Data Representation and Networking

Instructor: Dmitri A. Gusev

Spring 2007

CSC 120.02: Introduction to Computer Science

Lecture 3, January 30, 2007

Data Representation Topics Covered in Lecture 2 (recap+)

- Bits, Bytes, Words
- Analog (continuous) vs. Digital (discrete); extra term: *Pulse-code modulation (PCM)*
- Representing Images (*pixels*, *RGB*, *indexed color / palette*); *Color depth, HiColor* (16-bit), *TrueColor* (24-bit)
- Representing Sound (*digitizing*, *sampling*, *quantization, CD*)

We will discuss image and sound representation some more, after we consider…

Data Compression

Save storage space; speed up transmission.

Bandwidth: Bits (bytes) per second

Compression ratio: $\frac{size\_of\_the\_compressed\_data}{size\_of\_the\_uncompressed\_data}$

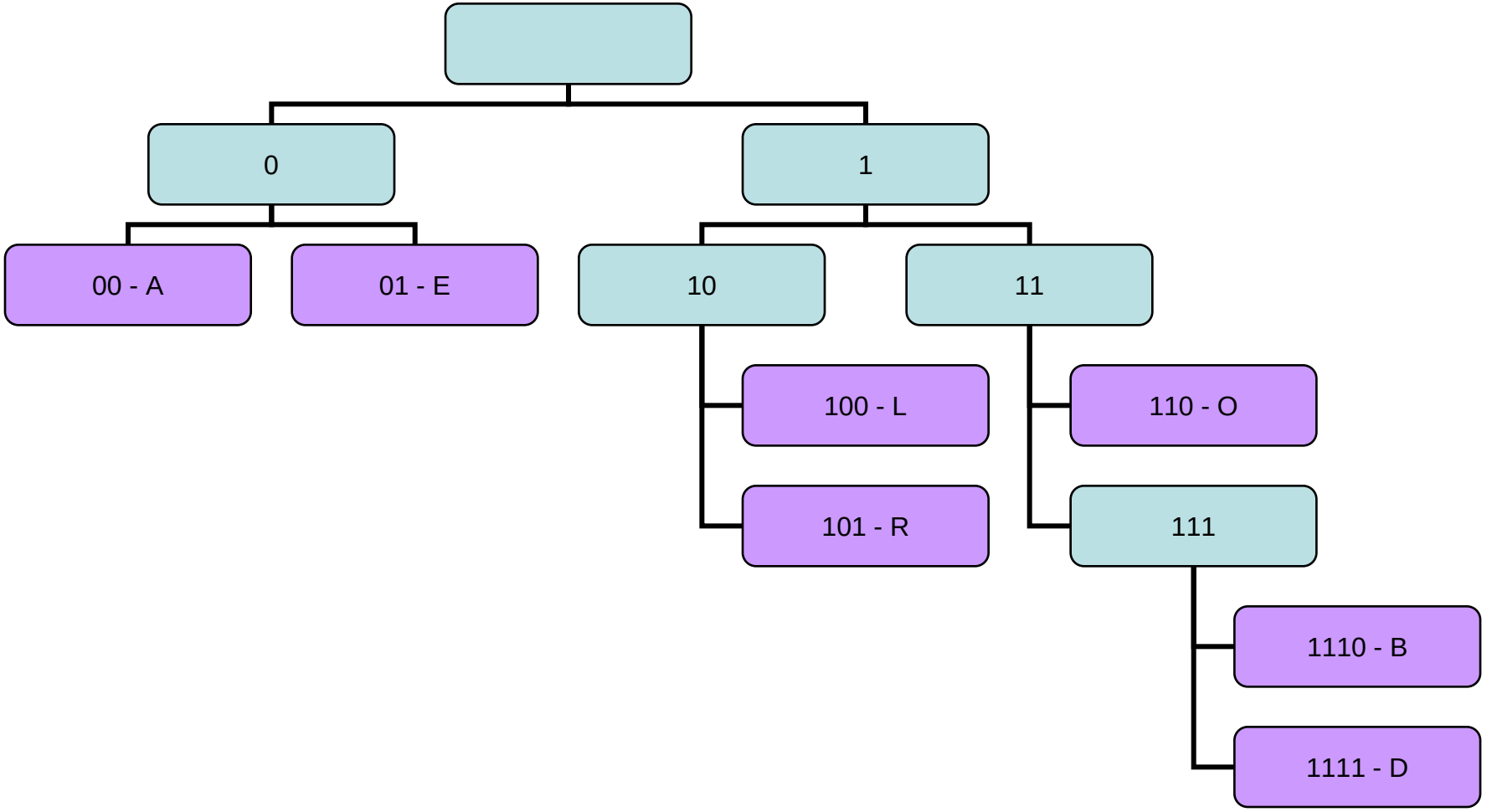
Lossless vs. *lossy* compression

Keyword encoding: Replace a popular word with a shorter code (“with” → “w/”, “without” → “w/o”)

Run-length encoding: AAAAAA → A6

Can combine the two.

Huffman Encoding



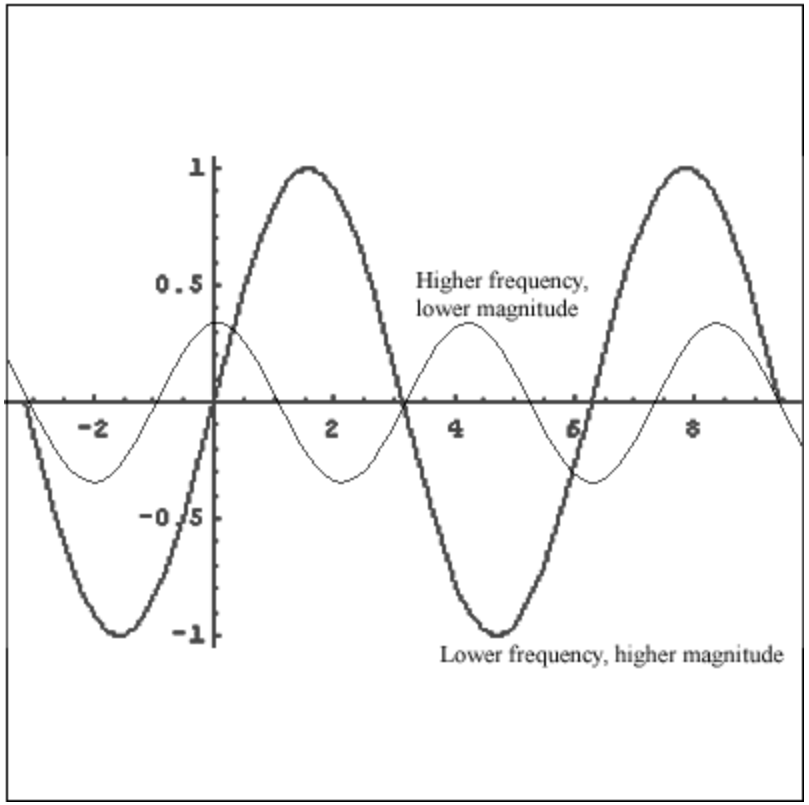
Why would anyone accept *lossy* compression?

Audio formats: WAV, MP3…

Video *codecs* (**co**mpressors / **dec**ompressors): MPEG-2, MPEG-4, DiVX… *Temporal compression* and *spatial compression*

Raster image formats: JPEG, JPEG2000, GIF, PCX, BMP, RAW…

Vector graphics: Images are described in terms of lines and geometric shapes



Signed-Magnitude Representation of Negative Numbers

Add an extra bit on the left to represent the sign.

Use 0 for the ‘+’ sign, 1 for the ‘-’ sign.

Example (3 bits allocated for the magnitude, 1 bit for the sign):

$0101=5_{10}$, $1101=-5_{10}$.

Problems with the *signed-magnitude* representation:

- Two representations of 0: 0000 and 1000;
- Special logic is required to perform addition, subtraction, multiplication and division.

Ten's Complement Representation of Negative Numbers

- Limit the maximum number of decimal digits by *k*.
- Interpret the first half of numbers (0,1,…,(10^k/2)-1) as natural numbers. Interpret the other numbers as

$$\text{Negative}(m) = 10^k - m$$

- Example, *k*=3: 123+(-455)=123+(1000-455)=668_{10c}=-332_{10}

Two's Complement Representation of Negative Numbers

- Representing

$$\text{Negative}(m) = 2^k - m,$$

where *k* is the number of bits used.

Example: *k*=8,

-125	10000011	(256-125=131=1+2+128)
+ 3	+ 00000011	
-122	10000110	"134"=Negative(122))

Overflow will occur if the result of addition exceeds 127: "128" (10000000) now serves as (-128)!

Representing Real Numbers

- *Scientific notation*:

  .00508259 = $5.08259*10^{-3} \rightarrow$ 5.08259E-3

  The *decimal point* is kept to the right of the most significant (non-zero) digit.
- *Floating point*: A real value in Base 10

  r = sign*mantissa*$10^{exponent}$

  The # of digits is fixed, but the point “floats”.
- In other bases, the analog of the decimal point is called a *radix point*.

Representing Real Numbers in Binary

r = sign*mantissa*$2^{exponent}$

How to convert the fractional part from decimal to binary? Keep multiplying by the base and reading off the digits. Example:

$17.875_{10} = 10001.111_2$

17/2=8.5, 0.5*2=1; 8/2=4; 4/2=2;2/2=1;

.875*2=1.75; 0.75*2=1.5; 0.5*2=1.

Representing Text

- Encoding characters vs. *formatting* (fonts, margins, tables, color, etc.)
- A *character set* is a list of characters and the codes used to represent them. How many characters do we need?..
- *ASCII* (American Standard Code for Information Interchange): Originally allowed 128 unique characters. The eighth bit was a *check bit*. *Latin-1 Extended ASCII character set*: 256 characters.

The Unicode Character Set

- 16 bits per character. 2^{16}=65536 unique characters can be represented.
- The first 256 characters in the Unicode set correspond to those of the extended ASCII character set. (“*Backward compatibility*”.)

Networking

- Connections: *cable* / *wireless*
- A *node*, or *host* is any device on the network
- *Data transfer rate* / *bandwidth*
- A *protocol* is a set of rules that defines how data is formatted and processed on a network
- The *client/server model*
- *File servers* and *web servers*

Types of Networks

- *Local-area network (LAN)*
  - *Ring topology* (a LAN configuration)
  - *Star topology* (another LAN configuration)
  - *Bus topology* (yet another LAN configuration)*: Ethernet* (the industry standard)
  - A special node on a LAN may serve as a *gateway*
- *Wide-area network (WAN)*
- *Metropolitan-area network (MAN)*
- The *Internet*

Internet Connections

- The *Internet backbone*
- An *Internet service provider (ISP)*
- Connection via a *cable modem*

} Broadband

- A *digital subscriber line (DSL)*
- A *dial-up* connection via a *phone modem*

The word *modem* stands for **mo**dulator/**dem**odulator.

Download / *upload*

Packet Switching

- Divide a message into *packets*, send them separately, have them collected and reassembled at their destination.
- *Routers* direct packets between networks.
- *Repeaters* strengthen and propagate signals along communication lines.