

Infrastructure Issues?

http://cs61.seas.harvard.edu/wiki/2015/Infrastructure

Git

Make sure you can:

1. Pull the handout code from code.seas
2. Make a commit (let's edit the README!)
3. Push to your repo on code.seas

For section, clone the cs61-sections repo so you can play with the code locally:

```
git clone git@code.seas.harvard.edu:cs61/cs61-sections.git
```

See here for more information: http://cs61.seas.harvard.edu/wiki/2015/Git

---

Assignment 1: Debugging Malloc

Motivation

Let's write some evil programs! List some possible memory bugs:

Memory Bugs:

a. Memory leaks
b. Invalid frees -- freeing memory on the stack
c. Use of uninitialized pointer or already freed pointer
d. Out-of-bounds write
e. Integer overflow
f. Double frees

Some Small Examples

Go to cs61-sections/2015/s01. For each of the membug*.c files, name the memory bugs present in each file. Before running, predict the effect of the memory bug -- will the program crash? Fail silently? Run "make", then "./membug1", "./membug2", etc. to observe the effects of each bug.

Point out the warnings that are generated by compiling these files as well

a. membug1.c: Invalid free: pointer not allocated
b. membug2.c: Memory leak
c. membug3.c: Invalid free: freeing memory on the stack.
d. membug4.c: Invalid free: freeing memory on the stack! ptr itself is allocated on the stack.
e. membug5.c: Out-of-bounds write
f. membug6.c: Use of uninitialized pointer and already freed pointer.
g. membug7.c: Double free and memory leak

A bigger example: a buggy heap.

A heap is a tree-based data structure that satisfies something called the *heap property*. This property states that all pairs of parent/child nodes are ordered in the same way. For example, in a max heap a parent node is always greater than all its children. Therefore, you can see that the maximum element in a max heap is the root node of the tree. Heaps have O(log(n)) insertion and deletion time for their elements.

The code heap.c will read in integers as command line arguments and build a valid max heap from these numbers. Run "make" and then "./heap 3 4 5". It prints out the resulting heap; the max element should be first, and then the following rows print lesser items. We get this:

```
$ ./heap 3 4 5
5
4   3
```

This looks good—the maximum item is on the first line, and the other items are there too. But the next line doesn't look good:

```
*** Error in `./heap': munmap_chunk(): invalid pointer: 0xbffbde0c ***
Aborted (core dumped)
```

Core dump!! Time to boot up GDB to figure out what happened.

Bug 1

- Open up GDB with the command "gdb heap". You can also pass in the -tui flag ("gdb -tui heap") in order to have a graphical interface to see the code alongside the debugger.
- Run the program by typing "run 3 4 5". GDB conveniently stops where the Segmentation Fault occurred.
- We can print the contents of all the variables in scope and try to figure out what went wrong. Do you see what happened? What is the bug in the code that caused this to occur?

SOLUTION: passing h instead of &h to heap_free; h is on the stack, not the heap. Use `bt` to get to user code, since we assume almost always that there is no bug in library code.

There are lots more bugs in this code - see how many more you can find on your own! We will pick up and continue with this example next time.

Writing our own debugger

Open m61.c in the pset1 directory. Discuss the functionality of m61_malloc, m61_free, m61_calloc, m61_realloc. Open m61.h in the pset1 directory. Discuss the interaction between m61_malloc and malloc function calls (ex: calling malloc without M61_DISABLE means malloc calls m61_malloc). What can we add to m61_* implementations to avoid and keep track of these memory bugs?

// have students look at the tests to see what you should be doing

Pointer Review

note to TFs: skip to Assignment 1 Hints if running low on time

Given the following definitions what does each function do, and what do they return?
(Assume the arguments are cast to the right types)

1	`int sum(int a, int b){`
	`  return a + b;`
	`}`
		`sum(8, 6) = 14`

2	`char* sum(char* a, int b) {`
	`  return &a[b];`
	`}`
		`sum((char*) 8, 6) = 14`

3	`int* sum(int* a, int b) {`
	`  return &a[b];`
	`}`
		`sum((int*) 8, 6) = 32`

Memory Layout

Top: memory location – This is the address where the bytes reside.
Middle: data – the bytes being pointed to.
Bottom: logical array index.

0x006	0x007	0x008	0x009	0x00A	0x00B	0x00C	0x00D	0x00E	0x00F
0x8B	0x14	0x24	0x08	0x05	0x44	0x24	0x04	0xC3	0x00
-2	-1	0	1	2	3	4	5	6	7

1. Which col/row is 'a' as it is used in '&a[b]' (in example #2)? `0x008`
2. What would '*a' equal? `0x24 (or 36 in decimal)`
3. How about a[6]? `0xC3 (or 195)`
4. Can you think of another way to write the char* version? Hint: how are arrays and ptrs related?

```
char* sum(char* a, int b) {
```

```
        return a + b;
}
```

5. What is the key to these char* versions working? Another way of putting that question is what's wrong with the int* version?

   Explanation: Conceptually we can use 'address of' (&) to move from the value in the second row (a[b]) to the address in the top row (&a[b]). That is a[6] is the data C3, the address of C3 is E which is our answer: 8+6=E (in Hex).

> More weird code syntax: what is the difference between a[5] and 5[a]?
>
> Although it looks a little funky, a[5] is equivalent to *(a+5) in C. Similarly, 5[a] is equivalent to *(5+a)! (That's not so in Java, Javascript, or C++.)

Pointer Equivalence

You can think of comparisons in three ways: data/value, point to the same memory, are the same memory.

If we have:

```
int c, d;
int *e, *f;
c = 10;
d = 10;
```

clearly c == d because we are comparing values, specifically the bytes 0x0A with 0x0A.

If we then say

```
e = &c;
f = &d;
```

then *e == *f, since 10 == 10, but e != f. Why??

Pointer comparison boils down to comparing the underlying addresses. Since the pointers point to different objects—e points to c, and f points to d—and C guarantees that different objects have distinct, non-overlapping addresses (with the exception of unions), we can see that e != f.

Equivalence Exercises

`int a = 5;`	`1. (a == b) ? True : False;`

```
int b = 5;
int* x = &a; // x==0xf4dc
int* y = &b; // y==0xf4d8
int* z = x;
int** p = &z; // p==0xf4d0
int array[10] = {5};
int yarra[6] = {1, 2, 3, 4, 5, 6};
int* w = array + 4;
int* group[3];
group[0] = array;
group[1] = yarra;
group[2] = y;
```

```
2. (x == y) ? True : False;
3. (y == &a) ? True : False;
4. (*z == a) ? True : False;
5. (*group[0] == a) ? True : False;
6. ((*group)[4] == (*(group+1))[4]) ? True : False;
7. ((*group)[0] == (*(group+1))[4]) ? True : False;
8. (*w == 4) ? True : False;
9. (*w == 0) ? True : False;
10. (w == 0) ? True : False;
11. (((*(group+1))[3]) == 1) ? True : False;
12. ((**(group+2)) == 5) ? True : False;
```

Symbol	Type	Value
z	int*	0xf4dc
(&z)-2	int**	0xf4c8
yarra[3]	int	4
&b	int*	0xf4d8
group[2][0]	int	5
array[6]	int	0
*p	int*	0xf4dc
**(x-3)	int	5
**(&(group[1])-1)	int	5
*w	int	0
group+2	int**	Unknown
*(group+3)	Undefined behavior	Undefined behavior

Extra Hints for Assignment 1

1. Use structs to avoid complicated pointer arithmetic.

Suppose ptr is a pointer to the start of where we are placing our metadata, and we want to place the following types in the metadata: int, int, char*. Let's look at this code:

```
// declare metadata values
int a = 5;
int b = 6;
char* c = // some pointer

void* ptr = // pointer to beginning of metadata;
*((int*) ptr) = a;
*(((int*) ptr) + 1) = b;
*(((char**) ptr) + 2) = c; // char * is 4 bytes on 32-bit machine
```

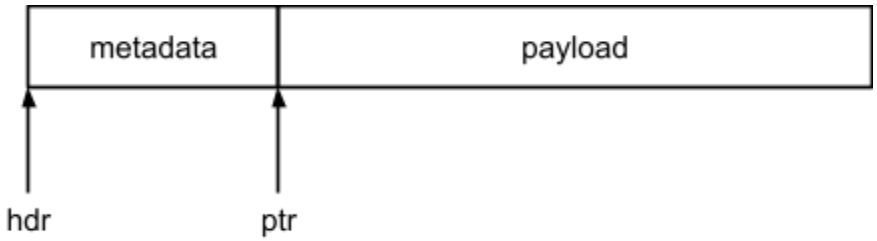
This is horrendous code. There is a much more elegant and robust solution using structs:

```
struct header {
  int a;
  int b;
  char* c;
};

void* ptr = // pointer to beginning of metadata;
struct header* hdr = (struct header*) ptr;
hdr->a = a;
hdr->b = b;
hdr->c = c;
```

2. Avoid pointer arithmetic on (void*) types!

Suppose ptr is a pointer to the beginning of the payload, and we want to move the pointer back to get to the beginning of the metadata:




```
void* ptr = // pointer to beginning of payload
struct header* hdr = (void*) (ptr - sizeof(struct header));
// avoid this!!! void* arithmetic is not well defined;
// this works on GCC and Clang in some modes but not in others
```

Instead:

```
void* ptr = // pointer to beginning of payload
struct header* hdr = ((struct header*) ptr) - 1; // much more robust
```

Identify the Undefined Behavior!

(there are also some behaviors that perhaps aren't undefined, but will cause compiler warnings/failures)

```
int lotsOfUndefinedBehavior {
    int i = 0;
    int j;
    char* message = "Hello World";
    message[0] = 'J';
      // Cannot change a string literal (could be in read-only memory along
      // with the rest of the code, for instance)
    printf("%s", message);

    i++;
    printf("New value is: %d", i);

    j++; // Use of uninitialized variable
    printf("New value is: %d", j);

    char* p = (char*) malloc(sizeof(char)/sizeof(int));
    p[0] = 'a'; // malloc 0 bytes means this is a null ptr dereference!
    printf("Our string is: %s", p);

    int* nats = (int*) malloc(sizeof(int)*42);
    nats[0] = 1;
    for( int k; k < 43; k++ ) { // k was never initialized!
        nats[k] = 2*nats[k-1];
            // when k >= 31, we get integer overflow -- note that nats[k] =
      2^k, and INT_MAX is 2^31 - 1.
            // when k = 42, accessing an out-of-bounds array element
    }
      // we don't return anything in a non-void function!
}
```