Structures and Unions in C

Alan L. Cox
alc@rice.edu
Administrivia

Assignment 1 is due tonight

Textbook
- Lectures begin covering material that is also covered by the textbook on 1/29
- Assignment 3 (assigned 1/31) requires use of the textbook
Objectives

Be able to use compound data structures in programs

Be able to pass compound data structures as function arguments, either by value or by reference

Be able to do simple bit-vector manipulations
Structures

Compound data:

A date is
  - an int month and
  - an int day and
  - an int year

Unlike Java, C doesn’t automatically define functions for initializing and printing ...
Structure Representation & Size

\[
\text{sizeof(struct ...)} = \ \text{sum of sizeof(field)} + \text{alignment padding}
\]

Processor- and compiler-specific

```c
struct CharCharInt {
    char c1;
    char c2;
    int i;
} foo;

foo.c1 = 'a';
foo.c2 = 'b';
foo.i = 0xDEADBEEF;
```

<table>
<thead>
<tr>
<th>c1</th>
<th>c2</th>
<th>padding</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>62</td>
<td></td>
<td>EF</td>
</tr>
</tbody>
</table>

x86 uses “little-endian” representation
Typedef

Mechanism for creating new type names

- New names are an alias for some other type
- *May* improve clarity and/or portability of the program

```
typedef long int64_t;
typedef struct ADate {
    int month;
    int day;
    int year;
} Date;
int64_t i = 100000000000;
Date d = { 1, 18, 2018 };
```
Constants

Allow consistent use of the same constant throughout the program

- Improves clarity of the program
- Reduces likelihood of simple errors
- Easier to update constants in the program

Preprocessor directive

Constant names are capitalized by convention

Define once, use throughout the program

```c
int array[10];
for (i=0; i<10; i++) {
  ...
}
```
Arrays of Structures

```c
Date birthdays[NFRIENDS];

bool check_birthday(Date today) {
    int i;
    for (i = 0; i < NFRIENDS; i++) {
        if ((today.month == birthdays[i].month) &&
            (today.day == birthdays[i].day))
            return (true);
    }
    return (false);
}
```
Pointers to Structures

Date
create_date1(int month, int day, int year)
{
    Date d;
    d.month = month;
    d.day = day;
    d.year = year;
    return (d);
}

void
create_date2(Date *d, int month, int day, int year)
{
    d->month = month;
    d->day = day;
    d->year = year;
}

Date today;
today = create_date1(1, 18, 2018);
cREATE_DATE2(&today, 1, 18, 2018);

Copies date
Pass-by-reference

Pointers to Structures (cont.)

```c
void create_date2(Date *d, int month, int day, int year)
{
    d->month = month;
    d->day   = day;
    d->year  = year;
}

void fun_with_dates(void)
{
    Date today;
    create_date2(&today, 1, 18, 2018);
}
```

```c
<table>
<thead>
<tr>
<th>Address</th>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x30A0</td>
<td>month</td>
<td>1</td>
</tr>
<tr>
<td>0x30A4</td>
<td>day</td>
<td>18</td>
</tr>
<tr>
<td>0x30A8</td>
<td>year</td>
<td>2018</td>
</tr>
<tr>
<td>0x3098</td>
<td>d</td>
<td>0x1000</td>
</tr>
</tbody>
</table>

```
Pointers to Structures (cont.)

```c
Date *
create_date3(int month,
    int day,
    int year)
{
    Date *d;
    d->month = month;
    d->day   = day;
    d->year  = year;
    return (d);
}
```

What is d pointing to?!?!
(more on this later)
Abstraction in C

From the #include file widget.h:

```
struct widget;

struct widget *widget_create(void);
int            widget_op(struct widget *widget, int operand);
void           widget_destroy(struct widget *widget);
```

From the file widget.c:

```
#include "widget.h"

struct widget {
    int x;
    ...
};
```
Collections of Bools (Bit Vectors)

Byte, word, ... can represent many Booleans
   One per bit, e.g.,  00100101 = false, false, true, ..., true

Bit-wise operations:

Bit-wise AND:  00100101 & 10111100 == 00100100
Bit-wise OR:   00100101 | 10111100 == 10111101
Bit-wise NOT:  ~ 00100101      == 11011010
Bit-wise XOR:  00100101 ^ 10111100 == 10011001
Operations on Bit Vectors

A mask indicates which bit positions we are interested in.

Always use C’s `unsigned` types for bit vectors.

Selecting bits:

```c
const unsigned int low_three_bits_mask = 0x7;
unsigned int bit_vec = 0x15;

important_bits = bit_vec & low_three_bits_mask;
```

Result = ?

0...00 0101 == 0...01 0101 & 0...00 0111
Operations on Bit Vectors

```c
const unsigned int low_three_bits_mask = 0x7;
unsigned int bit_vec = 0x15;
```

Setting bits:

```c
bit_vec |= low_three_bits_mask;
```

Result = ?

```
0...01 0111 == 0...01 0101 | 0...00 0111
```
Operations on Bit Vectors

const unsigned int low_three_bits_mask = 0x7;  \(0\ldots00\ 0111\)
unsigned int bit_vec = 0x15;  \(0\ldots01\ 0101\)

Clearing bits:

\[\text{bit_vec} \&= \sim\text{low}\_\text{three}\_\text{bits}\_\text{mask};\]

\[0\ldots01\ 0000 == 0\ldots01\ 0101 \& \sim0\ldots00\ 0111\]

Result = ?
Bit-field Structures

Special syntax packs structure values more tightly

Similar to bit vectors, but arguably easier to read

- Nonetheless, bit vectors are more commonly used.

Padded to be an integral number of words

- Placement is compiler-specific.

```c
struct Flags {
    int f1:3;
    unsigned int f2:1;
    unsigned int f3:2;
} my_flags;

my_flags.f1 = -2;
my_flags.f2 = 1;
my_flags.f3 = 2;
```

```
  f1 f2 f3
  1 1 0 1 1 0 ...
```
Unions

Choices:

An element is

- an int `i` or
- a char `c`

\[
\text{sizeof(union ...)} = \text{maximum of sizeof(field)}
\]

```c
union AnElt {
    int   i;
    char  c;
} elt1, elt2;

elt1.i = 4;
elt2.c = 'a';
elt2.i = 0xDEADBEF;
```

```
c | padding  
---|----------
EF | BE  AD  DE
i |
```
Unions

A union value doesn’t “know” which case it contains

union AnElt {
    int   i;
    char  c;
} elt1, elt2;

elt1.i = 4;
elt2.c = 'a';
elt2.i = 0xDEADBEEF;

if (elt1 currently has a char) ...

? How should your program keep track whether elt1, elt2 hold an int or a char?

? Basic answer: Another variable holds that info
Tagged Unions

Tag every value with its case

I.e., pair the type info together with the union
Implicit in Java, Scheme, ML, ...

```c
enum Union_Tag { IS_INT, IS_CHAR };
struct TaggedUnion {
    enum Union_Tag  tag;
    union {
        int   i;
        char  c;
    } data;
};
```

- Enum must be external to struct, so constants are globally visible.
- Struct field must be named.
Next Time

Memory Allocation