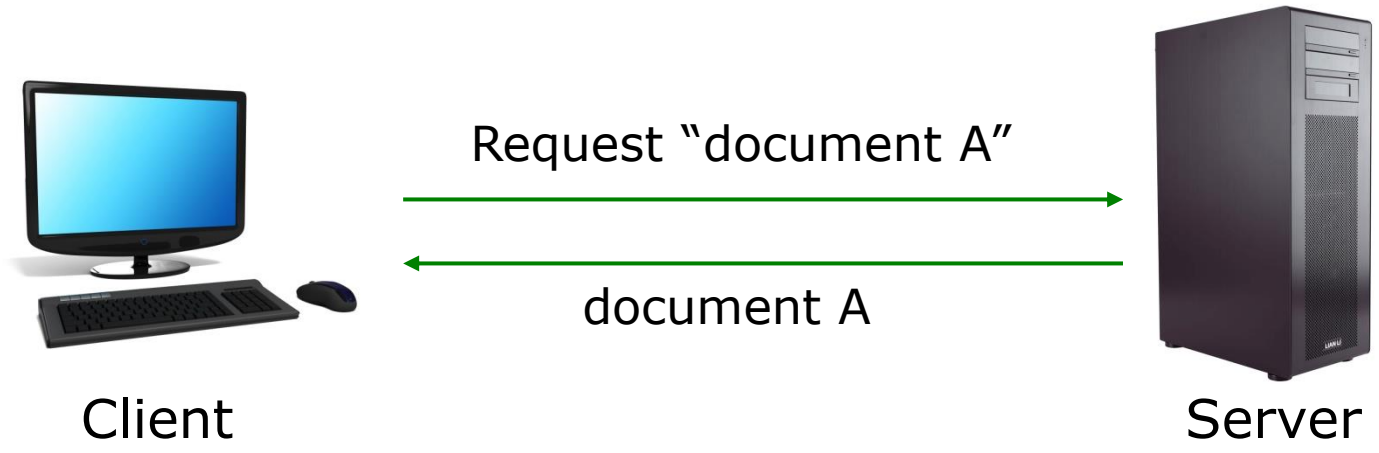
Web Essentials:
Clients, Servers, and Communication

PV219, spring 2024

Web Essentials

- **Client**: web browsers, used to surf the Web
- **Server** systems: used to supply information to these browsers
- Computer **networks**: used to support the browser-server communication



Internet vs. Web

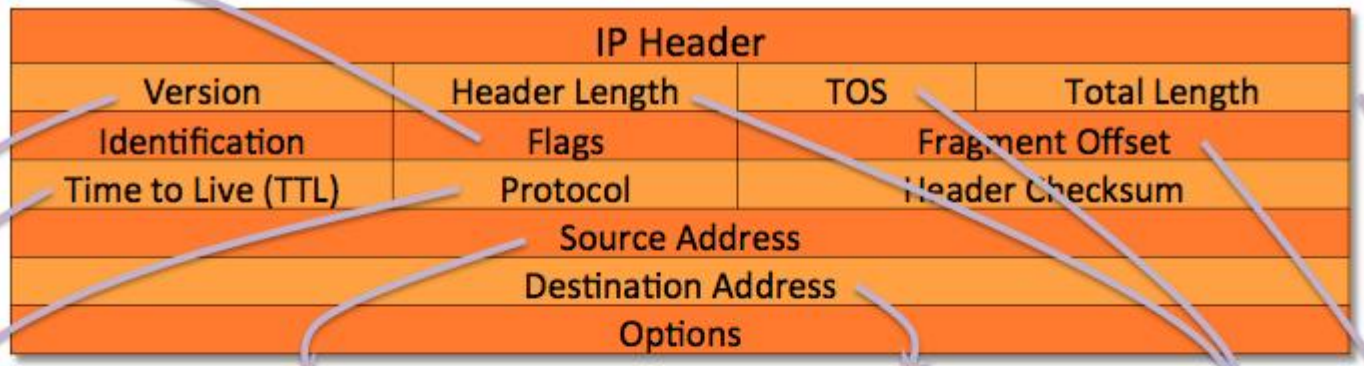
- **The Internet**: a inter-connected computer networks, linked by wires, cables, wireless connections, etc.
- **Web**: a collection of interconnected documents and other resources.
- The world wide web (**WWW**) is accessible via the Internet, as are many other services including email, file sharing, etc.

How does the Internet Work?

- Through communication protocols
- A **communication protocol** is a specification of how communication between two computers will be carried out
  - **IP** (Internet Protocol): defines the packets that carry blocks of data from one node to another
  - **TCP** (Transmission Control Protocol) and **UDP** (User Datagram Protocol): the protocols by which one host sends data to another.
  - Other application protocols: **DNS** (Domain Name Service), **SMTP** (Simple Mail Transmission Protocol), and **FTP** (File Transmission Protocol)

The Internet Protocol (IP)

- A key element of IP is **IP address**, a 32/64-bit number
- The Internet authorities assign ranges of numbers to different organizations
- IP is responsible for moving **packet** of data from node to node
- A packet contains information such as the data to be transferred, the source and destination IP addresses, etc.
- Packets are sent through different local network through **gateways**
- A **checksum** is created to ensure the correctness of the data; corrupted packets are discarded
- IP-based communication is **unreliable**

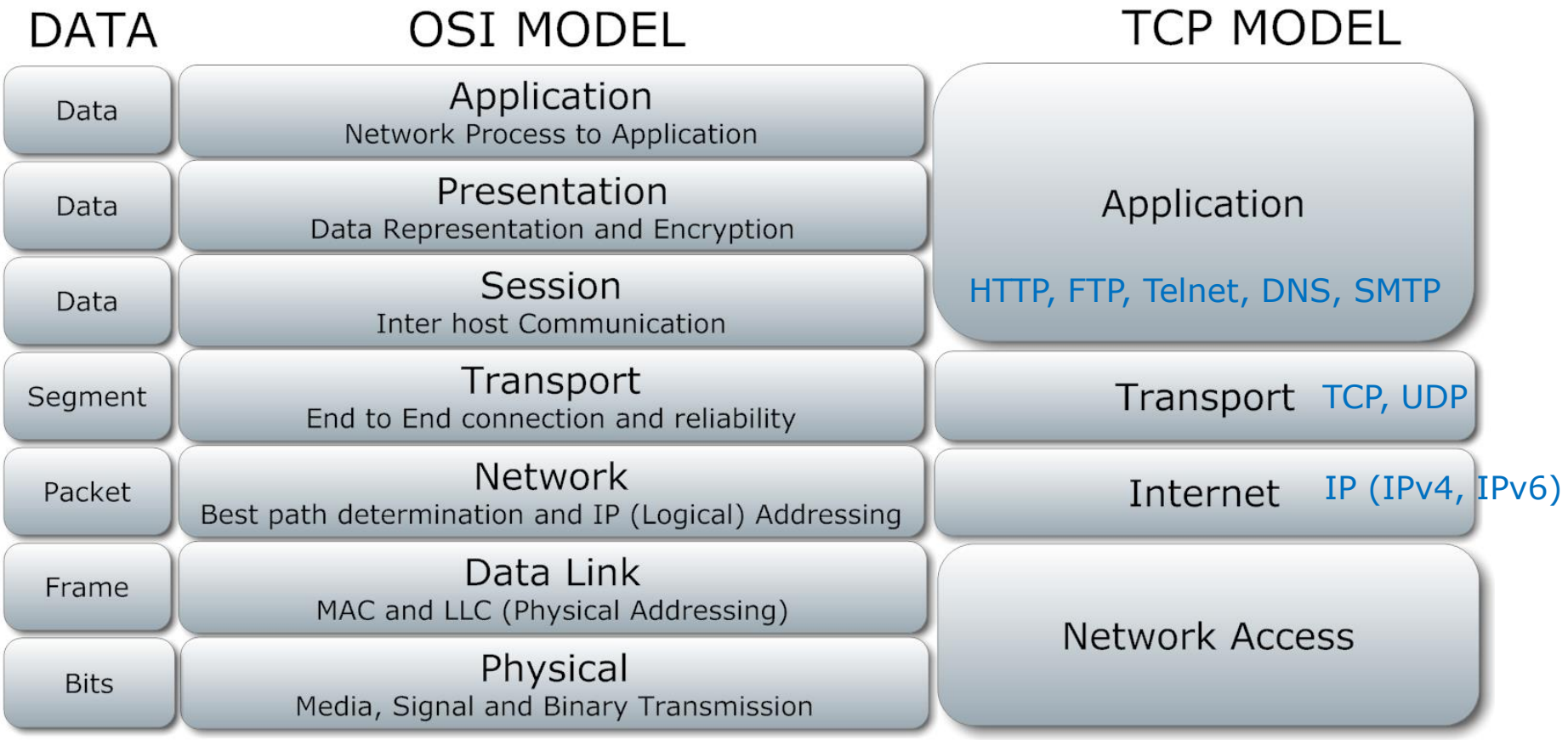


Internet Protocol Version 4, Src: 10.100.16.200 (10.100.16.200), Dst: 10.100.185.66 (10.100.185.66)
Version: 4
Header length: 20 bytes
Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
0000 00.. = Differentiated Services Codepoint: Default (0x00)
.... ..00 = Explicit Congestion Notification: Not-ECT (Not ECN-Capable Transport) (0x00)
Total Length: 1420
Identification: 0x126d (4717)
Flags: 0x02 (Don't Fragment)
0... .... = Reserved bit: Not set
.1.. .... = Don't fragment: Set
..0. .... = More fragments: Not set
Fragment offset: 0
Time to live: 255
Protocol: TCP (6)
Header checksum: 0x98ad [correct]
[Good: True]
[Bad: False]
Source: 10.100.16.200 (10.100.16.200)
Destination: 10.100.185.66 (10.100.185.66)

Transmission Control Protocol (TCP)

- TCP is a higher-level protocol that extends IP to provide additional functionality: **reliable** communication
- TCP adds support to detect errors or lost data and to trigger **retransmission** until the data is correctly and completely received
- Connection
- Acknowledgment

TCP/IP Protocol Suites



The World Wide Web

- **WWW** is a system of interlinked, hypertext documents that runs over the Internet
- Two types of software:
  - **Client**: a system that wishes to access the information provided by servers must run client software (e.g., web browser)
  - **Server**: an internet-connected computer that wishes to provide information to others must run server software
  - Client and server applications communicate over the Internet by following a protocol built on top of TCP/IP – **HyperText Transport Protocol** (**HTTP**)

WWW History

- **1989** - **Birth of WWW**
  - Tim Berners-Lee & his associates at CERN
- **1990** - **First Web Browser**
  - Used within CERN
- **1991** - **Public offering of WWW**
- **1993** - **Birth of Mosaic**
  - Graphical, multimedia browser from NCSA
- **1994** - **First commercial browser**
  - By Netscape communications founded by Jim Clark and Marc Andreessen

Basics of the WWW

- **Hypertext**: a format of information which allows one to move from one part of a document to another or from one document to another through **hyperlinks**
- Uniform Resource Locator (**URL**): unique identifiers used to locate a particular resource on the network
- **Markup language**: defines the structure and content of hypertext documents

Web Client: Browser

Makes HTTP requests on behalf of the user

- Reformat the URL entered as a valid HTTP request
- Use DNS to convert server's host name to appropriate IP address
- Establish a TCP connection using the IP address
- Send HTTP request over the connection and wait for server's response
- Display the document contained in the response
  - If the document is not a plain-text document but instead is written in HTML, this involves rendering the document

Web Servers

Main functionalities:

- Server waits for connect requests
- When a connection request is received, the server creates a new process to handle this connection
- The new process establishes the TCP connection and waits for HTTP requests (**stateless!**) – HTTP2+ is a solution
- The new process invokes software that maps the requested URL to a resource on the server
- If the resource is a file, creates an HTTP response that contains the file in the body of the response message
- If the resource is a program, runs the program, and returns the output

Static Web: HTML/XHTML, CSS

- **HTML** stands for **H**yper**T**ext **M**arkup **L**anguage
  - It is a text file containing small markup tags (elements) that tell the Web browser how to display the page
- **XHTML** stands for e**X**tensible **H**yper**T**ext **M**arkup **L**anguage
  - It is identical to HTML 4.01
  - It is a stricter and cleaner version of HTML
- **CSS** stands for **C**ascading **S**tyle **S**heets
  - It defines how to display HTML elements

Client-Side Programmability

- Scripting language: a lightweight programming language
- Browser scripting: **JavaScript** (by Netscape in 1995)
  - Designed to add interactivity to HTML pages
  - Usually embedded into HTML pages
  - What can a JavaScript do?
    - Put dynamic text into an HTML page
    - React to events
    - Read and write HTML elements (generally, DOM manipulation)
    - Validate data before it is submitted to a server
    - Asynchronously communicate with server
    - Create cookies
    - …

Server-Side Programmability

- The requests cause the response to be generated
- Server scripting:
  - **CGI/Perl**: Common Gate Way Interface (*.pl, *.cgi)
  - **PHP**: Open source, strong database support (*.php)
  - **ASP**: Microsoft product, uses .Net framework (*.asp)
  - **Java** via JavaServer Pages (*.jsp)
  - **JavaScript** via node.js (*.js)
  - …