

§8.0-8.3: Data Storage and Number Bases

27 Oct 2008

CMPT14x

Dr. Sean Ho

Trinity Western University

Announcements

- XSLT competition:
 - Wed 5 Nov 1:00-2:30pm
 - Register by Fri 31 Oct for free tutorial
 - Orion Ifland <orion.ifland@twu.ca>
 - www.twu.ca/xsltcompetition
- Math Sciences Club event:
 - Thu 30 Oct, 1:10-2:20pm Neu 41
 - Alma Barranco, “Electronic Learning”
 - Free pizza!

Essay / Paper

- Computing scientist as **Godly Christian Leader**:
 - Not just **knowledge** about tools, but
 - **Wisdom** of how to use tools
 - ◆ To **serve** others and
 - ◆ To give glory to **God**
- Write a short **essay** on a topic of your choosing about **computers** and **society**:
 - ◆ Approx **5 pages** typed double-spaced 12pt 1in margins
 - ◆ Submit half-page **topic** by **Fri 9Nov**
 - ◆ Paper **due** last week of class (**Mon 3Dec**)
 - Electronic submission (email, eCourses)

Sample paper topics

- **Censorship** and free speech
 - Pornography, gambling, hate groups, etc.
- **Violence** in video games (Columbine etc.)
- **Privacy**: online banking, ID theft, etc.
- **Blogs**: effect on politics, social interaction, etc.
- **File sharing**: Napster, Gnutella, etc.
- **Artificial intelligence**: the nature of sentience
- **Online dating** (e.g. eHarmony): pros/cons
- **Equity of access** / rural digital divide
- **TRINITY.** come up with your **own** topic!

Tips for essay writing

- Your essay should be a **position paper**:
 - The topic should have at least two **sides** (e.g. pro/con)
 - You should state (in the introductory paragraph) what your **position** is (**thesis**)
 - You should have at least 2-3 points, each, both **for** and **against** your position
 - ◆ It is not necessary to **rebut** every point that contradicts your position:
 - ◆ Be honest about the **faults**/limitations of your thesis
 - Summary **intro/conclusion** paragraphs

Ch8: Data storage and I/O

- As programmers, you're already expert **users** of various datatypes and file I/O
- Now we peek **under the hood** to see what the compiler and the OS are really doing to implement these
- Every variable we declare takes up space in **memory** (RAM):
 - How much **space** does each variable need?
 - How is our data **stored**?

Binary numbers



- At the lowest level, all computer data are stored using logical **bits**: each bit can be either 0 or 1
 - **High voltage** (1) vs. **low** voltage (0)
 - Most memory chips use a big bank of tiny **capacitors**: has charge (1) vs. no charge (0)
- We use groups of bits to **represent** data (numbers, characters, strings, etc.):
 - e.g., this pattern of eight **bits**: 0 1 0 0 0 0 1 1
 - ◆ Could represent the decimal **number** 35
 - ◆ Or it might represent the **character** “#”
 - ◆ Or something else – depends on how we **interpret**

Number bases

- God gave us 10 fingers; so we often count in **base 10**:
 - “5927” interpreted as a **decimal** number:
 - ◆ 5 units of ($10^3 = 1000$)
 - ◆ 9 units of ($10^2 = 100$)
 - ◆ 2 units of ($10^1 = 10$)
 - ◆ 7 units of ($10^0 = 1$)
- Counting in **binary** is similar:
 - “0110” interpreted as a binary number:
 - ◆ 0 unit of ($2^3 = 8$)
 - ◆ 1 unit of ($2^2 = 4$)
 - ◆ 1 unit of ($2^1 = 2$)
 - ◆ 0 unit of ($2^0 = 1$)



Hexadecimal, octal

- **Hexadecimal** is base **16**: we use 'A'..'F' to represent the “digits” ten, eleven, twelve, etc.
 - “BEEF” as a hexadecimal number:
 - ◆ B (11) units of ($16^3 = 4096$) \Rightarrow 45056
 - ◆ E (14) units of ($16^2 = 256$) \Rightarrow 3584
 - ◆ E (14) units of ($16^1 = 16$) \Rightarrow 224
 - ◆ F (15) units of ($16^0 = 1$) \Rightarrow 15
 - ◆ Total: BEEF (hex) \Rightarrow 48879 (dec)
- There's also **octal**, base 8:
 - only the digits 0..7 are used

Using bases in Python

- Python has special **notation** for expressing integer literals in hexadecimal and octal:

- **Hexadecimal**: prefix “0x”

```
hexNum = 0xBEEF    # 48879
```

- **Octal**: prefix “0”

```
octNum = 0115    #  $1(8^2) + 1(8^1) + 5(8^0) = 77$ 
```

- Convert into strings with hexadecimal/octal notation:

```
hexStr = hex(48879)    # '0xbeef'
```

```
octStr = oct(77)      # '0115'
```

Bits, bytes, nibbles, words

- One hexadecimal digit can be represented by **four bits**: one **nibble**
- Two nibbles (**eight bits**) is called a **byte**
 - One byte can be used to store one CHAR
- A group of bytes can be used to represent one datum: this is called a **word**
 - Pentium CPUs generally use 4-byte words (**32 bits**)
 - Newer CPUs can use 8-byte words (**64 bits**)
 - Word is the smallest **unit of data** the machine can store or retrieve

Accessing memory



- A computer's **main memory** (generally, RAM) stores everything it needs to do its current tasks
- A location within memory is uniquely identified by its **address**
 - Most modern CPUs use 32-bit words to **store** memory addresses
 - This means there is a maximum of 2^{32} unique memory addresses (the **address space**)
 - If each location stores one byte of data, then there is 2^{32} bytes = 4GB of **addressable**

Units of measure

■ SI abbreviations:

- K = kilo = 1,000
- M = mega = 1,000,000
- G = giga = 1,000,000,000

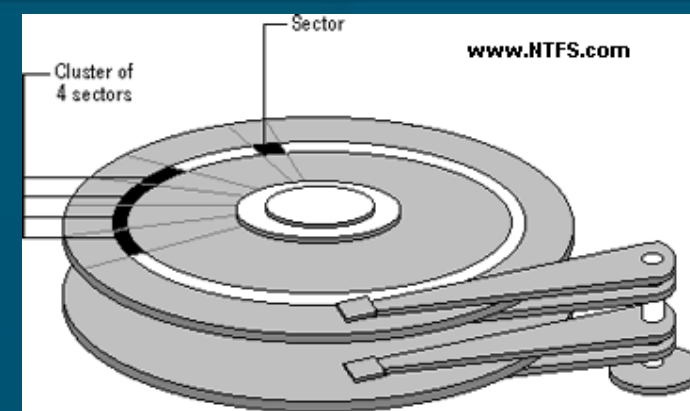
■ When working with binary data:

- KB = kilobyte = 1,024 bytes = 2^{10} bytes
- MB = megabyte = 1,024,576 = 2^{20} bytes
- GB = gigabyte = 1,073,741,824 = 2^{30} bytes
- But hard drive manufacturers use SI abbrevs

Units of measure, cont.

- Kilobytes vs. kilobits:
 - **KB** = kilobyte = 1,024 bytes = 8192 bits
 - **Kb** = kilobit = 1,024 bits
 - RAM chip manufacturers often use kilobits
- Also, in SI abbreviations,
 - **M** = mega = 10^6 : e.g., megawatt = 10^6 watt
 - **m** = milli = 10^{-3} : e.g., milliwatt = 10^{-3} watt
- But not everyone is consistent, so be careful

Storage



- A **page** of memory is generally 256 bytes
- A **sector** is a unit of disk storage, also commonly 256 bytes (but sometimes 512 bytes)
- A **block** of disk storage is usually 512 bytes
- Hard disks are made up of **platters**, accessed by magnetic **heads** on movable arms
- The platters have concentric tracks that (across all heads) make up **cylinders**
- Hard drive geometry is often expressed in **C/H/S: # cylinders / # heads / # sectors per track**

Summary of today (§8.0-8.3)

- Number bases:
 - Binary
 - Hexadecimal (0xBEEF)
 - Octal (0115)
- Units of measure of **memory**:
 - Bits, nibbles, bytes, words, pages
- Units of measure for **hard disks**:
 - C/H/S geometry
- SI **units** vs binary units, KB vs. Kb, etc.