CSC 402 – Internet Technology

Recap

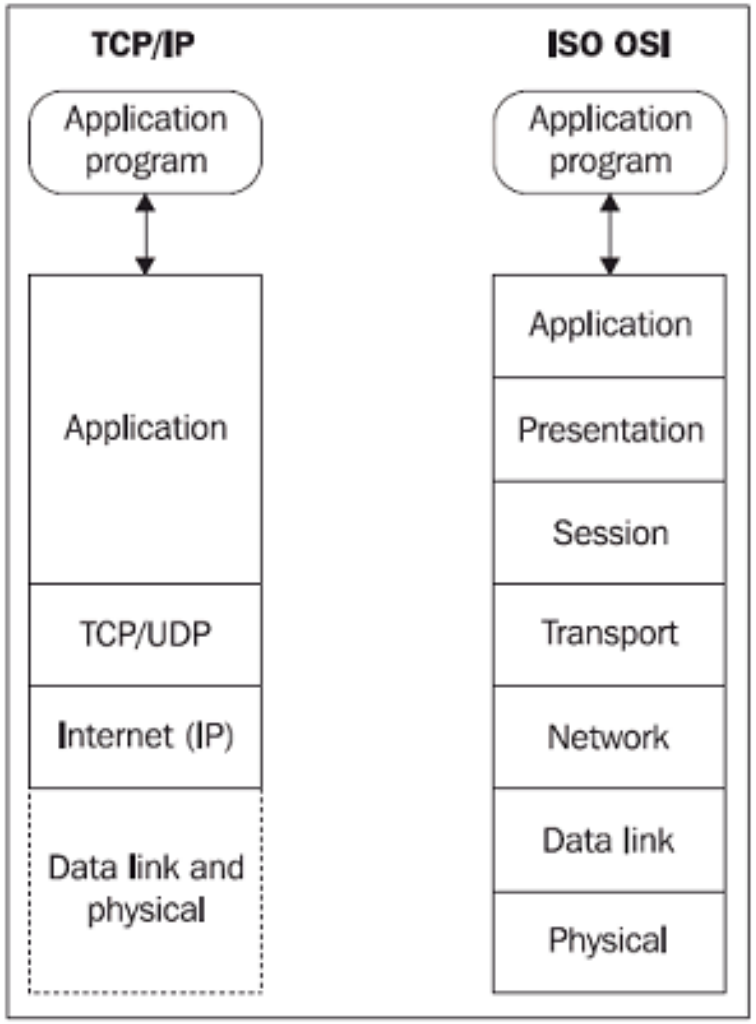
- VLSM

Application Layer Protocol

- DNS
- HTTP
- FTP
- E-mail – SMTP, MIME, POP3, IMAP
- E-mail Security - PGP
- Remote terminal – Telnet, SSH

Application Layer

- Application layer
  - 7th layer according to ISO/OSI stack
  - Interface to the presentation layer
  - 4th layer according to TCP/IP stack
  - The actual Internet Protocol suite
  - Interfaces with transport layer protocols (e.g. TCP or UDP – multiplexing over different port numbers)
  - Covers also the presentation and
  - session layers known from ISO/OSI stack
- Application layer protocols
  - Provide interfaces between application processes and underlying network protocols
  - Data translation and interpretation services



Introduction

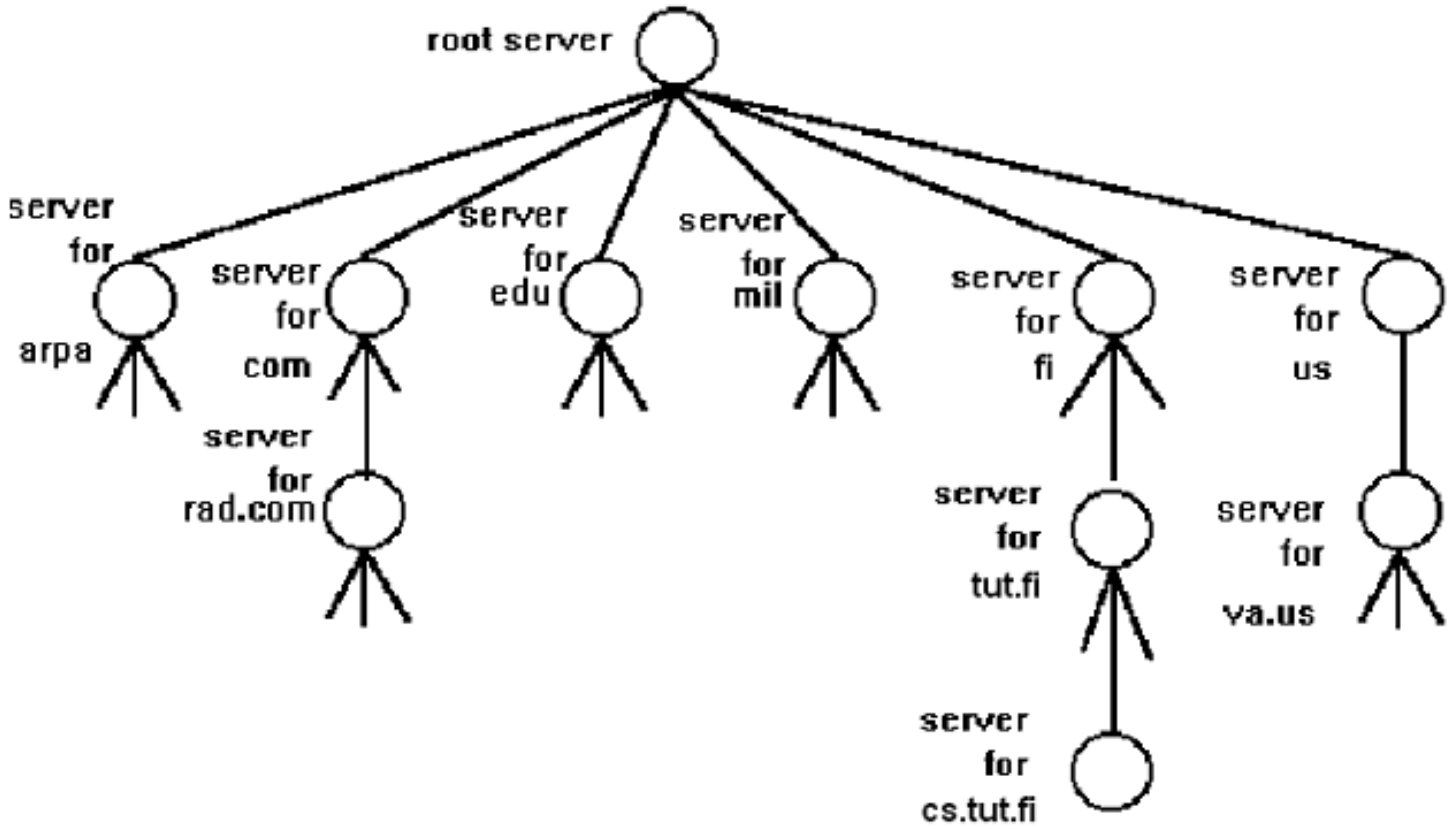
- Communication approaches
  - Client-server concept.
  - ‘Traditional’ approach.
  - 2 separate application implementations.
  - Connection is initiated by the client side.
  - Server is ‘serving’ client’s request.
  - Used in majority of applications.
- Peer-to-peer concept
  - Only one implementation exists.
  - Peers can both initiate and accept connections.
  - Gained popularity due to p2p file sharing services.
- Hybrid solutions
  - E.g. Skype is based on both concepts.

DNS - Domain Name System

- The problem of host identification
  - By hostname - mnemonic, e.g. www.deerwalk.edu.np.
    - People prefer mnemonic names.
  - By IP address, e.g. 130.230.4.103.
  - The best way for machines (hardware operates on binary data, that includes the addresses) sort of mapping tool has to be introduced.
- Domain Name System
  - Distributed database implemented in a hierarchy of name servers, used by TCP/IP applications.
  - An application-level protocol that allows hosts and name servers to communicate in order to provide the translation service.
  - Runs between communicating ends using client-server paradigm.
  - Relies on an underlying transport protocol (TCP or UDP on port 53).
- Advantages
  - Hostnames also allow independence from knowing the physical location of a host.
  - A host may be moved to a different network, while the users continue to use the same (logical) host name.

DNS

- Hierarchical DNS Organization



DNS – Name Space

- Root
- TLD, Top Level Domains - Around 200 domains (Domain – a logical set of computers)
  - Public: .com, .net…
  - Special: .mil, .gov, .edu, .org…
  - Infractructure TLD: .arpa and .root (latter one has never been used)
  - National: .np, .de, .fr, .es, .uk -> always 2 letters long, Maintained by ICANN (Internet Corporation for Assigned Names and Numbers): www.icann.org register almost any name
- After TLD one may register any name
  - Not more than 127 levels after the root.
  - Name length at every level is not more than 63 symbols.
  - Total length of the name should not be longer than 255 symbols.
  - New domain registration.
    - To register aa.bb.np, ISP responsible for bb.np should perform the registration.
- The structure of domains does not represent the physical outline of a network, but the logical one.
  - i.e., the company structure. E.g. np -> dw.np -> csit.dw.np -> class16.csit.edu.np.

DNS – issue

- Whenever a new system is installed in a zone the DNS administrator for the zone allocates a name and an IP address for the new system and enters these into the Name Server's database.
  - A Name Server (NS) is said to have authority for one zone or multiple zones.
- A person responsible for a zone must provide a primary NS for that zone and one or more secondary NSs.
- Difference between a primary and a secondary servers.
  - The primary loads all the information for the zone from disk files, the appropriate info (name and IP address) must be added manually to a configuration file; after each change the NS reloads its files.
- The secondary ones obtain all the information from the primary one through a zone transfer, querying it on a regular basis (e.g. every 3 hours).

DNS – Server types

- Local name server (LNS), typically close to the client.
  - Each ISP (university, company, etc.) has a local name server.
- Root Name Server (RNS)
  - 13 root servers in the world and every primary name server has to know the address of one of root server (next slide).
  - http://www.root-servers.org, http://www.open-rsc.org, http://as112.net.
  - Names in the “form letter.root-servers.net”, letter ranging from A to M.
  - When a local name server cannot immediately satisfy a query from a host, the LNS behaves as a DNS client (resolver) and queries one of the RNS.
  - If the RNS has a record for the hostname, it sends reply message to LNS.
  - Then LNS sends a DNS reply to the querying host.
- Authoritative Name Server (ANS)
  - Every host is registered with its ANS.
  - Typically, the ANS for host is a name server in the host’s local ISP.
  - Each host is required to have at least two ANSs, in case of failures.
  - By definition: A name server is authoritative for host if it always has a DNS record that translates the host’s hostname to that host’s IP address.
  - When ANS is queried by a root server, the ANS responds with a DNS reply that contains the requested mapping.

DNS - Services

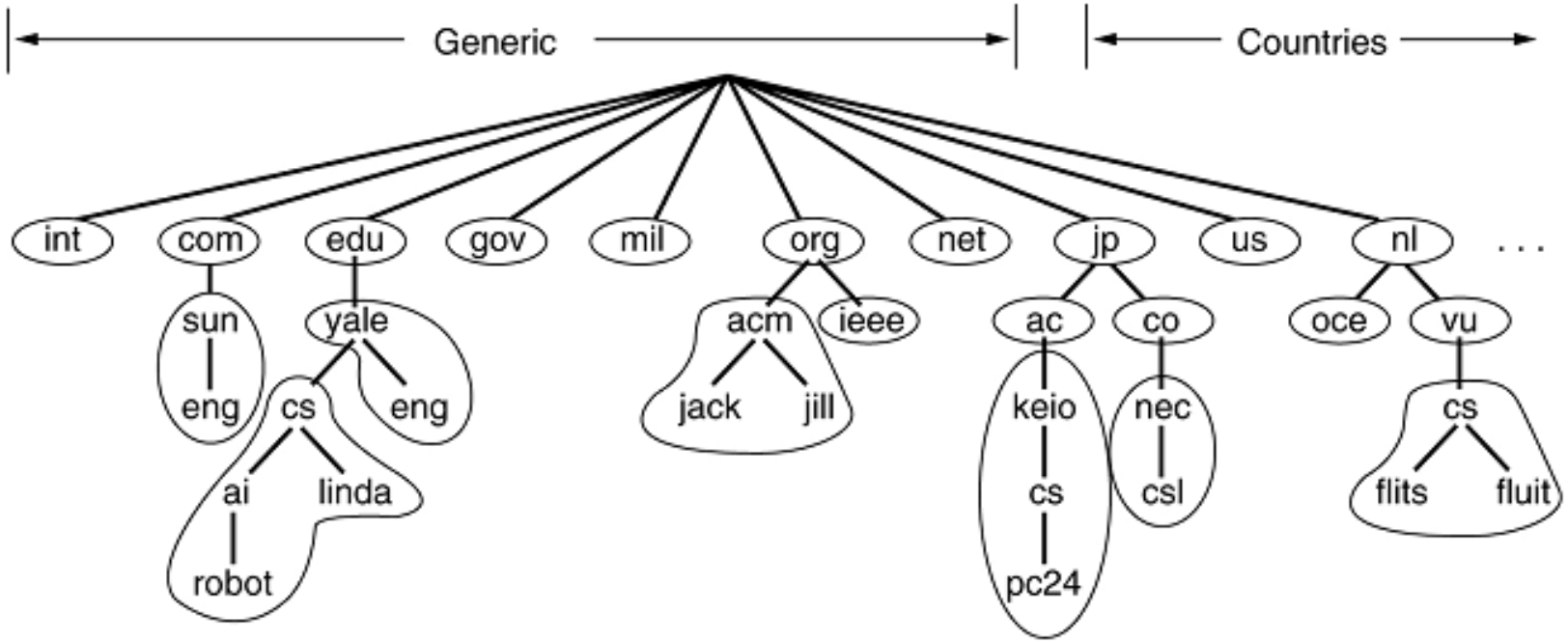
- Host aliasing
  - A host with complicated hostname can have one or more alias names.
  - Hostname such as relay1.west-coast.enterprise.com could have two aliases such as enterprise.com and www.enterprise.com.
  - Alias hostnames are typically more mnemonic than a canonical hostname.
- Mail Server aliasing
  - It's desirable that e-mail address be mnemonic.
  - E.g. e-mail address user@hotmail.com, hotmail.com – mnemonic address.
  - The canonical hostname of the Hotmail mail server could be more complicated and much less mnemonic than simply hotmail.com.
  - E.g. col0-omc3-s2.col0.hotmail.com.
- Load distribution
  - DNS is also being used to perform load distribution among replicated servers (such as replicated Web servers).
  - Busy sites, such as cnn.com, are replicated over multiple servers.
  - The DNS database contains this set of IP addresses.

DNS

- DNS Caching
  - Fundamental property of the DNS
  - When a name server receives information about particular destination, it caches the mapping information (name and IP address).
  - A later query for the same destination can use the cached result, and not result in additional queries to other servers.
  - DNS extensively exploits caching in order to improve the delay performance and to reduce the number of DNS messages in the network.
- DNS Request
  - Relatively complicated procedure.
  - Caching decreases the complexity and load.
  - Scenario: Client sends a request to local DNS server.
    - The local DNS server checks whether it is an authoritative server for the domain to whom the client belongs, and checks availability of the DNS information.
    - If the information does not exist: It sends a DNS request towards root DNS server of the corresponding domain.
    - If root DNS server does not have the requested information the request is forwarded to lower-layer authoritative (or intermediate) DNS server(s).

DNS – Request


Originator
VU CS name server
Edu name server
Yale name server
Yale CS name server
flits.cs.vu.nl
cs.vu.nl
edu-server.net
yale.edu
cs.yale.edu
1
2
3
4
5
6
7
8


Generic
Countries
int
com
edu
gov
mil
org
net
jp
us
nl
. . .
sun
eng
yale
cs
eng
ai
linda
robot
acm
ieee
jack
jill
ac
co
keio
cs
pc24
nec
csl
oce
vu
cs
flits
fluit

HTTP

- Hypertext Transfer Protocol (HTTP)
- The Web's application-layer protocol.
  - http://www.google.com/ - indicates that HTTP is to be used
- Implemented in two programs.
  - A client program.
  - A server program.
  - Both of programs are executing on different end systems, talk to each other by exchanging HTTP messages.
- HTTP defines
  - The structure of these messages.
  - How the client and server exchange the messages.

HTTP

- Web Objects: A Web page (called a document) consists of objects (files)
  - HTML file, JPEG/GIF image, Java applet, audio clip, etc. – each addressable by a single URL.
  - Most web pages consist of a base HTML file and several referenced objects.
    - A Web page containing HTML text and five referenced JPEG images consists of six objects.
  - The base HTML file references the other objects in the page with the objects' URLs.
  - Each URL has two components: the host name of the server that houses the object and the object's path name.
  - For example, the URL www.some_school.fi/some_department/picture.gif has:
    - www.some_school.fi for a host name.
    - /some_department/picture.gif for a path name.
- Web Browser: User Agent for the web
  - Implemented on Client Side.
    - E.g. Microsoft Internet Explorer, Mozilla Firefox, Opera, Google Chrome etc.
- Web Server: Houses Web objects, each addressable by a URL
  - Implemented on Server side.
    - Apache, Microsoft Internet Information Server to name a few.
- HTTP defines how
  - Web clients (browsers) request Web pages from servers (Web servers).
  - Servers transfer Web pages to clients.
- The general idea of interaction
  - When a user requests a Web page (for example, clicks on a hyperlink), the browser sends HTTP request messages for the objects in the page to the server.
  - The server receives the requests and responds with HTTP response messages that contain the objects.

HTTP – How it works

- The HTTP client first initiates a TCP connection with the server
  - TCP connection is established.
  - Browser and the server processes access TCP through their socket interfaces.
  - The client sends HTTP request messages and receives HTTP response messages.
    - Communications through the socket interfaces.
  - Similarly, the HTTP server receives request messages and sends response messages.
  - Once the client sends a message to its socket interface, the message is “out of the client's hands” and is “in the hands of TCP”.
- Each HTTP request message emitted by a client process eventually arrives intact at the server.
- Each HTTP response message emitted by the server process eventually arrives intact at the client.
- The great advantages of a layered architecture.
  - HTTP doesn’t have to worry about lost data, or the details of how TCP recovers from loss or reordering of data within the network.
  - That is the job of TCP and the protocols in the lower layers of the protocol stack.
- The server sends requested files to clients without storing any state information about the client.
  - If a client asks for the same object twice, the server does not respond by saying that it just served the object to the client.
    - instead, the server resends the object as it has completely forgotten what it did before.
  - Because an HTTP server maintains no information about the clients, HTTP is said to be a **stateless protocol**
- It can use both Persistent (HTTP/1.1) and Non-persistent connection (HTTP/1.0).

HTTP

- HTTP server is stateless
  - It simplifies server design.
  - Permit to develop very high-performing Web servers.
- It is often desirable for a Web site to identify users.
  - either because the server wishes to restrict user access or
  - because it wants to serve content as a function of the user identity
- HTTP provides two mechanisms to help a server identify a user.
  - Authentication.
  - Cookies.

HTTP: Authentication

- Web sites may require user to provide a username and a password in order to access the content
  - This requirement is referred to as authentication.
  - HTTP provides authentication.
- Suppose a client requests an object from a server, and the server requires user authorization
  - The client first sends an ordinary request message.
  - The server responds with empty entity body and with a 401 Authorization Required status code.
  - In this response message the server includes the WWW-Authenticate: header, which specifies the details about how to perform authentication.
    - Typically, it indicates that the user needs to provide a username and a password.
  - The client receives the response message and prompts the user for a username and password
  - The client resends the request message, but this time includes an Authorization: header line, which includes the username and password.
  - After obtaining the first object, the client continues to send the username and password in subsequent requests for objects on the server.
    - This typically continues until the client closes the browser.
    - While the browser remains open, the username and password are cached, so the user is not prompted for a username and password for each object it requests.

HTTP – Cookies

- Cookies are an alternative mechanism that sites can use to keep track of users
  - Defined in RFC 2109.
- E.g. suppose a client contacts a Web site for the first time, and this site uses cookies.
  - The server's response will include a “Set-cookie:” header.
  - Often this header line contains an identification number generated by the Web server. For example, the header line might be: “Set-cookie: 1678453”.
  - When the HTTP client receives the response message, it checks the “Setcookie:” header and identification number.
  - It appends a line to a special cookie file that is stored in the client machine.
    - This line includes the host name of the server and user's associated identification number.
  - In subsequent requests to the same server the client includes a “Cookie:” request header.
    - This header line specifies the identification number for that server.
    - In the current example, the request message includes the header line “Cookie: 1678453”.

HTTP – Cookies

- In such manner the server.
  - Does not know the username of the user.
  - Does know that this user is the same user that made a specific request before.
- Web servers use cookies for many different purposes.
  - If a server requires authentication but doesn't want to hassle a user with a username and password prompt every time the user visits the site.
  - If a server wants to remember a user's preferences so that it can provide targeted advertising during subsequent visits.
  - If a user is shopping at a site (for example, buying several CDs), the server can use cookies to keep track of the items that the user is purchasing, that is, to create a virtual shopping cart.
- Drawbacks
  - Cookies won’t work for a nomadic user who accesses the same site from different machines.
    - The site will treat the user as a different user for each different machine used.
  - Privacy issues.
  - E.g. banner ads cookies.

FTP

- FTP (File Transfer Protocol) is a protocol for transferring a file from one host to another host.
  - Dates back to 1971.
  - Described in RFC 959.
- In a typical FTP session.
  - User wants to transfer files to or from a remote host.
    - A user identification and a password is required.
- After providing this authorization information, the user can transfer files from the local file system to the remote file system and vice versa.

FTP

- User interacts with FTP through an FTP user agent.
- The user provides the hostname (IP address or domain name) of the remote host.
- The FTP client application establishes a TCP connection with the FTP server in the remote host.
- The user then provides the username (login) and password, which get sent over the TCP connection as part of FTP commands.
- Once the server has authorized the user, the user can perform operations on the files stored in the remote file system.
  - List folders/files, upload, download, create folders etc.

FTP

- Example FTP session: Client first sets up a control TCP connection on server port number 21.
- The user ID and password are sent over this connection.
- Sending commands to change the remote directory.
- Request or upload a file.
- Client opens a TCP data connection on server port 20.
- FTP sends exactly one file over the data connection and then closes the data connection.
- During the same session a new data connection is opened for each sent file (non-persistent).
- The control connection remains open (persistent connection).

FTP

- Throughout a session, the FTP server must maintain state about the user.
  - The server must associate the control connection with a specific user account.
  - The server must keep track of the user's current directory as the user is browsing the remote directory tree.
    - Keeping track of the state information for each ongoing user session significantly constrains the total number of sessions that FTP server can handle simultaneously.
- Addressing – sample FTP URLs.
  - ftp://user:password@host:port/path.
  - ftp://anonymous:my@email.edu.np@ftp.funet.fi:21/pub/standards/RFC/rfc2166.txt
  - ftp://ftp.funet.fi/pub/standards/RFC/rfc2166.txt

FTP Vs. HTTP

- Both file transfer protocols.
- FTP uses two parallel TCP connections to transfer a file, a control connection and a data connection.
  - The control connection is used for sending control information between the hosts – such as user identification, password, commands to change remote directory, and commands to "put" and "get" files.
  - The data connection is used to actually send a file.
- Because FTP uses a separate control connection, FTP is said to send its control information out-of-band.