

CS2204 Fundamentals of Internet Applications Development

- 1. The internet**
 - 1.1. History of development**
 - 1.2. What is TCP/IP?**
 - 1.2.1. TCP/IP layers
 - 1.3. IP address**
 - 1.4. Domain Name**
 - 1.5. Common internet applications**
 - 1.5.1. Email
 - 1.5.2. World Wide Web
 - 1.5.2.1. W3C, WHATWG and Web technologies
 - 1.6. The universal resource locator (URL)**
 - 1.6.1. Port number
 - 1.7. Architecture of Web Applications**
 - 1.7.1. Typical architecture
 - 1.7.2. Browser
 - 1.7.3. Browser usage share
 - 1.7.4. Web Server
 - 1.7.5. Evolution of Web Technologies
 - 1.8. Miscellaneous useful concepts of Internet**
 - 1.8.1. Intranet
 - 1.8.2. Firewall
 - 1.8.3. Extranet & VPN
 - 1.9. Conclusions**

1. The internet

Isn't it just another computer network? What exactly is it?

A computer network is a connection of computers through which each computer can talk to each other

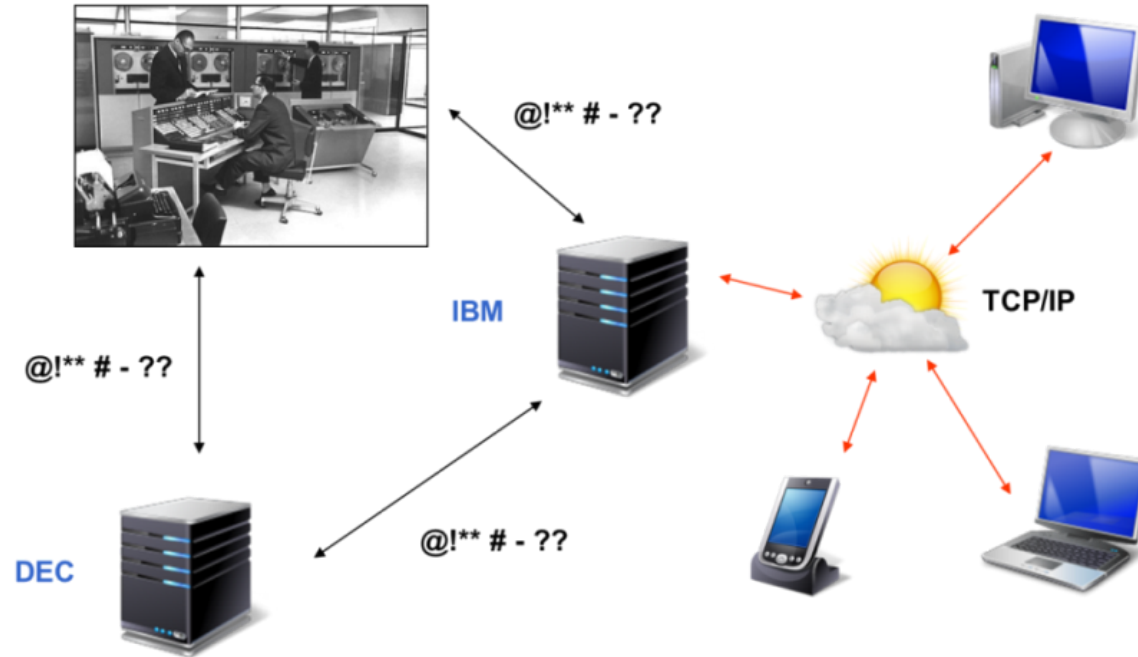
The left hand side network in the diagram below shows a pre-Internet network. Different brands of computers use different languages to talk to each other. Easy to communicate between same kind of computer but could not understand each other for different brands

Right hand side network is modern day Internet. Different brands, and even different kinds (notebook, servers, mobile devices) can talk to each other seamlessly

Reason? They all speak the same language. This is a case in which the power of **standards** can be seen. A more specific technical term for the language is **protocol**. It is a well defined set of agreement for information exchange

The protocol is **TCP/IP**

Programmed Data Processor (PDP)



1.1. History of Development

1.1. History of Development

The Advanced Research Project Agency (ARPA) of US wanted to develop a new **decentralized** network. 4 universities were invited: University of Utah; Stanford Research Institute; University of California, Santa Barbara; and University of California, Los Angeles

Provide remote login, file transfer and remote printing service

In 1983, TCP/IP was adopted as the standardized protocol and the Internet was born

Number of sites rapidly grew to more than 500

In 1990 private and commercial use took over academia and research. Tremendous growth in use followed.

1.2. What is TCP/IP?

1.2. What is TCP/IP?

The full name is Transport Control Protocol/Internet Protocol. This is in fact not a single protocol but a suite (**stack**) of different protocols

Defining a protocol between 2 computers is complicated and involves different components. It is easier to break it down into layers (hence the name stack) for understanding and reuse.

Application Layer	TELNET, FTP, SMTP	PING, TFTP
Transport Layer	TCP	UDP
Network Layer	IP, ICMP	
Link Layer	Device Driver and Interface Card	

1.2.1. TCP/IP layers

1.2.1. TCP/IP layers

There is a layer not shown - physical layer, describing the physical hardware which can be a network interface card (NIC), a wifi hardware or SIM card

Link layer interfaces with the physical layer. A layer can be considered as a software library with program call interfaces provided to the upper layer. Upper layers such as network/transport layers can access the hardware through the link layer without knowing the detail hardware characteristics.

Network layer is responsible for device/computer identification and routing (first locate the device then find the path to it).

Transport layer is used to actually send and receive data in binary form through the set up path.

Application layer addresses the issue that binary data are difficult to use, therefore more high level (meaningful to human) messages are created. These sets of messages are application specific, i.e. a particular application would use one kind of application protocol. Many applications have been created since the creation of Internet. Sometimes TCP/IP only refer up to the Transport layer.

1.3. IP address

There are different ways to identify a device/computer, e.g. serial number or MAC address. Since the Internet would connect all the devices in the world together, the identification has to be unique and allocated under a common system. This identification is **IP Address**

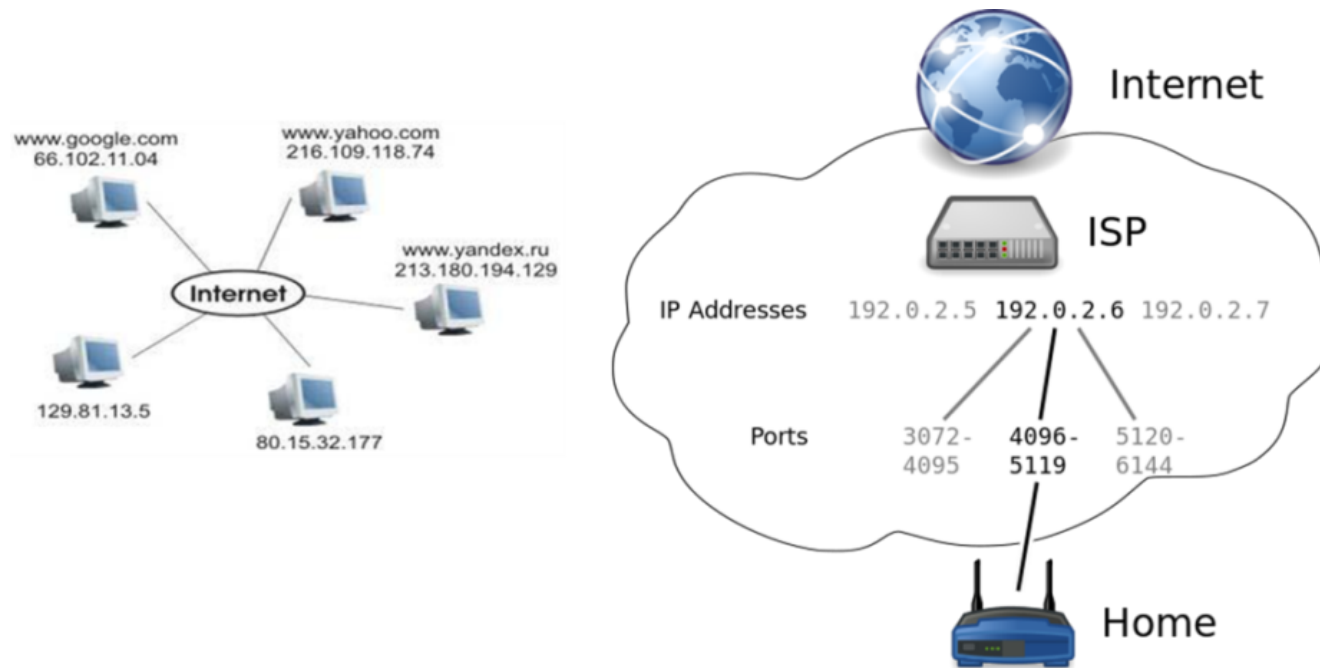
The allocation is managed by a no. of organizations, Internet Assigned Names Authority, iana.org (IANA), Internet Corporation for Assigned Names and Numbers, icann.org (ICANA) and regional centers, such as Asia Pacific center.

It is a 32-bits number, divided into four octets and written in a dotted-decimal notation (IPv4), e.g. 144.214.36.117 with some numbers reserved for special use:

127.0.0.1 – local host

192.168.1.1 – private (home) network

How many devices can be identified with this scheme? Not much, only 4G because of 32 bits. Another scheme is therefore needed (IPv6) using 128 bits in 8 sections.



1.4. Domain Name

1.4. Domain Name

The IP address is a number used in protocols but it is difficult to remember by human. A more meaning text based naming is therefore developed.

Note that the IP address is in sections (no matter it is IPv4 or IPv6) which are in levels, e.g. 144.214 represents all devices in CityU.

Domain name is text version of a section of IP address, e.g. cityu.edu.hk, microsoft.com, etc.

Common domains are: top-level domain names for organizations, .com, .edu, .gov, .mil, .net, .org; Geographical name domains: .hk, .ca, .to, .jp, .kr

When domain names are used by us, program (e.g. browser) will ask other computer systems to help translate to the numeric IP address. These computers are known as **Domain Name Servers**.

1.5. common internet applications

To understand better how the Internet operates, let's check out some common Internet applications and their protocols. Internet applications are not normal computer system applications used in stand-alone computers or Apps in mobile phone. They all use TCP/IP as the underlying protocol.

Remote login (protocol telnet) - used to login a remote computer (**host**) and interact with commands, e.g. the famous Unix computer system.

File transfer (ftp) - used to send (big) binary files to remote computers.

Email (smtp, imap and pop) - used to send and receive emails

World Wide Web (http) - the Internet Application that changes the world and is the focus of this course

Note that all Internet Applications operate in a **client-server** mode involving two devices, a connection and an application protocol. In this mode, programs are commonly named as such: e.g. ftp & ftpd, telnet & telnetd, browser & httpd etc. where d stands for the server program.

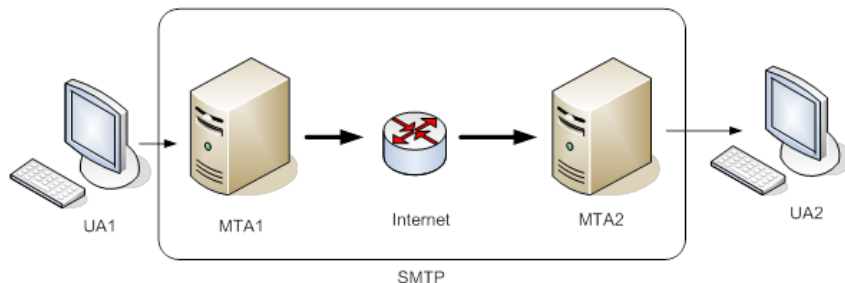
1.5.1. email

Ftp and telnet protocols will be further demonstrated in tutorials. Let's use email to elaborate the concepts of client-server, connection and protocols.

The path of email flow can be shown in the diagram below. UA (User Agent) is a program that interfaces to the user (the client). Through the UA (e.g. UA1), a user creates an email and sends to an Outgoing Mail Server (the server).

The email receiver could be very far away and the email may go through many intermediate MTAs in the Internet until it reaches the receiver's mailbox (identified by the email address) which is located in an Incoming Mail Server. To relay these message from one MTA to another, smtp protocol is used.

When the email receiver wants to read the email, another UA (UA2) has to be used to connect to the Incoming Mail Server using protocols imap or pop.



1.5.2. World Wide Web

Can be said as one of the most important Internet Application using Hyper Text Transfer Protocol (HTTP). Browser is the client and Web server is the server. Using the Web, multimedia documents can be located, linked and viewed (i.e. why called the Web).

Developed by Tim Berners-Lee of the European Laboratory for Particle Physics (CERN) in 1990. He wrote the first versions of HTTP and HTML standards.

Another famous person is Marc Andreessen who wrote the first widely used browser – Mosaic, later became Netscape. Before that, browsers could only use text commands without images!



1.5.2.1. W3C, WHATWG and Web technologies

Tim Berners-Lee found the World Wide Web Consortium (W3C) in 1994, www.w3.org with a primary goal to make the Web **universally** accessible regardless of disability, language and culture. Knowing the importance of standardization but at the same time encouraging diversity and competition, W3C is devoted to develop nonproprietary, interoperable technologies for the Web.

W3C initially was composed of 3 hosts : MIT, INRIA (Institut National de Recherche en Informatique et Automatique) & Keio University, with over 400 members from all over the world.

Standardized technologies are proposed and managed in stages known as Technical Report, Development, Working Draft, Candidate Recommendation and Recommendation (also known as a specification or standard).

WHATWG (Web Hypertext Application Technology Group) is another group working on Web standards, sometimes along side with W3C. It was formed in 2004 initially because of the disagreement with W3C's direction and slow process. HTML5 is the main outcome of this group.

1.6. the universal resource locator (URL)

A resource is a useful entity in the Web. It could be a Web page, an Excel sheet, a movie or song. Since every resource can be linked in the Web, we need a way to identify, point to and access the resource.

A URL - <http://wbi.cityu.edu.hk:3638/.../index.html>, consists of 4 parts:

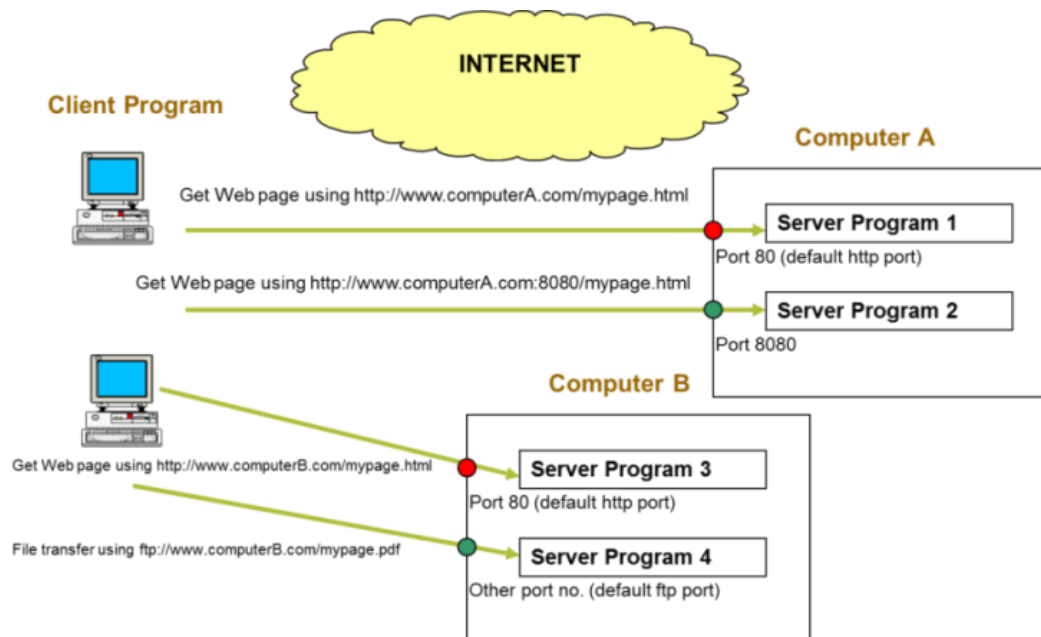
- 1st part, (application) protocol designator
- 2nd part, domain name: e.g. www.cs.cityu.edu.hk
- 3rd part, port number: 23, 80, 3638
- 4th part, path of a specific path of folder structure to access the resource, e.g. .../cs2204/example/image/ref01.png

There are many different protocols, common ones are:

- FTP Service: <ftp://ftp.cityu.edu.hk>
- Mail Service: <mailto:csmtchan@cityu.edu.hk>
- Local file: <file:///c:/myfile/myfirst.html>

1.6.1. port number

The diagram below shows the meaning of port number. The domain name in a URL only points to a host(computer) but inside there could be more than one server program running to provide service, e.g. you can run two Web servers in one computer. Port number can then be used to differentiate the services. Usually a default port number has been allocated for common applications/protocols like ftp, email or http.



1.7. Architecture of Web Applications

1.7. Architecture of Web Applications

Any program that uses Internet technologies i.e. TCP/IP can be regarded as Internet Application but in this course we only focus on one - **Web Apps**

You should appreciate advantages of Web Apps and be able to explain their popularity:

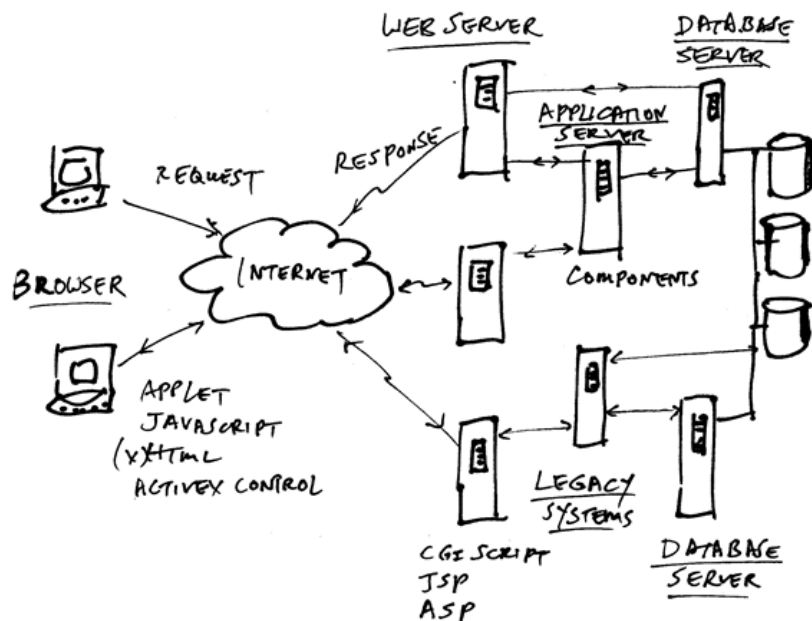
- Standard user interface - nearly everyone knows how to use a browser
- Write once, can be used in different devices
- Can link to existing DBMS and legacy systems - wrap a Web interface and deliver the service to the world
- Highly scalable - keep adding servers to expand capacity
- Distributed, component based and reusable - do you know which Google server you are using?

1.7.1. typical architecture

1.7.1. typical architecture

A typical architecture of Web App System is called N-tier system:

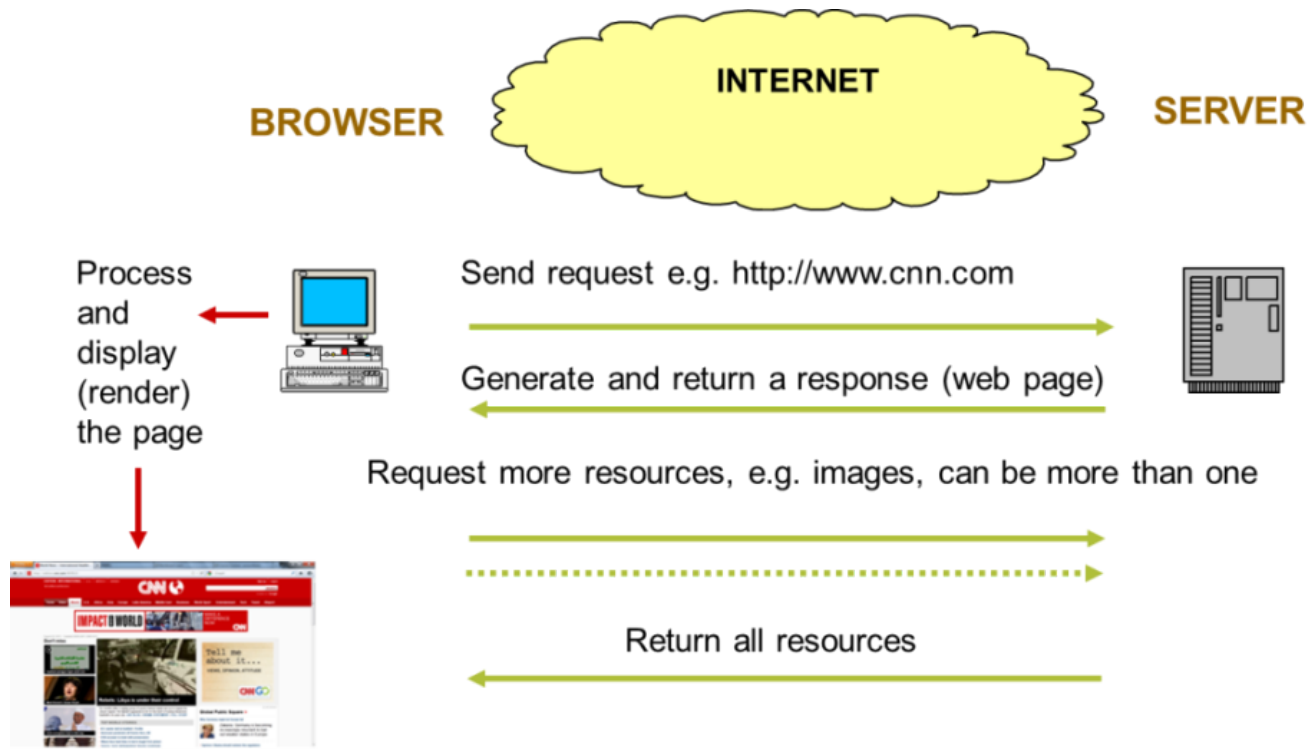
- Browser is the 1st tier (client)
- Web servers - 2nd tier (server)
- Other servers (database and application servers) linked to Web servers are N tiers



1.7.2. Browser

Browser is a program that can talk to a Web server using http, render the Web page sent from server and interpret (run) Javascript. It becomes the most successful universal client to access different kinds of information in the Web. Common products : Internet Explorer/Edge, Opera, Firefox, Safari, Google Chrome, and many others.

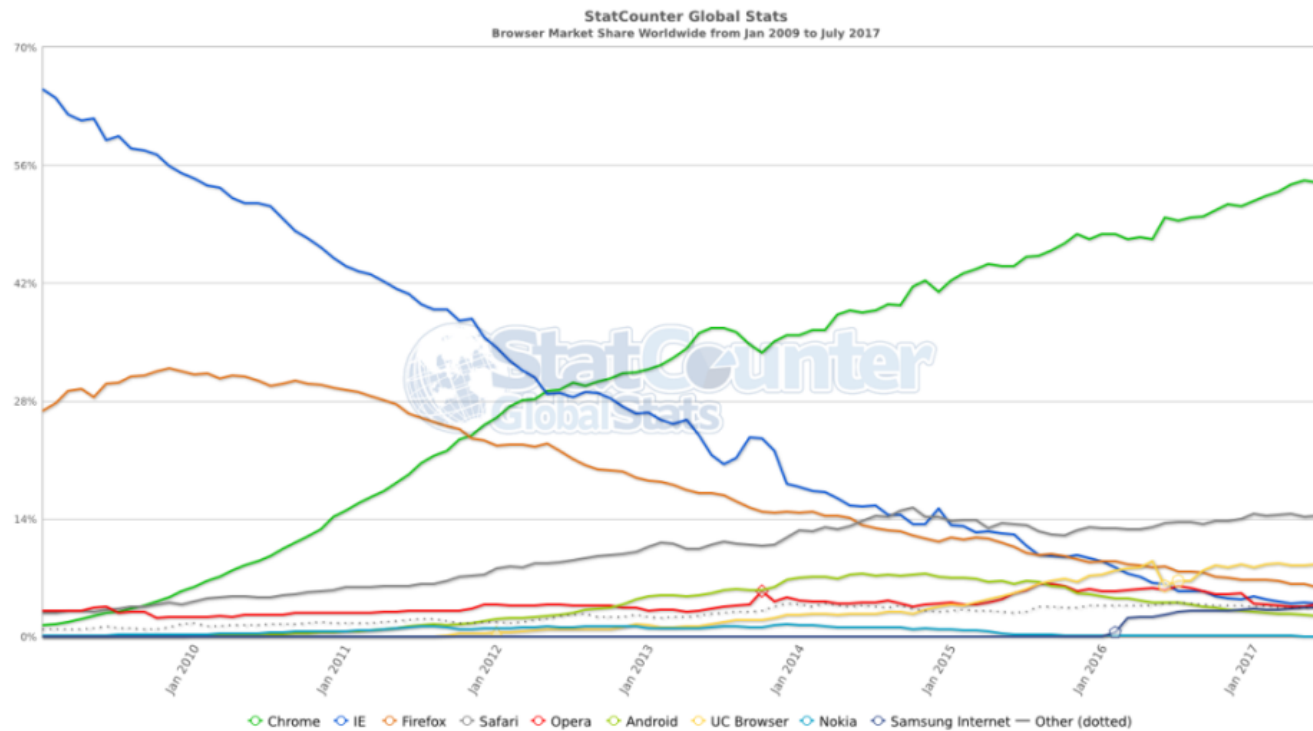
The main issue when considering browser in our course is compatibility. You should test your pages with all browsers your promised to support and always use standard techniques as far as possible.



1.7.3. Browser usage share

1.7.3. Browser usage share

Understand which browser should be supported

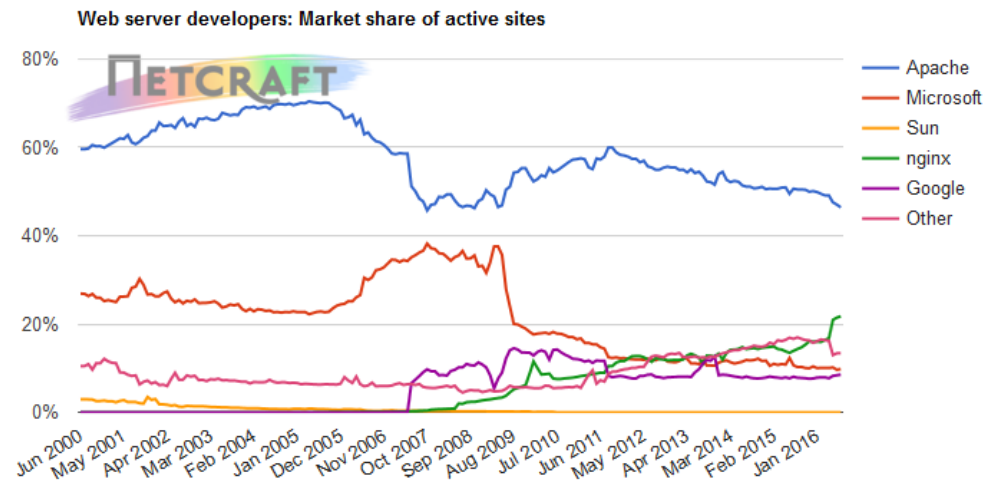


1.7.4. Web Server | Browser Usage - https://en.wikipedia.org/wiki/Usage_share_of_web_browsers

1.7.4. Web Server

Web Server is a server program (always running until shutdown) that reads the HTTP port (default 80), delivers or builds the Web page by running server side scripts such as CGI scripts/JSP/ASP or activates components in Application Servers

Common Products : Microsoft IIS, Apache, Oracle iPlanet, IBM WebSphere, Tomcat, node.js and nginx etc. The main issue here is choice of back end technologies. Should also consider the usage share.



1.7.5. Evolution of Web Technologies

Some highlights from the beginning around 1990 :

- Static page - retrieval of unchanged information
- Enhanced browser processing - Applet, ActiveX, scriptlet, DHTML, plug-ins
- Dynamic page, connecting to DBMS server-side scripting, CGI scripts, ASP/JSP, use of server objects, JDBC/ODBC, ADO, JDO
Extensive use of components - DCOM, Java Bean, EJB
- Web services - XML programmatic interface, .NET, WSDL, SOAP, UDDI
- RIA (Rich Internet Applications) - ajax, Silverlight, HTML5, etc. Javascript takes over full programming in the client side
- Mobile device Webapp HTML5/CSS3/Javascript with native device library

Scripts refer to command languages usually executed by interpretation (script engine), e.g. Unix Shell script, Windows batch file, Javascript & VBScript, etc.

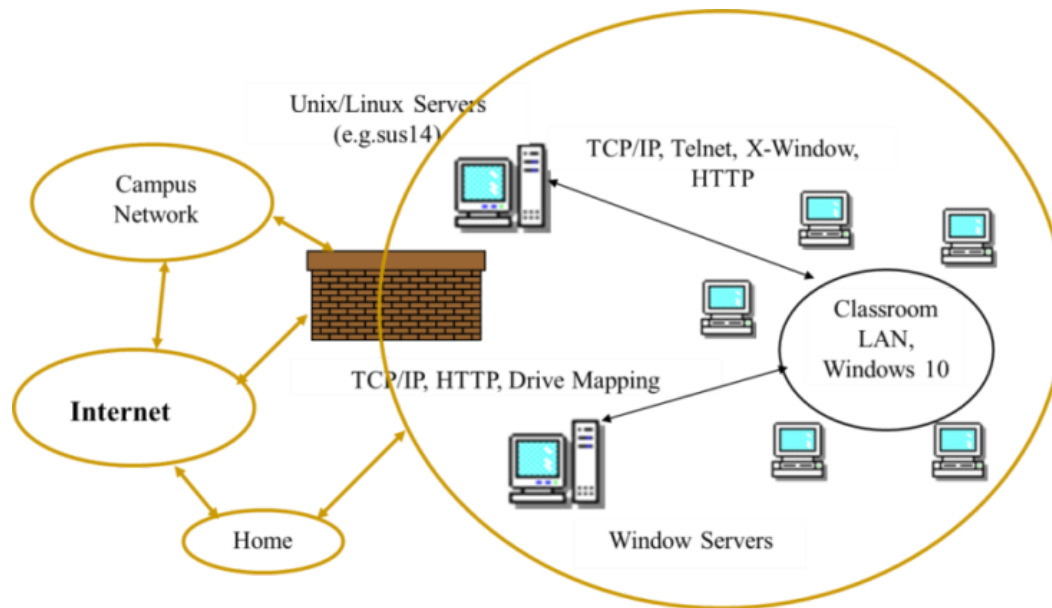
- Client-Side Scripting - scripts are embedded in the HTML and sent to the browser, executed in the client side (browser)
- Server-Side Scripting - executed in the server side (Web Server), scripts are embedded in the HTML file, interpreted by Web server first to produce result

(usually HTML also) which are then sent to the browser; mainly used to produce dynamic page content

1.8. Miscellaneous useful concepts of Internet

1.8. Miscellaneous useful concepts of Internet

Use our CS lab classroom network as an example.



1.8.1. Intranet

1.8.1. Intranet

Using the standard Internet technologies on a private network (behind firewall), can be a small LAN or a large network of LANs, all using of TCP/IP & Web technologies

Using Firewall to protect the private information from the public Internet access.

Same set of applications can be deployed within organization or outside organization but need to be aware of different security requirements.

1.8.2. Firewall

1.8.2. Firewall

A gateway server located at a network's perimeter that protects the resources of a private network from outside users.

Can control incoming and outgoing traffic by opening or closing ports. May also serve as Proxy Server providing a caching mechanism to speed up the Internet connectivity and filtering web content

Can exist as a server, router or even a program in a computer (like in Windows)

1.8.3. Extranet & VPN

1.8.3. Extranet & VPN

Extranet is a network connecting various Intranets of **different locations**. The Intranets can be from the same organization, business partners, customers, suppliers, and other independent enterprises who want to share data and applications over the Internet with protection.

It is connecting different Intranets through the use of public Internet instead of setting up a private network. VPN technologies will be used.

Virtual Private Network (VPN) is a private data network that makes use of a tunneling protocol and security procedures, primarily encryption, to maintain privacy over the Internet, typically for employee remote access of Intranet resources.

1.9. Conclusions

Some critical thinking questions:

- Which browser is better?
- Which browser should I choose?
- Why should I follow the trend of technologies and how?

Externsion readings :

- Browser history
- The World Wide Web: past, present and future
- Evolution of the Web