Memory, Data, & Addressing II
CSE 351 Spring 2020

Instructor:
Ruth Anderson

Teaching Assistants:
Alex Olshanskyy
Rehaan Bhimani
Callum Walker
Chin Yeoh
Diya Joy
Eric Fan
Edan Sneh
Jonathan Chen
Jeffery Tian
Millicent Li
Melissa Birchfield
Porter Jones
Joseph Schafer
Connie Wang
Eddy (Tianyi) Zhou

http://xkcd.com/138/
Administrivia

- Assignments – Nothing **Due** this week!
- Lecture Video to watch before Monday (4/06)
  - Not a full lecture, but part of one to help us retain time in lectures for small group interaction.
  - Posted later today, should be watched before lecture on Monday
- Pre-Course Survey, hw0, hw1, hw2 due Mon (4/06) – 11:59pm
- Lab 0 due Tuesday (4/07) – 11:59pm
- hw3 due Wednesday (4/08) – **11am**
  - Autograded, unlimited tries, no late submissions
- Lab 1a released today, due next Monday (4/13)
  - Pointers in C
  - Reminder: last submission graded, *individual* work
- Feedback for today: [https://catalyst.uw.edu/webq/survey/rea2000/387932](https://catalyst.uw.edu/webq/survey/rea2000/387932)
Late Days

- You are given **7 late days** for the whole quarter
  - Late days can only apply to Labs & Unit Summaries
  - No benefit to having leftover late days
- Count lateness in **days** (even if just by a second)
  - **Special**: weekends count as **one day**
  - No submissions accepted more than two days late
- Late penalty is 20% deduction of your score per day
  - Only late work is eligible for penalties
  - Penalties applied at end of quarter to *maximize* your grade
- Use at own risk – don’t want to fall too far behind
  - Intended to allow for unexpected circumstances
Memory, Data, and Addressing

- Representing information as bits and bytes
  - Binary, hexadecimal, fixed-widths
- Organizing and addressing data in memory
  - Memory is a byte-addressable array
  - Machine “word” size = address size = register size
  - Endianness – ordering bytes in memory
- Manipulating data in memory using C
  - Assignment
  - Pointers, pointer arithmetic, and arrays
- Boolean algebra and bit-level manipulations
Addresses and Pointers in C

- `&` = “address of” operator
- `*` = “value at address” or “dereference” operator

```c
int* ptr;  // Declares a variable, `ptr`, that is a pointer to (i.e. holds the address of) an `int` in memory
int x = 5;  // Declares two variables, `x` and `y`, that hold `ints`, and initializes them to 5 and 2, respectively
int y = 2;
ptr = &x;  // Sets `ptr` to the address of `x` ("`ptr points to x""

y = 1 + *ptr;  // "Dereference `ptr""
```

Sets `y` to “1 plus the value stored at the address held by `ptr`.” Because `ptr` points to `x`, this is equivalent to `y=1+x`;

What is `*(&y)`?

* is also used with variable declarations
Assignment in C

- A variable is represented by a location
- Declaration ≠ initialization (initially holds “garbage”)
- `int x, y;`
  - `x` is at address 0x04, `y` is at 0x18

<table>
<thead>
<tr>
<th>Address</th>
<th>0x00</th>
<th>0x01</th>
<th>0x02</th>
<th>0x03</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x00</td>
<td>A7</td>
<td>00</td>
<td>32</td>
<td>00</td>
</tr>
<tr>
<td>0x04</td>
<td>00</td>
<td>01</td>
<td>29</td>
<td>F3</td>
</tr>
<tr>
<td>0x08</td>
<td>EE</td>
<td>EE</td>
<td>EE</td>
<td>EE</td>
</tr>
<tr>
<td>0x0C</td>
<td>FA</td>
<td>CE</td>
<td>CA</td>
<td>FE</td>
</tr>
<tr>
<td>0x10</td>
<td>26</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>0x14</td>
<td>00</td>
<td>00</td>
<td>10</td>
<td>00</td>
</tr>
<tr>
<td>0x18</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>0x1C</td>
<td>FF</td>
<td>00</td>
<td>F4</td>
<td>96</td>
</tr>
<tr>
<td>0x20</td>
<td>DE</td>
<td>AD</td>
<td>BE</td>
<td>EF</td>
</tr>
<tr>
<td>0x24</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
</tbody>
</table>
Assignment in C

- A variable is represented by a location
- Declaration ≠ initialization (initially holds “garbage”)
- `int x, y;`
  - `x` is at address 0x04, `y` is at 0x18

### 32-bit example (pointers are 32-bits wide)

<table>
<thead>
<tr>
<th>Address</th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x00</td>
<td></td>
</tr>
<tr>
<td>0x04</td>
<td>00 01 29 F3</td>
</tr>
<tr>
<td>0x08</td>
<td></td>
</tr>
<tr>
<td>0x0C</td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td></td>
</tr>
<tr>
<td>0x14</td>
<td></td>
</tr>
<tr>
<td>0x18</td>
<td>01 00 00 00</td>
</tr>
<tr>
<td>0x1C</td>
<td></td>
</tr>
<tr>
<td>0x20</td>
<td></td>
</tr>
<tr>
<td>0x24</td>
<td></td>
</tr>
</tbody>
</table>

little-endian
Assignment in C

- left-hand side = right-hand side;
  - LHS must evaluate to a location
  - RHS must evaluate to a value (could be an address)
  - Store RHS value at LHS location
- `int x, y;`
- `x = 0;`

32-bit example (pointers are 32-bits wide)

& = “address of”
* = “dereference”
Assignment in C

- left-hand side = right-hand side;
  - LHS must evaluate to a location
  - RHS must evaluate to a value (could be an address)
  - Store RHS value at LHS location

- `int x, y;`
- `x = 0;`
- `y = 0x3CD02700;`

32-bit example (pointers are 32-bits wide)

& = "address of"
* = "dereference"

little endian!
Assignment in C

- left-hand side = right-hand side;
  - LHS must evaluate to a location
  - RHS must evaluate to a value (could be an address)
  - Store RHS value at LHS location

- `int x, y;`
- `x = 0;`
- `y = 0x3CD02700;`
- `x = y + 3;`
  - Get value at `y`, add 3, store in `x`

32-bit example (pointers are 32-bits wide)

& = “address of”
* = “dereference”
Assignment in C

- left-hand side = right-hand side;
  - LHS must evaluate to a *location*
  - RHS must evaluate to a *value* (could be an address)
  - Store RHS value at LHS location

- `int x, y;`
- `x = 0;`
- `y = 0x3CD02700;`
- `x = y + 3;`
  - Get value at `y`, add 3, store in `x`
- `int* z;`
  - `z` is at address 0x20

---

32-bit example
(pointers are 32-bits wide)

& = “address of”
* = “dereference”
Assignment in C

- left-hand side = right-hand side;
  - LHS must evaluate to a *location*
  - RHS must evaluate to a *value* (could be an address)
  - Store RHS value at LHS location

- **int** `x, y;`
- `x = 0;`
- `y = 0x3CD02700;`
- `x = y + 3;`
  - Get value at `y`, add 3, store in `x`
- **int** `*z = &y + 3;`
  - Get address of `y`, “add 3”, store in `z`

<table>
<thead>
<tr>
<th>Address</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x00</td>
<td>0x00</td>
</tr>
<tr>
<td>0x04</td>
<td>0x04</td>
</tr>
<tr>
<td>0x08</td>
<td>0x08</td>
</tr>
<tr>
<td>0x0C</td>
<td>0x0C</td>
</tr>
<tr>
<td>0x10</td>
<td>0x10</td>
</tr>
<tr>
<td>0x14</td>
<td>0x14</td>
</tr>
<tr>
<td>0x18</td>
<td>0x18</td>
</tr>
<tr>
<td>0x1C</td>
<td>0x1C</td>
</tr>
<tr>
<td>0x20</td>
<td>0x20</td>
</tr>
<tr>
<td>0x24</td>
<td>0x24</td>
</tr>
</tbody>
</table>

32-bit example (pointers are 32-bits wide)

`&` = “address of”
`*` = “dereference”

**Pointer arithmetic**
Pointer Arithmetic

- Pointer arithmetic is scaled by the size of target type
  - In this example, \(\text{sizeof}(\text{int}) = 4\)

- \(\text{int}^* z = \&y + 3;\)
  - Get address of \(y\), add \(3 \times \text{sizeof}(\text{int})\), store in \(z\)
  - \(\&y = 0x18 = 1 \times 16^1 + 8 \times 16^0 = 24\)
  - \(24 + 3 \times (4) = 36 = 2 \times 16^1 + 4 \times 16^0 = 0x24\)

- Pointer arithmetic can be dangerous!
  - Can easily lead to bad memory accesses
  - Be careful with data types and \textit{casting}
Assignment in C

- `int x, y;`
- `x = 0;`
- `y = 0x3CD02700;`
- `x = y + 3;
  - Get value at `y`, add 3, store in `x`
- `int* z = &y + 3;`
  - Get address of `y`, add 12, store in `z`
- `*z = y;`
  - What does this do?
Assignment in C

- `int x, y;`
- `x = 0;`
- `y = 0x3CD02700;`
- `x = y + 3;`
  - Get value at `y`, add 3, store in `x`
- `int* z = &y + 3;`
  - Get address of `y`, add 12, store in `z`
- `*z = y;`
  - Get value of `y`, put in address stored in `z`
Arrays in C

Declaration: `int a[6];`

- **Element type**
- **Name**
- **Number of elements**

Arrays are adjacent locations in memory storing the same type of data object.

`a` (array name) returns the array’s address.

64-bit example (pointers are 64-bits wide)
Arrays in C

Declaration: `int a[6];`

Indexing:  
- `a[0] = 0x015f;`  
- `a[5] = a[0];`

Arrays are adjacent locations in memory storing the same type of data object:
- `a` (array name) returns the array’s address
- `&a[i]` is the address of `a[0]` plus `i` times the element size in bytes

![Array Memory Diagram](image-url)
Arrays in C

Declaration: `int a[6];`

Indexing:  
- `a[0] = 0x015f;`
- `a[5] = a[0];`

No bounds  
- `a[6] = 0xBAD;`

cHECKING:  
- `a[-1] = 0xBAD;`

Arrays are adjacent locations in memory storing the same type of data object.

- `a` (array name) returns the array’s address.
- `&a[i]` is the address of `a[0]` plus `i` times the element size in bytes.

---

```
0x00 0x08 0x10 0x18 0x20 0x28 0x30 0x38 0x40 0x48
0x00 0x80 0x10 0x18 0x20 0x28 0x30 0x38 0x40 0x48
0x00 0x01 0x00 0x00 0x01 0x00 0x00 0x01 0x00 0x00
```

0x00 0x08 0x10 0x18 0x20 0x28 0x30 0x38 0x40 0x48
AD 0B 00 00 5F 01 00 00 5F 01 00 00 AD 0B 00 00
Arrays in C

Declaration: `int a[6];`

Indexing:
- `a[0] = 0x015f;`
- `a[5] = a[0];`

No bounds:
- `a[6] = 0xBAD;`

Checking:
- `a[-1] = 0xBAD;`

Pointers:
- `int* p;`

Equivalent:
- `p = a;`
- `p = &a[0];`
- `*p = 0xA;`

Arrays are adjacent locations in memory storing the same type of data object.

**a (array name)** returns the array’s address.

**&a[i]** is the address of **a[0]** plus i times the element size in bytes.
Arrays in C

Declaration: `int a[6];`

Indexing: 
- `a[0] = 0x015f;`
- `a[5] = a[0];`

No bounds 
- `a[6] = 0xBAD;`

checking: 
- `a[-1] = 0xBAD;`

Pointers: 
- `int* p;`
- equivalent: 
  - `p = a;`
  - `p = &a[0];`
  - `*p = 0xA;`

Array indexing = address arithmetic (both scaled by the size of the type)
- equivalent: 
  - `p[1] = 0xB;`
  - `*(p+1) = 0xB;`
  - `p = p + 2;`
Arrays in C

Declaration: `int a[6];`

Indexing: `a[0] = 0x015f;`  
`a[5] = a[0];`

No bounds `a[6] = 0xBBAD;`  
checking: `a[-1] = 0xBBAD;`

Pointers: `int* p;`  

equivalent  

```
p = a;  
p = &a[0];  
*p = 0xA;  
```

array indexing = address arithmetic  
(both scaled by the size of the type)

```
p[1] = 0xBB;  
*(p+1) = 0xBB;  
p = p + 2;  
```

`*p = a[1] + 1;`
Question: The variable values after Line 3 executes are shown on the right. What are they after Line 4 & 5?

- Vote at [http://pollev.com/rea](http://pollev.com/rea)

```c
void main() {
    int a[] = {5, 10};
    int* p = a;
    p = p + 1;
    *p = *p + 1;
}
```

| (A) | 101 10 5 10 then | 101 11 5 11 |
| (B) | 104 10 5 10 then | 104 11 5 11 |
| (C) | 100 6 6 10 then | 101 6 6 10 |
| (D) | 100 6 6 10 then | 104 6 6 10 |
Representing strings

- C-style string stored as an array of bytes (char*)
  - Elements are one-byte ASCII codes for each character
  - No “String” keyword, unlike Java

<table>
<thead>
<tr>
<th>ASCII</th>
<th>Code</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>48</td>
<td>space</td>
</tr>
<tr>
<td>33</td>
<td>49</td>
<td>!</td>
</tr>
<tr>
<td>34</td>
<td>50</td>
<td>&quot;</td>
</tr>
<tr>
<td>35</td>
<td>51</td>
<td>#</td>
</tr>
<tr>
<td>36</td>
<td>52</td>
<td>$</td>
</tr>
<tr>
<td>37</td>
<td>53</td>
<td>%</td>
</tr>
<tr>
<td>38</td>
<td>54</td>
<td>&amp;</td>
</tr>
<tr>
<td>39</td>
<td>55</td>
<td>'</td>
</tr>
<tr>
<td>40</td>
<td>56</td>
<td>(</td>
</tr>
<tr>
<td>41</td>
<td>57</td>
<td>)</td>
</tr>
<tr>
<td>42</td>
<td>58</td>
<td>*</td>
</tr>
<tr>
<td>43</td>
<td>59</td>
<td>+</td>
</tr>
<tr>
<td>44</td>
<td>60</td>
<td>,</td>
</tr>
<tr>
<td>45</td>
<td>61</td>
<td>-</td>
</tr>
<tr>
<td>46</td>
<td>62</td>
<td>.</td>
</tr>
<tr>
<td>47</td>
<td>63</td>
<td>/</td>
</tr>
<tr>
<td>64</td>
<td>80</td>
<td>@</td>
</tr>
<tr>
<td>65</td>
<td>81</td>
<td>A</td>
</tr>
<tr>
<td>66</td>
<td>82</td>
<td>B</td>
</tr>
<tr>
<td>67</td>
<td>83</td>
<td>C</td>
</tr>
<tr>
<td>68</td>
<td>84</td>
<td>D</td>
</tr>
<tr>
<td>69</td>
<td>85</td>
<td>E</td>
</tr>
<tr>
<td>70</td>
<td>86</td>
<td>F</td>
</tr>
<tr>
<td>71</td>
<td>87</td>
<td>G</td>
</tr>
<tr>
<td>72</td>
<td>88</td>
<td>H</td>
</tr>
<tr>
<td>73</td>
<td>89</td>
<td>I</td>
</tr>
<tr>
<td>74</td>
<td>90</td>
<td>J</td>
</tr>
<tr>
<td>75</td>
<td>91</td>
<td>K</td>
</tr>
<tr>
<td>76</td>
<td>92</td>
<td>L</td>
</tr>
<tr>
<td>77</td>
<td>93</td>
<td>M</td>
</tr>
<tr>
<td>78</td>
<td>94</td>
<td>N</td>
</tr>
<tr>
<td>79</td>
<td>95</td>
<td>O</td>
</tr>
<tr>
<td>96</td>
<td>97</td>
<td>`</td>
</tr>
<tr>
<td>97</td>
<td>98</td>
<td>a</td>
</tr>
<tr>
<td>98</td>
<td>99</td>
<td>b</td>
</tr>
<tr>
<td>99</td>
<td>100</td>
<td>c</td>
</tr>
<tr>
<td>100</td>
<td>101</td>
<td>d</td>
</tr>
<tr>
<td>101</td>
<td>102</td>
<td>e</td>
</tr>
<tr>
<td>102</td>
<td>103</td>
<td>f</td>
</tr>
<tr>
<td>103</td>
<td>104</td>
<td>g</td>
</tr>
<tr>
<td>104</td>
<td>105</td>
<td>h</td>
</tr>
<tr>
<td>105</td>
<td>106</td>
<td>i</td>
</tr>
<tr>
<td>106</td>
<td>107</td>
<td>j</td>
</tr>
<tr>
<td>107</td>
<td>108</td>
<td>k</td>
</tr>
<tr>
<td>108</td>
<td>109</td>
<td>l</td>
</tr>
<tr>
<td>109</td>
<td>110</td>
<td>m</td>
</tr>
<tr>
<td>110</td>
<td>111</td>
<td>n</td>
</tr>
<tr>
<td>111</td>
<td>112</td>
<td>o</td>
</tr>
<tr>
<td>112</td>
<td>113</td>
<td>p</td>
</tr>
<tr>
<td>113</td>
<td>114</td>
<td>q</td>
</tr>
<tr>
<td>114</td>
<td>115</td>
<td>r</td>
</tr>
<tr>
<td>115</td>
<td>116</td>
<td>s</td>
</tr>
<tr>
<td>116</td>
<td>117</td>
<td>t</td>
</tr>
<tr>
<td>117</td>
<td>118</td>
<td>u</td>
</tr>
<tr>
<td>118</td>
<td>119</td>
<td>v</td>
</tr>
<tr>
<td>119</td>
<td>120</td>
<td>w</td>
</tr>
<tr>
<td>120</td>
<td>121</td>
<td>x</td>
</tr>
<tr>
<td>121</td>
<td>122</td>
<td>y</td>
</tr>
<tr>
<td>122</td>
<td>123</td>
<td>z</td>
</tr>
<tr>
<td>123</td>
<td>124</td>
<td>{</td>
</tr>
<tr>
<td>124</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>126</td>
<td>}</td>
</tr>
<tr>
<td>126</td>
<td>127</td>
<td>~</td>
</tr>
</tbody>
</table>

**ASCII**: American Standard Code for Information Interchange
Null-Terminated Strings

- **Example**: "Donald Trump" stored as a 13-byte array

<table>
<thead>
<tr>
<th>Decimal:</th>
<th>68</th>
<th>111</th>
<th>110</th>
<th>97</th>
<th>108</th>
<th>100</th>
<th>32</th>
<th>84</th>
<th>114</th>
<th>117</th>
<th>109</th>
<th>112</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hex:</td>
<td>0x44</td>
<td>0x6F</td>
<td>0x6E</td>
<td>0x61</td>
<td>0x6C</td>
<td>0x64</td>
<td>0x20</td>
<td>0x54</td>
<td>0x72</td>
<td>0x75</td>
<td>0x6D</td>
<td>0x70</td>
<td>0x00</td>
</tr>
<tr>
<td>Text:</td>
<td>Donald Trump</td>
<td>\0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Last character followed by a 0 byte ("\0")
  (a.k.a. "null terminator")
  - Must take into account when allocating space in memory
  - Note that '0' ≠ '\0' (*i.e.* character 0 has non-zero value)

- How do we compute the length of a string?
  - Traverse array until null terminator encountered
Endianness and Strings

char s[6] = "12345";

- Byte ordering (endianness) is not an issue for 1-byte values
  - The whole array does not constitute a single value
  - Individual elements are values; chars are single bytes

- IA32, x86-64 (little-endian)
- SPARC (big-endian)

C (char = 1 byte)
Examining Data Representations

- Code to print byte representation of data
  - Any data type can be treated as a *byte array* by **casting** it to `char`
  - C has **unchecked casts**  !! DANGER !!

```c
void show_bytes(char* start, int len) {
    int i;
    for (i = 0; i < len; i++)
        printf("%p\t0x%.2x\n", start+i, *(start+i));
    printf("\n");
}
```

**printf directives:**
- `%p` Print pointer
- `\t` Tab
- `%x` Print value as hex
- `\n` New line
Examining Data Representations

- Code to print byte representation of data
  - Any data type can be treated as a byte array by casting it to char
  - C has unchecked casts  !! DANGER !!

```c
void show_bytes(char* start, int len) {
    int i;
    for (i = 0; i < len; i++)
        printf("%p\t0x%.2x\n", start+i, *(start+i));
    printf("\n");
}
```

```c
void show_int(int x) {
    show_bytes( (char *) &x, sizeof(int));
}
```
show_bytes Execution Example

```c
int x = 12345; // 0x00003039
printf("int x = %d;\n", x);
show_int(x); // show_bytes((char *) &x, sizeof(int));
```

- **Result (Linux x86-64):**
  - **Note:** The addresses will change on each run (try it!), but fall in same general range

```c
int x = 12345;
0x7fffb7f71dbc 0x39
0x7fffb7f71dbd 0x30
0x7fffb7f71dbe 0x00
0x7fffb7f71dbf 0x00
```
Summary

- Assignment in C results in value being put in memory location
- Pointer is a C representation of a data address
  - & = “address of” operator
  - * = “value at address” or “dereference” operator
- Pointer arithmetic scales by size of target type
  - Convenient when accessing array-like structures in memory
  - Be careful when using – particularly when casting variables
- Arrays are adjacent locations in memory storing the same type of data object
  - Strings are null-terminated arrays of characters (ASCII)