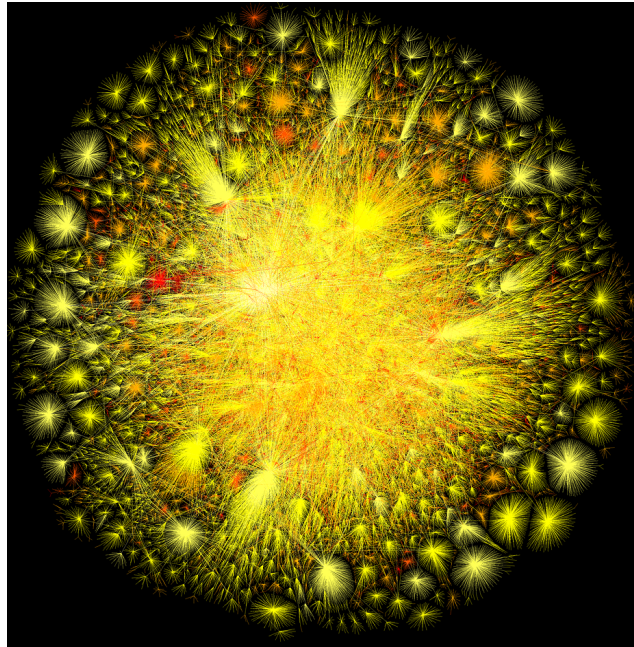


THE INTERNET



A Visualisation of the networks

THE INTERNET PROTOCOL STACK

Layers	Protocols	Offer
Applications	HTTP, FTP, SMTP, DNS	
Transport	TCP, UDP, ICMP	sessions, reliability...
Network	IP (v4, v6)	routing, addressing
Link	Ethernet, 802.11 (ARP)	addressing local machines
Physical	Ethernet, 802.11 (physical)	

IP

INTERNET PROTOCOL

- Defined by IETF in 1981
- **Addressing** machines and **routing** over the Internet
- Two versions of the IP protocol on the Internet:
 - IPv4 (very well spread)
 - IPv6 (not that well-spread yet)
- IPv4
 - 4-byte addresses assigned to each computer, e.g., 137.194.2.24.
 - Institutions are given ranges of such addresses, to assign as they will.
 - Problem:
 - only 2^{32} possible addresses
 - a large number of them cannot be assigned to new hosts.
 - many hosts connected to the Internet do not have an IPv4 address (see [IPv4 Address Exhaustion](#))
 - some network address translation (NAT) occurs.
- IPv6:
 - 16-byte addresses;
 - much larger address space! Addresses look like 2001:660:330f:2::18 (meaning 2001:0660:0330f:0002:0000:0000:0000:0018).
 - Other nice features (multicast, autoconfiguration, etc.).

TCP

TRANSMISSION CONTROL PROTOCOL

- Defined also by IETF in 1981
- One of the two main transport protocols used on IP, with UDP (User Datagram Protocol)
- Contrarily to UDP, provides reliable transmission of data (acknowledgments)
- Data is divided into small datagrams (\leq MTU) that are sent over the network, and possibly reordered at the end point
- Port Number
 - Like UDP, each TCP transmission indicates a source and a destination port number (between 0 and 65535) to distinguish it from other traffic
 - A client usually select a random port number for establishing a connection to a fixed port number on a server
 - The port number on a server conventionally identifies an application protocol on top of TCP/IP: 22 for SSH, 25 for SMTP, 110 for POP3. . .

DNS

DOMAIN NAME SYSTEM

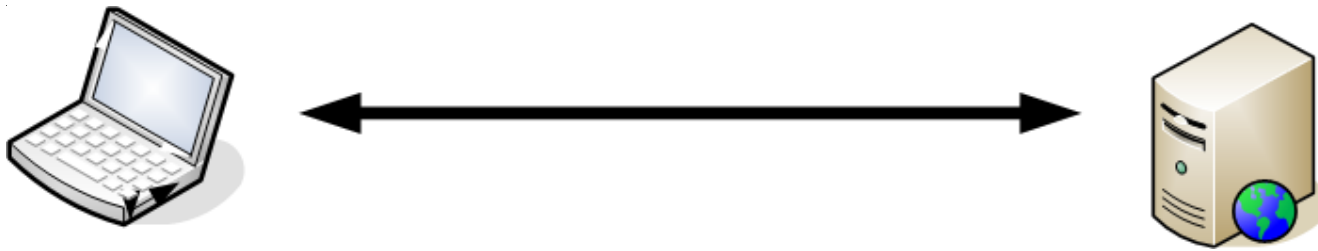
- Defined and modified by IETF
- IPv4 addresses are hard to memorize, and a given service (e.g., a Web site) may change IP addresses (e.g., new Internet service provider)
 - Even more so for IPv6 addresses!
- DNS: a UDP/IP-based protocol for associating human-friendly names (e.g., `www.google.com`, `weather.yahoo.com`) to IP addresses
- Hierarchical domain names:
 - `com` is a top-level domain (TLD), `yahoo.com` is a subdomain thereof, etc.
- Hierarchical domain name resolution:
 - root servers with fixed IPs know who is in charge of TLDs, servers in charge of a domain know who is in charge of a subdomain, etc.
- Nothing magic with `www` in `www.google.com`: just a subdomain of `google.com`.

INTERNET VS. THE WEB

- Internet
 - physical network of computers (or hosts)
- World Wide Web, Web, WWW:
 - logical collection of hyperlinked documents
 - static and dynamic
 - accessible from the Internet
 - each document (or Web page, or resource) identified by a URL
- Intranet
 - Network of computers not accessible from the general internet
 - Private Web Pages, not accessible from the Internet

THE WEB

A CLIENT/SERVER ARCHITECTURE



- A variety of clients are used:
 - graphical browsers
 - textual browsers: w3m, lynx ... used by visually-impaired people when sites are accessible
 - browsers with speech-synthesis engines
 - crawlers, spiders, robots ...
- Servers deliver content to the clients:
 - static content (pages, images, ...)
 - dynamically generated content (php, js, asp, ...)
- Compromise: light-client/heavy-client

WEB SERVERS

Server	Share	Distribution
Apache	53%	Windows, Mac OS, Linux, Unix
Microsoft IIS	17%	with some versions of Windows
nginx	15%	Windows, Mac OS, Linux, Unix
lighttpd	<1%	Windows, Mac OS, Linux, Unix

- Market share:
 - according to various studies, precise numbers do not really mean anything.
- Many large software companies have either their own Web server or their own modified version of Apache (notably, GWS for Google).
- nginx and lighttpd are lighter (i.e., less feature-rich, but faster in some contexts) than Apache.
- The versions of Microsoft IIS released with consumer versions of Windows are very limited.

WEB SEARCH ENGINES

- A large number of different search engines, with market shares varying a lot from country to country.
- At the world level:
 - Google vastly dominating (around 80% of the market; more than 90% market share in France!)
 - Yahoo!+Bing still resists to its main competitor (around 10% of the market)
- In some countries, local search engines dominate the market
 - Baidu with 75% in China,
 - Naver in Korea,
 - Yahoo! Japan in Japan,
 - Yandex in Russia
- Other search engines mostly either use one of these as backend (e.g., Google for AOL) or combine the results of existing search engines
- Many others: DuckDuckGo, Exalead, ...