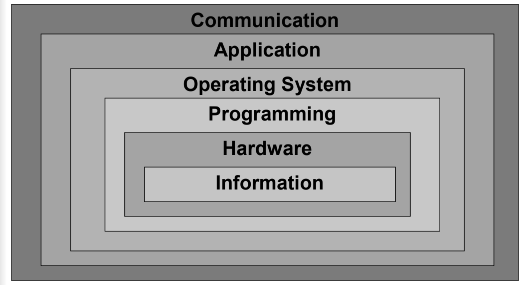
Principles of Computer Science I

Prof. Nadeem Abdul Hamid

CSC 120A - Fall 2004

Lecture Unit 4

Layers of a Computing System

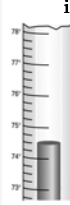


Storing and Using Data

- Computers must deal with many types of data
  - Numbers
  - Text
  - Audio
  - Images and graphics
  - Video
- Computers are finite
  - Must balance computational limits vs. our perception of sight and sound

Representing Information

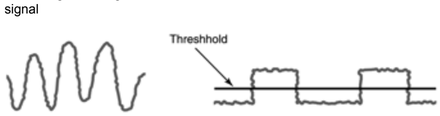
- Analog data
  - a continuous representation, analogous to the actual information it represents
- Digital data
  - a discrete representation, breaking the information up into separate elements



- A mercury thermometer is an analog device– the mercury rises in a continuous flow in the tube in direct proportion to the temperature
- Computers do not work well with analog data– we must *digitize* by breaking range of values into pieces

Electronic Signals

An analog and a digital signal



Degradation of analog and digital signals

Binary Representation

- Devices that store and manage data are less expensive and more reliable if they only have to represent one of two possible values
- One bit can represent two possibilities (0 or 1)
- *N* bits can represent 2^N possibilities
- Integers (we've discussed already)
  - Base-2 numbers
  - Two's complement notation
  - Overflow

Representing Real Numbers

- "Floating point" notation
  - Number of digits fixed but radix point floats
  - Formula: *sign* * *mantissa* * 10^{exp} (base-10)

Real value	Floating point
12001.00	$+12001 * 10^{0}$
-120.01	$-12001 * 10^{-2}$
0.12000	$+12000 * 10^{-5}$
-123.10	$-12310 * 10^{-2}$
155555000.00	$+15555 * 10^{4}$

- Binary floating point uses 2^{exp}

Representing Text

- Use a character set: list of characters and codes to represent each one
- ASCII (American Standard Code for Information Interchange)
  - Originally used seven bits to represent each character, allowing for ??? unique characters
  - Later evolved so that all eight bits were used which allows for ??? characters

ASCII Character Set

Left Digit(s) / *Right Digit*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
						ASCII				
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT
1	LF	VT	FF	CR	SO	SI	DLE	DC1	DC2	DC3
2	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS
3	RS	US	□	!	"	#	$	%	&	'
4	(	)	*	+	,	-	.	/	0	1
5	2	3	4	5	6	7	8	9	:	;
6	<	=	>	?	@	A	B	C	D	E
7	F	G	H	I	J	K	L	M	N	O
8	P	Q	R	S	T	U	V	W	X	Y
9	Z	[	\	]	^	_	`	a	b	c
10	d	e	f	g	h	i	j	k	l	m
11	n	o	p	q	r	s	t	u	v	w
12	x	y	z	{	|	}	~	DEL		

Unicode Character Set

- Uses 16 bits per character
  - Can represent 2^{16}, or over 65 thousand, characters
- Designed to be a superset of ASCII
  - First 256 characters correspond to ASCII

Code (Hex)	Character	Source
0041	A	English (Latin)
042F	Я	Russian (Cyrillic)
OE09	ฉ	Thai
13EA	Ꮺ	Cherokee
211E	℞	Letterlike Symbols
21CC	⇌	Arrows
282F	⠯	Braille
345F	㑟	Chinese/Japanese/Korean (Common)

Data Compression

- Reducing the amount of space (bits and bytes) needed to store a piece of data
- Compression ratio
  - size of the compressed data divided by the size of the original data
- Lossless: data can be retrieved without losing any of the original information
- Lossy: some information is lost in the process of compaction
- Common techniques
  - keyword encoding
  - run-length encoding
  - Huffman encoding

Keyword Encoding

- Frequently used words replaced with a single character
- Characters used to encode cannot be part of the original text

Word	Symbol
as	^
the	~
and	+
that	$
must	&
well	%
those	#

Run-Length Encoding

- Single character may be repeated over and over again in a long sequence
  - Not in English text, but often occurs in large binary data streams
- Sequence of repeated characters is replaced by
  - a *flag character*,
  - the repeated character,
  - a single digit indicating how many times the character is repeated.

RLE Example

- AAAAAAA would be encoded as: *A7
- *n5*x9ccc*h6 some other text *k8eee would be decoded into the following original text:

nnnnnxxxxxxxxxccchhhhhh some other text kkkkkkkkeee

  - Original text: 51 characters
  - Encoded string: 35 characters
  - Compression ratio of 35/51 (approximately 0.68)

- Can we encode repetition lengths greater than 9?

Huffman Encoding

- Why should the character "X", seldom used in text, take up the same number of bits as the blank, used very frequently?
- Huffman codes using variable-length bit strings to represent each character
  - Few characters may be represented by five bits, another few by six bits, yet another few by seven bits, and so on…

Huffman Code Example

Huffman Code	Character
00	A
01	E
100	L
110	O
111	R
1010	B
1011	D

- DOORBELL would be encoded as: 1011110110111101001100100

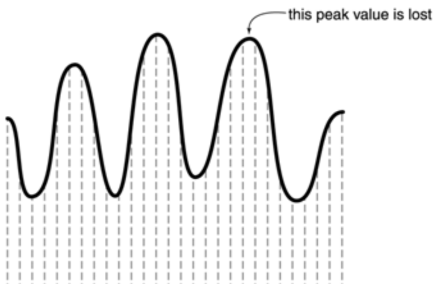
Huffman Encoding (cont.)

- Using fixed-size bit string to represent each character (say, 8 bits), the binary form of the original string would be 64 bits
- Huffman encoding is 25 bits long
- Compression ratio of 25/64 (approximately 0.39)
- Important characteristic
  - no bit string used to represent a character is the prefix of any other bit string used to represent a character
- To generate codes, figure out frequencies…

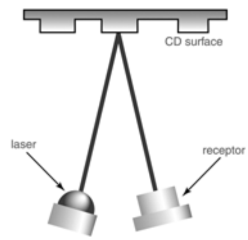
Representing Audio

- We perceive sound when a series of air compressions vibrate a membrane in our ear
- Stereo sends an electrical signal to a speaker to produce sound
  - Signal is analog representation of sound wave
  - Voltage in signal varies in direct proportion to the sound wave
- Digitizing the signal: *sampling*
  - Periodically measure voltage of signal and record appropriate numeric value
  - Sampling rate of around 40,000 times/second enough to create reasonable sound reproduction

Sampling (in general)



Storing Audio on a CD



- Surface of the compact disk (CD) has microscopic pits to represent binary digits
- Low intensity laser pointed as the disk
- Laser light reflects strongly if surface is smooth; reflects poorly if surface is pitted

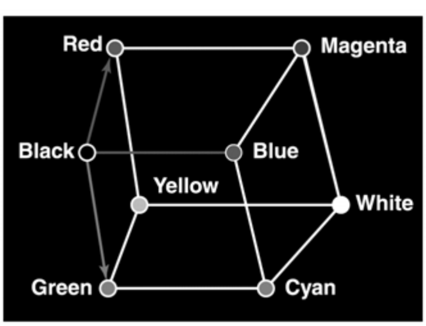
Audio Formats

- Popular formats: WAV, AU, AIFF, RA, MP3
- Currently dominant format for compressing audio data is MP3
- MP3 uses both lossy and lossless compression
  - Analyzes the frequency spread and compares it to mathematical models of human psychoacoustics (the study of the interrelation between the ear and the brain)
  - Discards information that can't be heard by humans
  - Bit stream is then compressed using form of Huffman encoding

Representing Images and Graphics

- Color is our perception of various frequencies of light reaching our retinas
- Three types of color photoreceptor cone cells in the retina
  - Respond to different sets of frequencies
  - Correspond to the colors of red, green, and blue
- Color on a computer: RGB value (three numbers indicating relative contribution of each primary color)
- (255, 255, 0) maximizes contribution of red and green, and minimizes contribution of blue; results in a bright yellow

Color Cube



RGB Values

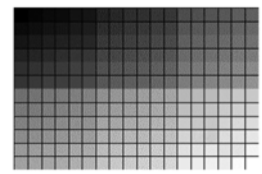
RGB Value			Actual Color
Red	Green	Blue	
0	0	0	black
255	255	255	white
255	255	0	yellow
255	130	255	pink
146	81	0	brown
157	95	82	purple
140	0	0	maroon

Color Depth

- Amount of data used to represent a color
- *HiColor* indicates a 16-bit color depth
  - Five bits used for each number in RGB value and extra bit sometimes used for transparency
- *TrueColor* indicates a 24-bit color depth
  - Each number in RGB value gets eight bits
- More bits used for RGB values = more different colors can be represented

Indexed Color

- Particular application may support only a certain number of specific colors
  - Creates a palette from which to choose
  - *E.g.*, Netscape Navigator's color palette:



Digitizing Images and Graphics

- We can digitize an image by representing it as a matrix of dots, called pixels
- Each pixel stores an RGB value
- The number of pixels used is called the resolution
- This method of representing an image (on a pixel-by-pixel basis) is called raster graphics format
- Common raster file formats
  - BMP (Windows bitmap), GIF, JPEG

Vector Graphics

- An alternative to raster graphics format
- Describe an image in terms of lines and geometric shapes
  - Define shapes using mathematical functions
  - Specify line directions, colors, thicknesses, *etc.*
- Files can be smaller in size because we don't specify every pixel in the image
- Can be resized (mathematically) without distortion
- But not really suitable for real-world images

Representing Video

- Video is a series of images
  - 30 or 60 frames per second, for example
  - Huge amount of data in uncompressed form
- Not feasible to represent without compression
- A video codec (COmpressor/DECompressor) refers to methods used to shrink the size of a movie
- Almost all video codecs use lossy compression to minimize the huge amounts of data associated with video
- Types of compression
  - Temporal compression: looks for differences between consecutive frames (don't store repeated information)
  - Spatial compression: removes redundant information within a frame (compress individual images)