

# 1 Internet Fundamentals

Tim Margush

Based on Programming the WWW by Robert Sebesta, 4<sup>th</sup> Edition

## 2 WWW

- World Wide Web
  - What is it?
  - When was it created?
  - Who created it?
  - Why was it created?
  - What technologies make it possible?
  - How big is it?

## 3 What is the WWW?

- A collection of documents and the software and protocols used to access them over computer networks
- The WWW is a software layer that relies on the Internet for functionality

## 4 WWW: When? Who? Why?

- 1989
- Tim Berners-Lee at CERN
  - European Laboratory for Particle Physics
  - First WWW page
    - <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>
- To locate and retrieve scientific documents anywhere on the Internet

## 5 Technologies

- Primarily the Internet
- Web Browsers and Servers
- HTTP
- URL
- MIME

## 6 Internet?

- Collection of computers connected via a communications network
- A common protocol for communications
  - TCP/IP
- Origins go back to 1969 as ARPAnet and NSFnet
  - Competing networks were assimilated
    - Resistance is futile

## 7 TCP/IP

- Transmission Control Protocol / Internet Protocol
- IP Address 130.101.10.1
- Domain Name [www.cs.uakron.edu](http://www.cs.uakron.edu)
  - Domains are a collection hierarchy
  - Top-level domains: com, edu, net, se, uk, ...

## 8 [www.cs.uakron.edu](http://www.cs.uakron.edu)

## 9 IP Address

- 32 bit number – Bytes expressed in decimal and joined with dots
  - 103.101.10.1
- Addresses are assigned to organizations in blocks and they assign internal addresses.
  - uakron.edu is assigned 130.101.x.x
  - This is a class B address

## 10 IP Address Classes

- Class A now called /8 blocks
  - $0 \leq 1^{\text{st}} \text{ byte} \leq 126$  (0xxxxxxx)
  - 16,777,214 internal addresses
    - General Electric 3.0.0.0 through 3.255.255.255
- /16 blocks (65536 internal) (Class B)
  - $128 \leq 1^{\text{st}} \text{ byte} \leq 192$  (10xxxxxx)
    - University of Akron **130.101.0.0** through **130.101.255.255**
- /24 blocks (Class C)
  - $193 \leq 1^{\text{st}} \text{ byte} \leq 233$  (110xxxxx)
    - Marietta College **199.218.109.0** through **199.218.109.255**

## 11 IPv6

- The 32 bit IP addresses currently in use are also called IPv4
  - Deployed 1983
  - Projections that all addresses would be used up by 2003 were incorrect
  - Classless IP assignments are extending the life of IPv4
  - Current projections for the lifetime of IPv4 vary
- IPv6 is a new 128 bit IP address assignment thought necessary to support a growing Internet
  - Deployed in 1999 but not widely used
  - Expressed in eight 2-byte (hex) sections
    - 1080:0:0:0:8:800:200C:417A

## 12 DNS Server

- www.cs.uakron.edu → 130.101.10.1 ?
- ICANN
  - The Internet Corporation for Assigned Names and Numbers
  - Manages 13 root servers holding the addresses of the top level domain (TLD) registries
- Domain names are translated using a hierarchical database system
  - Thousands of Domain Name Resolvers cache data from the root servers – translation requests go to one of these – my resolver is currently 65.24.7.3 (dns-cac-lb-01.ohiordc.rr.com)
  - Resolver asks the edu registry for the IP of uakron.edu and receives 130.101.5.4 back (dax.cc.uakron.edu)
  - Then it asks the local DNS server (dax.cc.uakron.edu) for the address of www.cs.uakron.edu and gets back **130.101.10.1**

## 13 Other Protocols

- ftp – file transfer
- smtp – simple mail transport
- Usenet – user network discussion boards
- telnet – telecommunication network
- Gopher – "go for" document retrieval
- snmp – simple network management

## 14 HTTP

- Hypertext Transfer Protocol
  - Standard for communications between browsers and WWW servers
- The Mosaic browser ushered in the graphical WWW in 1993
  - Based on Gopher
  - Funded by Al Gore's Bill (HPCA) in 1991
    - "During my service in the United States Congress I took the initiative in creating the internet. I took the initiative in moving forward a whole range of initiatives that have proven to be important to our country's economic growth and environmental protection, improvements in our educational system." Al Gore in an interview on March 9 1999

## 15 MIME

- Multipurpose Internet Mail Extensions
  - type/subtype
    - text/plaintext/html
    - image/gifimage/jpg
    - audio/mp4
  - Allows browser to determine what application is needed to render the resource
    - Servers generally add MIME types based on the filename's extension
      - text/plainsomething.txt
      - txt/html something.html
      - image/gifsomething.gif

## 16 Browser

- Program that primarily communicates with a WWW server using HTTP
- Browsers make HTTP requests to retrieve resources and display (render) their contents
  - For some resources, another application is loaded to handle content

## 17 WWW Server

- A program designed to handle HTTP requests and serve the requested resource
  - Monitors an Internet port
  - Responds to HTTP commands in the form of a URL
- The two most common servers are
  - Apache
  - Microsoft Internet Information Server (IIS)

## 18 URL

- Uniform Resource Locator
  - Uniquely identify resources on the Internet
- scheme:object-address
  - http://www.cs.uakron.edu/index.html
  - file://C:/docs/www/root
- HTTP object-address is in two parts
  - //Fully qualified domain name
    - www.cs.uakron.edu
  - /Path to document
    - index.html

## 19 Missing Parts?

- Servers add details to the path information
  - A directory prefix is added to resolve the root of the www server file space
    - http://www.cs.uakron.edu/index.html
    - **/usr/local/www/data/index.html**
  - A standard filename might be added
    - http://www.cs.uakron.edu
    - http://www.cs.uakron.edu/**index.html**

## 20 HTTP Protocol - Request

- Request
  - Method    path            HTTP version
  - Header fields
  - Blank line
  - Message body
- Methods
  - GET HEAD POST PUT DELETE

## 21 Telnet HTTP GET

1

```
telnet www.cs.uakron.edu http
GET /index.html HTTP/1.0
```

```
HTTP/1.1 200 OK
Date: Mon, 31 Dec 2007 03:29:50 GMT
Server: Apache/2.0.46 (Red Hat)
Last-Modified: Sun, 09 Jan 2005 03:00:18 GMT
ETag: "3a0004-153-a6d29480"
Accept-Ranges: bytes
Content-Length: 339
Connection: close
Content-Type: text/html; charset=ISO-8859-1
```

2

```
<HTML>
<HEAD>
<TITLE>University of Akron Computer Science</TITLE>
</HEAD>
<FRAMESET COLS = "20%,*">
  <FRAME NAME = "left" SRC = "left.html" FRAMEBORDER = "no" MARGINHEIGHT = "0"
    MARGINWIDTH = "0">
  <FRAME NAME = "right" SRC = "right.html" FRAMEBORDER = "no" MARGINHEIGHT = "0"
    MARGINWIDTH = "0">
</FRAMESET>
</HTML>
```

Connection to host lost.

## 22 HTTP Response

1

- Status line

- Response Header Fields

2

```
■ Blank Line
■ Response Body
HTTP/1.1 200 OK
Date: Mon, 31 Dec 2007 03:29:50 GMT
Server: Apache/2.0.46 (Red Hat)
Last-Modified: Sun, 09 Jan 2005 03:00:18 GMT
ETag: "3a0004-153-a6d29480"
Accept-Ranges: bytes
Content-Length: 339
Connection: close
Content-Type: text/html; charset=ISO-8859-1
```






```
<HTML>
<HEAD>
```

## 23 Security

- Encryption is one of the fundamental security tools used in Internet transactions
- Before 1976, an encryption key had to be secretly communicated between parties, creating a chicken/egg problem
- Diffie and Hellman invented public key encryption system that uses two-part keys
  - The public key is freely shared with the world
  - Each person keeps their own private part of the key

## 24 Public Key Encryption

- Joe wants to send a message to Mary
- Joe uses Mary's public key to encrypt the message.
- The message is sent over public channels.
- Only Mary can decrypt it with her private key

- 25  **Public Key Authentication**
- Mary needs to be sure the message is from Joe
  - Joe encrypts the encrypted message using his private key.
  - The double-encrypted message is publically sent.
  - Mary uses Joe's public key to decrypt the message... then her private key to decrypt it again
    - Only Joe could have applied the outer encryption
- 26  **WWW Documents**
- HTML – Hypertext Markup Language
    - Marks are placed around content to indicate how it is to be formatted
  - XHTML 1.1 is the current hypertext markup language standard
    - Most WWW documents do not conform to the standard
    - You should learn to write strict XHTML but not be afraid to violate the standards
- 27  **W3C**
- 1 ■ The WWW Consortium develops and maintains standards for XHTML and a variety of other WWW technologies
    - HTML was introduced in 1989 by Tim Berners-Lee
  - 2 1993 the first working draft published  
1996 HTML 2.0  
Jan 1997 HTML 3.2 – fonts, tables, applets  
Dec 1997 HTML 4.0 – style sheets  
Dec 1999 HTML 4.01 – Final form of HTML  
Jan 2000 XHTML 1.0 – Reformulates HTML 4.01 in XML  
May 2001 XHTML 1.1 – Modularizes XHTML for various devices
- 28  **XHTML Document Creation**
- Text file editor
    - Notepad, PSPad
  - Dedicated XHTML editors (WYSIWYG)
    - DreamWeaver, FrontPage, PageMill
  - Conversion Filters or Plug-ins
    - Convert documents created in various applications to HTML format
      - Word, WordPerfect, Excel, PowerPoint ...
- 29  **Other WWW Languages**
- XML
    - eXtensible Markup Language
  - JavaScript
  - Java
  - Perl
  - PHP