Ultimate Extensible Distributed System

Jim Fawcett

Software Modeling

Your Assignment

- Your supervisor just handed you a spec for implementation of a distributed system with universal connectability using sockets that:
  - Can process an open-ended variety of documents
  - Is expandable by five orders of magnitude in 10 years
  - Can add new tools easily
  - Supports 50 million users a day without gridlock
- You say NO WAY!
- Well, maybe.

Introduction to Internet and Web

- This presentation addresses two questions:
  - Is that possible?
    - Well yes—look over there—the web!
  - How was it accomplished?
    - Processing structure and protocols
    - Programming tools
      - Web servers and browsers that host:
        - Script languages, e.g., Javascript, VBScript, Perl, Ruby, …
        - Programming languages:
          Visual Basic, Java, C++, C#, …
    - And, of course, some very smart people

Goals

- Build distributed system to share documents
- Support expansion by five orders of magnitude in 10 years—200 hosts to 500 million hosts
- Manage communication between hundreds of millions of machines every day without collapsing from congestion
- Provide for arbitrary extensions:
  - From static text documents to graphics, dynamic content, streaming video, programmable interfaces, voice, …

Original Goals of the Web

- Universal readership
  - When content is available, it should be accessible from any type of computer, anywhere.
- Interconnecting all things
  - Hypertext links everywhere
  - Simple authoring

Internet Design Principles

- Goal is connectivity
- Achieved with Internet Protocol (IP)
  - Stateless so survives failures—no need to backup
- Made scalable with end-to-end intelligence
  - Transport Control Protocol (TCP)
    - Sender does not send until receipt is acknowledged.
    - Amount sent is based on receiver's current available buffer size, so receiver won't be flooded.
  - Be strict when sending and tolerant when receiving
- Protocol Specific Packet Headers
- Internet Design
- Robustness and the Internet

Web Design Principles

- Universal
- Decentralized
- Modular
- Extensible
- Scalable
- Accessible
- Forward/backwards compatibility
- Architecture of World Wide Web

Basic Concepts

- Client/server model
- Universal addressing
  - TCP/ IP, DNS
- Search engines
- Universal protocols
  - HTTP, URLs, HTML, FTP
- Format negotiation through HTTP
- Hypertext → Hypermedia via HTML → XHTML
  - Support for text, images, sound, and scripting

Internet and Web History

Internet History

- 1961 – First paper on packet-switching theory, Kleinrock, MIT
- 1969 – ARPANet goes online
  - Four hosts, each connected to at least two others
- 1974 – TCP/IP, Berkeley Sockets invented
- 1983 – TCP/IP becomes only official protocol
- 1983 – Name server developed at University of Wisconsin
- 1984 – Work begins on NSFNET
- 1990 – ARPANET shutdown and dismantled
- 1990 – ANSNET takes over NSFNET
  - Nonprofit organization—MERIT, MCI, IBM
  - Starts commercialization of the Internet
- 1995 – NSFNET backbone retired
- 1998 – DNS transferred from Dept of Commerce to ICANN
- 2000 – Web size estimates surpass 1 billion indexable pages

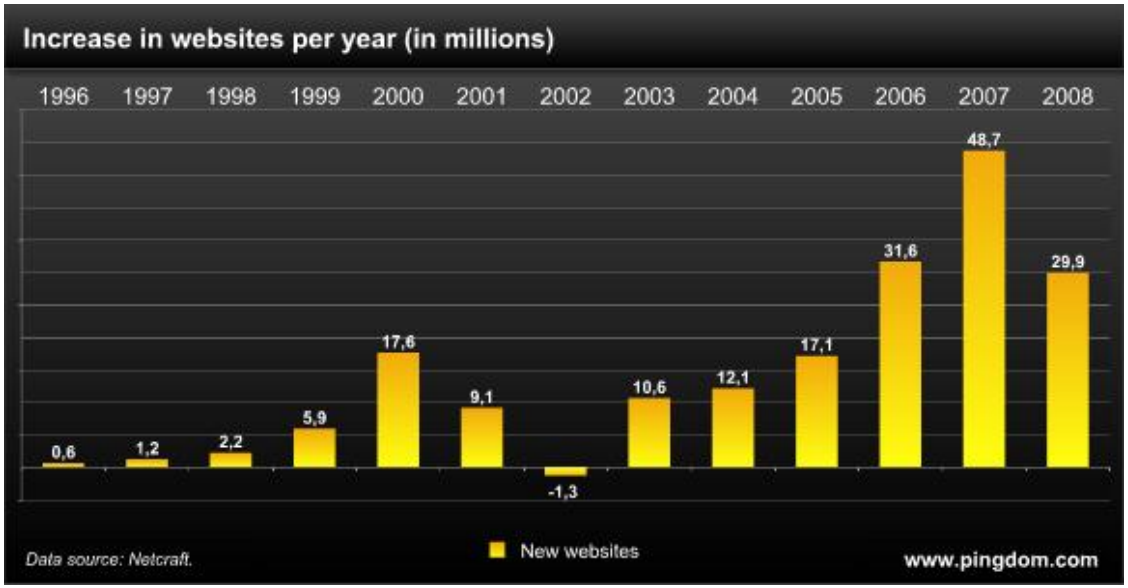
Web History

- 1990 – World Wide Web project
  - Tim Berners-Lee starts project at CERN
  - Demonstrates browser/editor accessing hypertext files
  - HTTP 0.9 defined, supports only hypertext, linked to port 80
- 1991 – first web server outside Europe
  - CERN releases WWW, installed at SLAC
- 1992 – HTTP 1.0, supports images, scripts as well
- 1993 – Growth phase
- 1994 – CERN and MIT agree to set up WWW Consortium
- 1999 – HTTP 1.1, supports open-ended extensions

Web Growth Phase: 1993

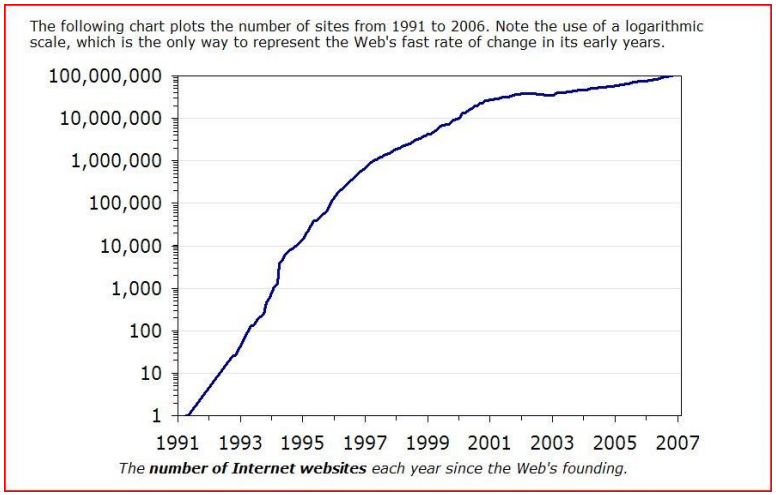
- InterNIC created to provide registration services
- WWW (port 80 HTTP) traffic is 1% of NSFNET traffic
- 200 known HTTP servers
- Article on WWW in *New York Times*
- Mosaic first release

Web Growth



http://www.techcrunch.com/2009/05/08/is-the-growth-of-the-web-slowing-down-or-just-taking-a-breather/

http://www.useit.com/alertbox/web-growth.html

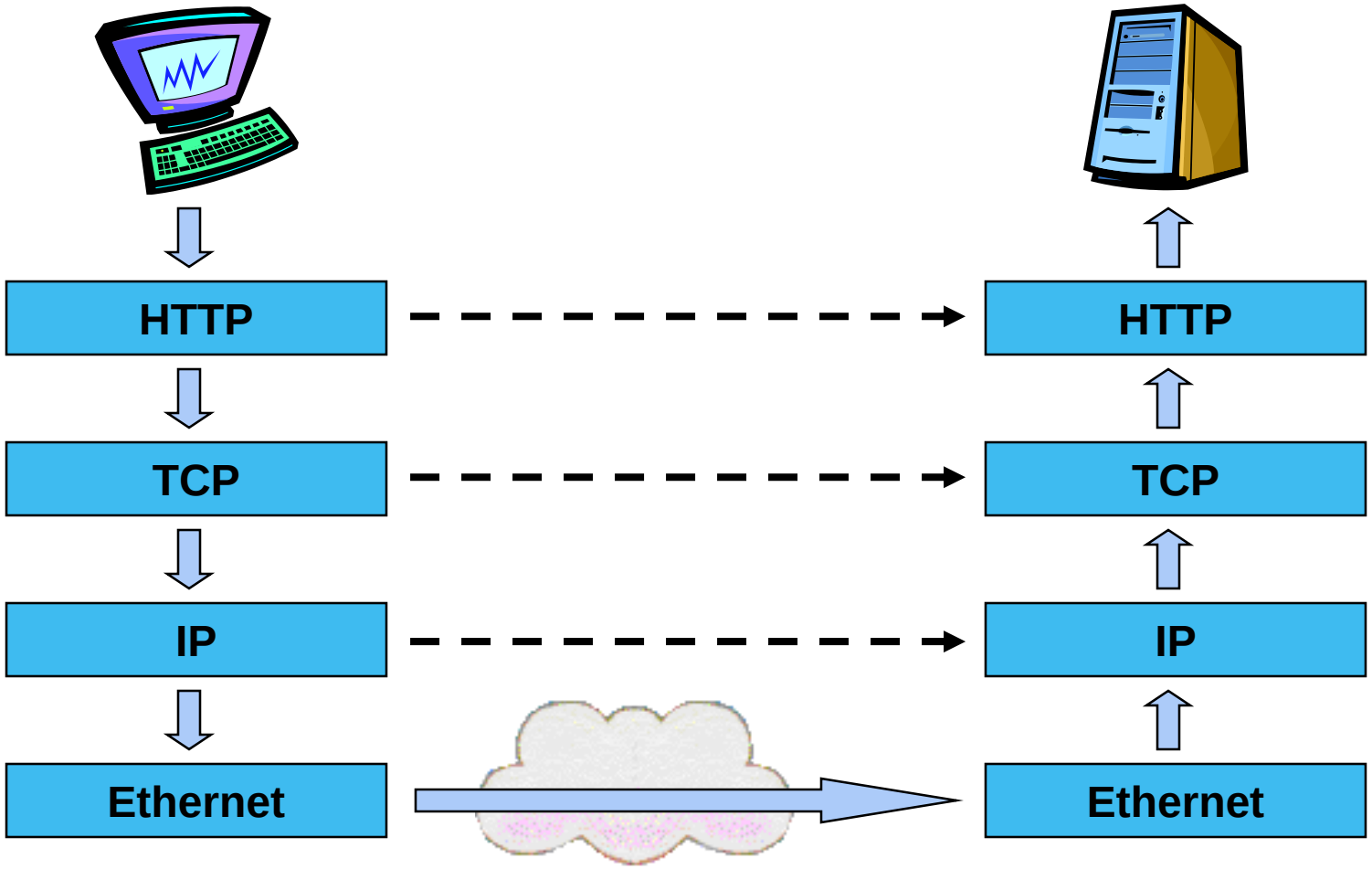


Web Technologies

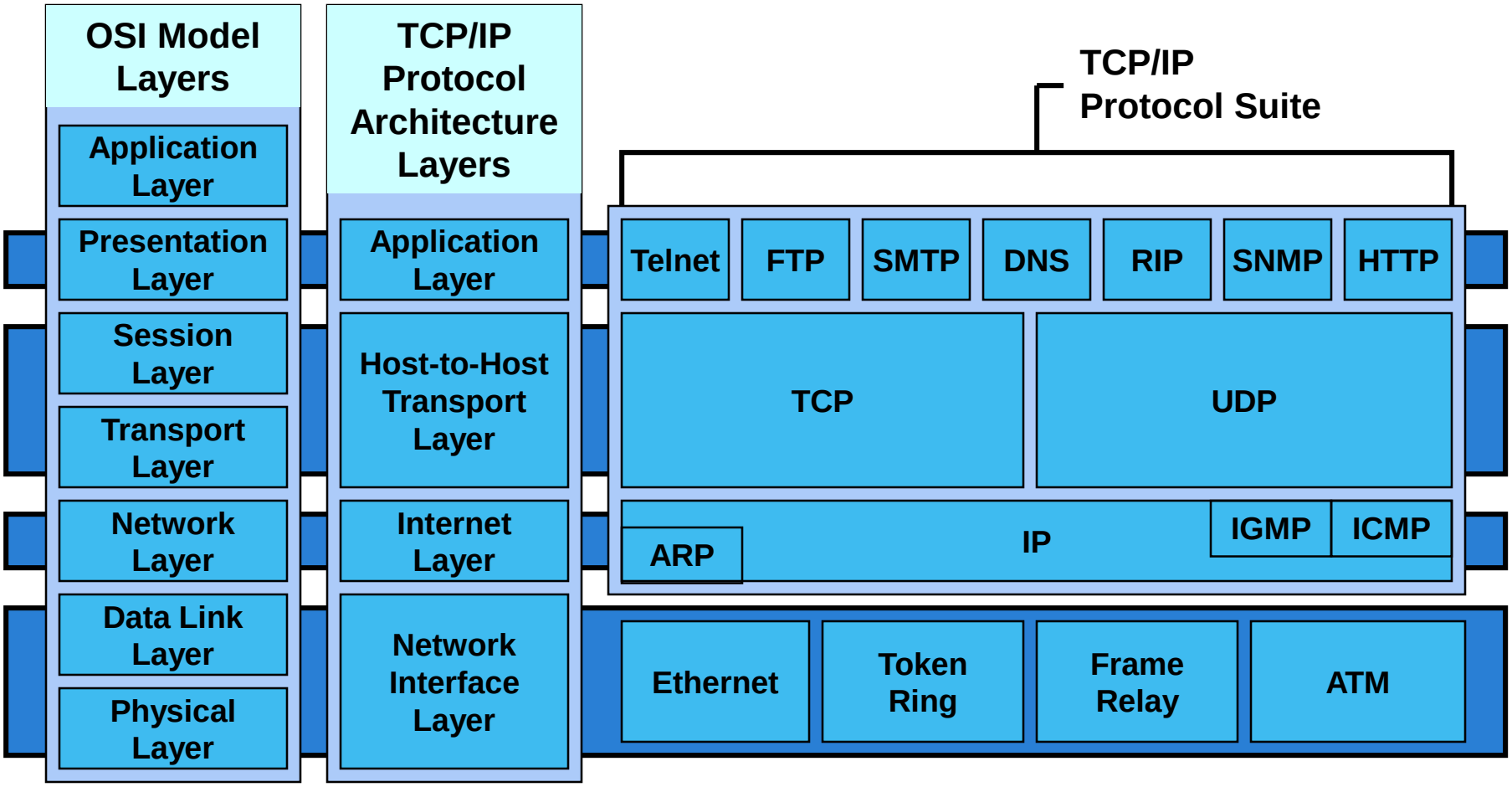
Tools: Servers on the Internet

- **HTTP:** Hypertext Transport Protocol
  - JSP and ASP add dynamic content
  - Web services add RPC program interface
- **FTP:** File Transport Protocol
- **Gopher:** Text and menus
- **NNTP:** Network News Transfer Protocol
- **DNS:** Distributed Name Service
- **Telnet:** Log into a remote computer
- **New tools:** If they use TCP/IP, just add them

Network Protocol Stack



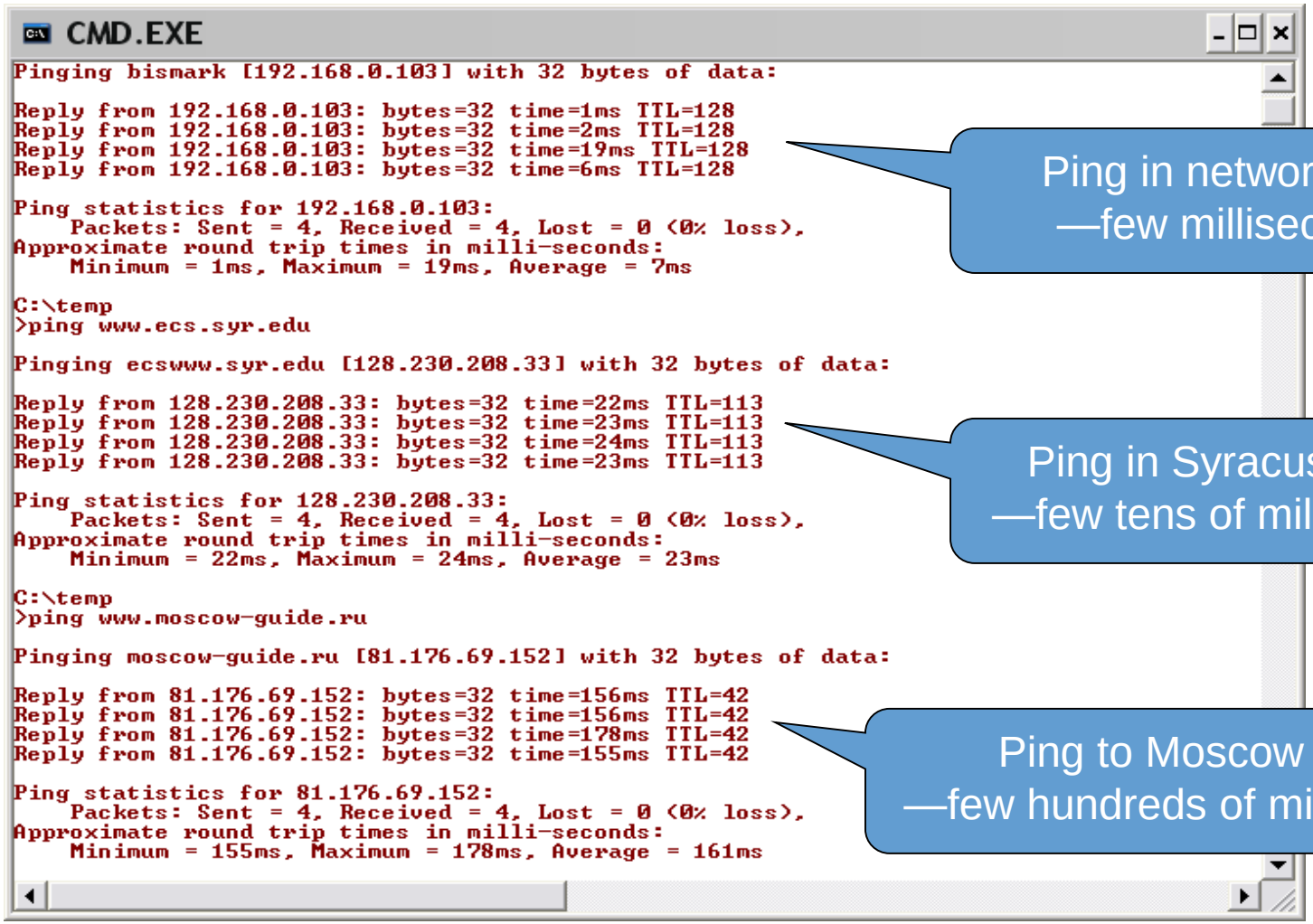
Network Protocols



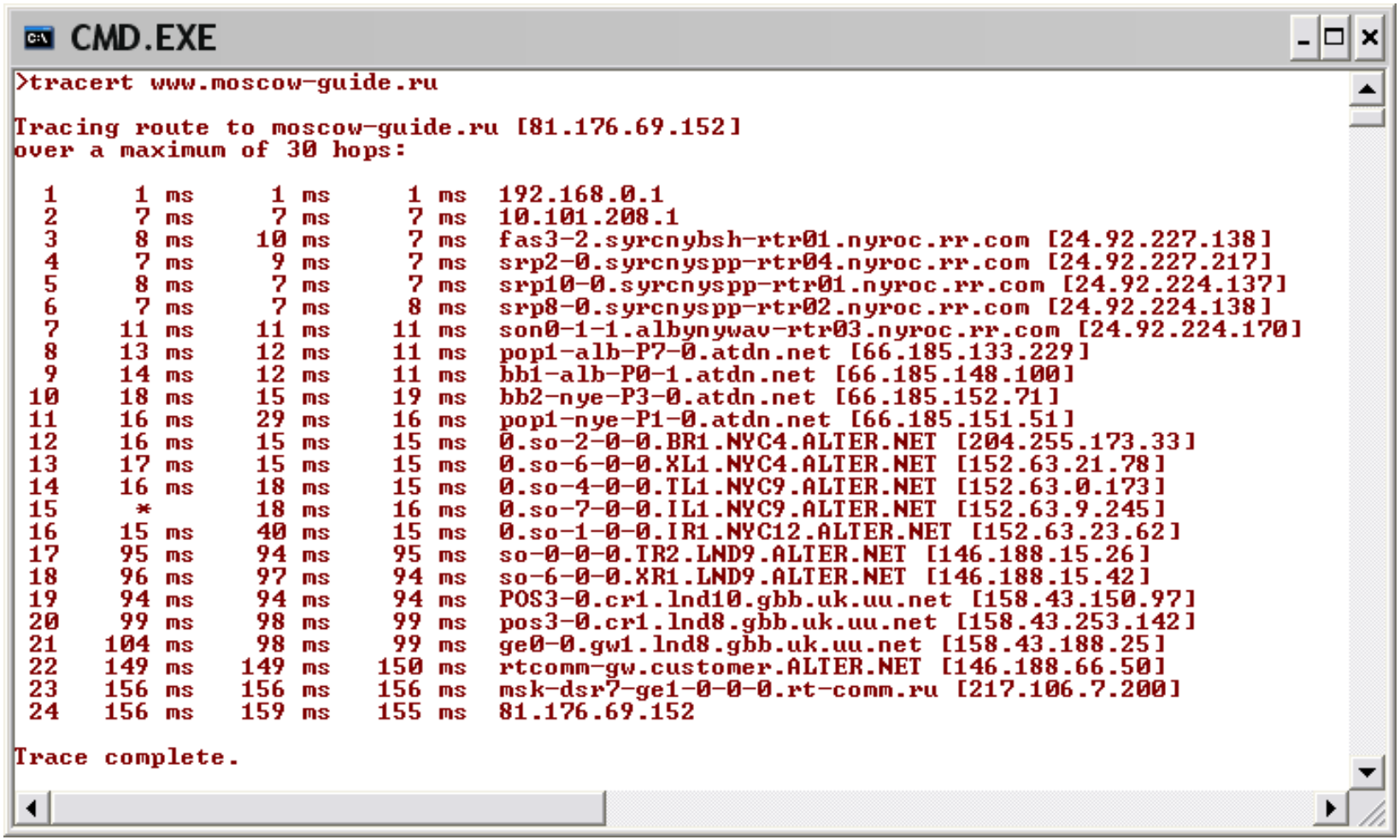
Communication between Networks

- Internet Protocol (IP)
  - Routable, connectionless datagram delivery
  - Specifies source and destination
  - Does not guarantee reliable delivery
  - Large message may be broken into many datagrams, not guaranteed to arrive in the order sent
- Transport Control Protocol (TCP)
  - Reliable stream transport service
  - Datagrams are delivered to the receiving application in the order sent
  - Error control is provided to improve reliability

Pinging Various URLs



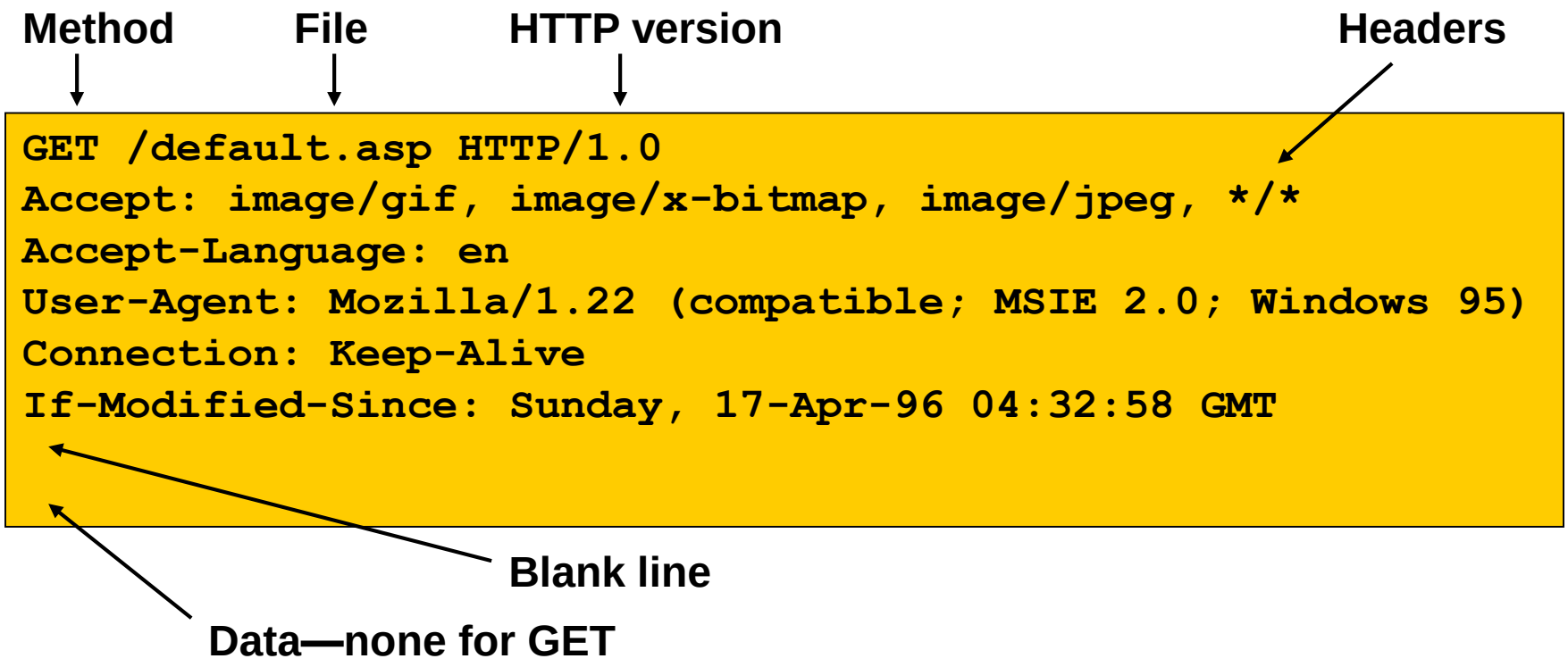
Tracing HTTP Message with Tracert



HTTP Methods

- GET request-URI HTTP/1.1
  - Retrieves entity specified in request-URI as body of response message
- POST request-URI HTTP/1.1
  - Sends data in message body to the entity specified in request-URI
- PUT request-URI HTTP/1.1
  - Sends entity in message body to become newly created entity specified by request-URI
- HEAD request-URI HTTP/1.1
  - Same as GET except the server does not send specified entity in response message
- DELETE request-URI HTTP/1.1
  - Request to delete entity specified in request-URI.
- TRACE request-URI HTTP/1.1
  - Request for each host node to report back

HTTP Request



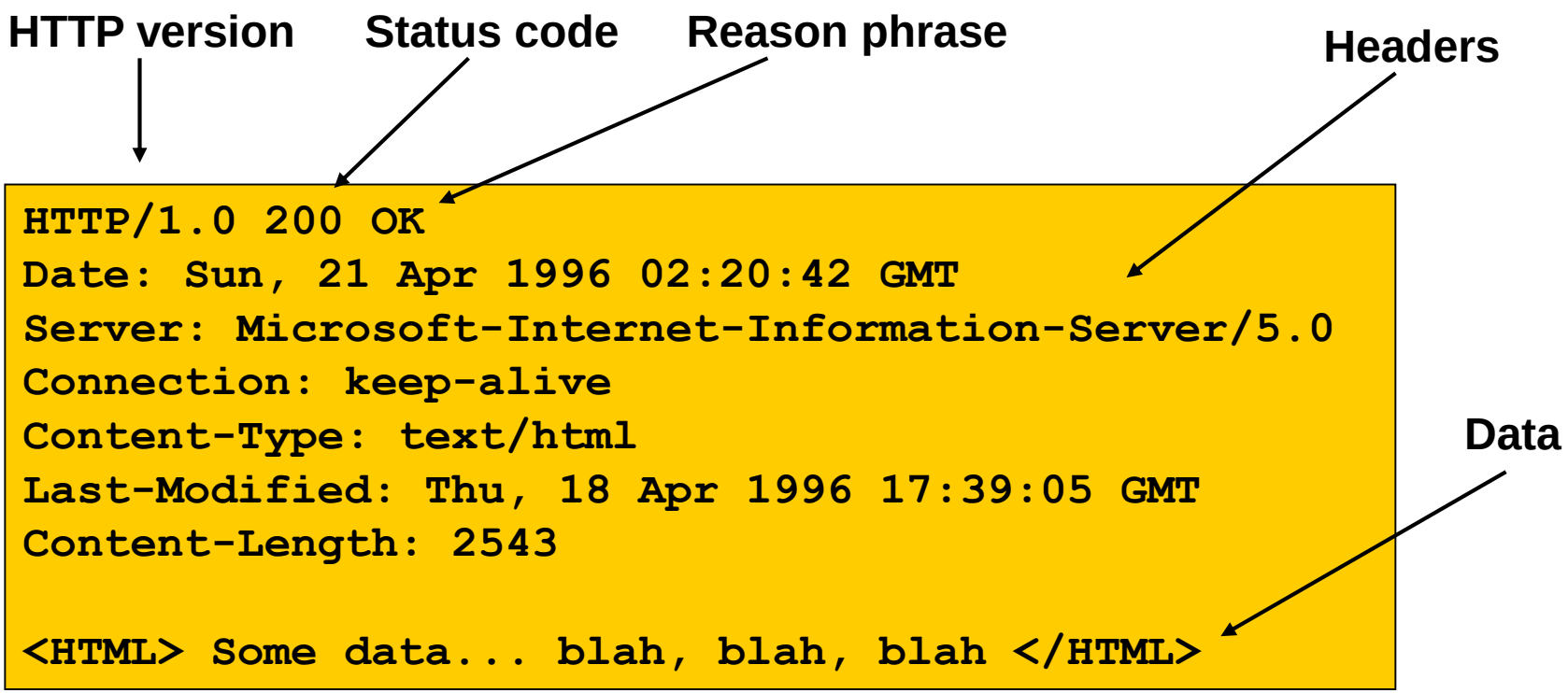
Multipurpose Internet Mail Extensions (MIME)

skip to HTTP Response

- Defines types of data/documents
  - text/plain
  - text/html
  - image/gif
  - image/jpeg
  - audio/x-pn-realaudio
  - audio/x-ms-wma
  - video/x-ms-asf
  - application/octet-stream

HTTP Response

skip to Programming the Web



Status Codes

200 OK
201 Created
202 Accepted
204 No Content
301 Moved Permanently
302 Moved Temporarily
304 Not Modified
400 Bad Request
401 Unauthorized
403 Forbidden
404 Not Found
500 Internal Server Error
501 Not Implemented
502 Bad Gateway
503 Service Unavailable

Classes:

1xx: Informational	- not used, reserved for future
2xx: Success	- action was successfully received, understood, and accepted
3xx: Redirection	- further action needed to complete request
4xx: Client Error	- request contains bad syntax or cannot be fulfilled
5xx: Server Error	- server failed to fulfill an apparently valid request

Programming the Web

Web Programming Model

- Packaged functionality
  - Web server supports default and user supplied controls
- Dynamic content display
  - ASP, JSP generates HTML using server data
  - Browser interprets client side scripts
- Machine-to-machine
  - Web services provide RPC interface

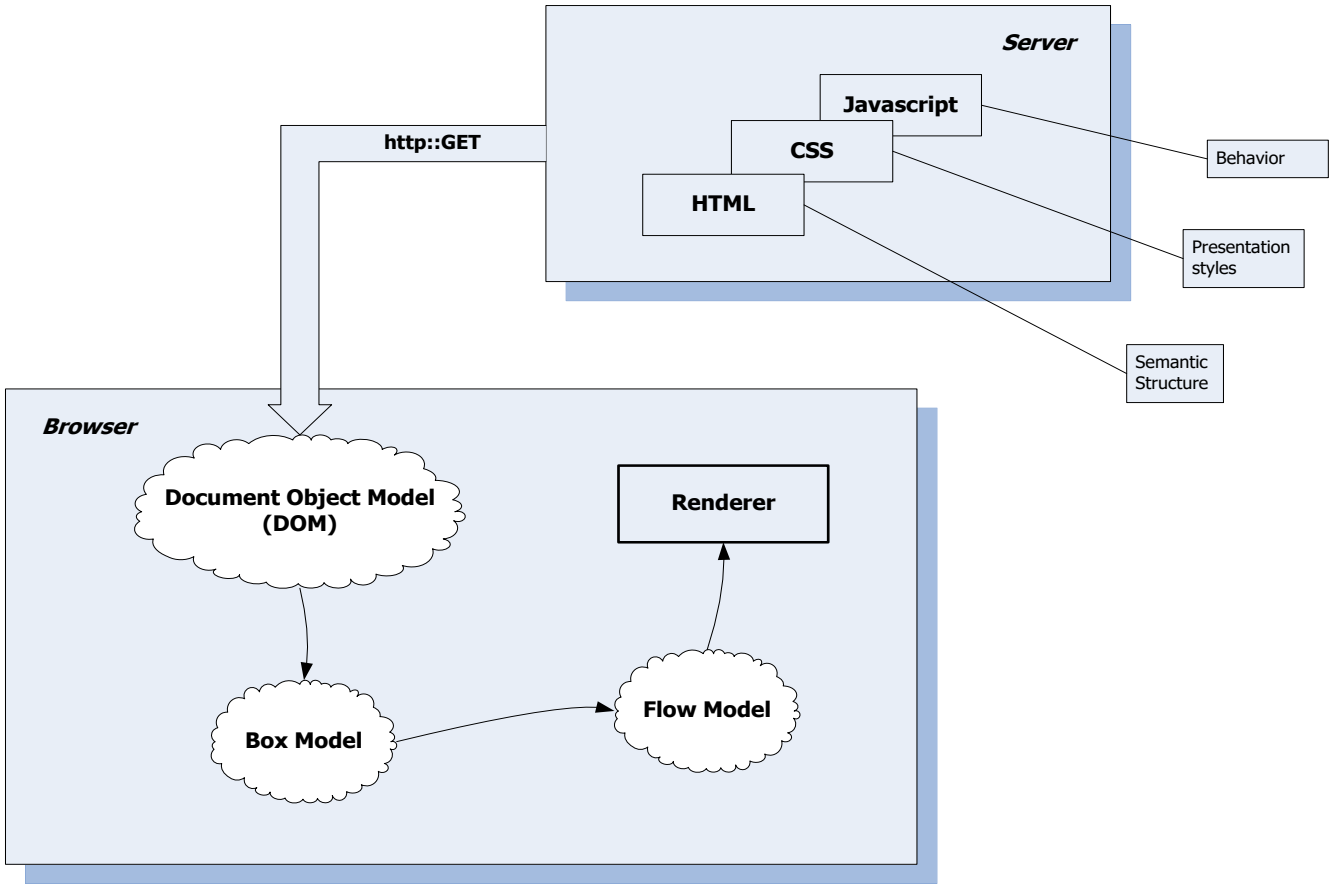
Programming the Web

- Client-side programming
  - JavaScript
    - Can modify HTML document using scripts sent from server and interpreted by client.
  - Silverlight uses C# in embedded CLR in browser plug-in
  - .Net controls, Java applets—need permissions
- Server-side programming
  - ASP script, C# code-behind
  - Server components
    - Session, application, ADO, FileSystem, …
  - Web controls used on ASPX pages
  - Web services

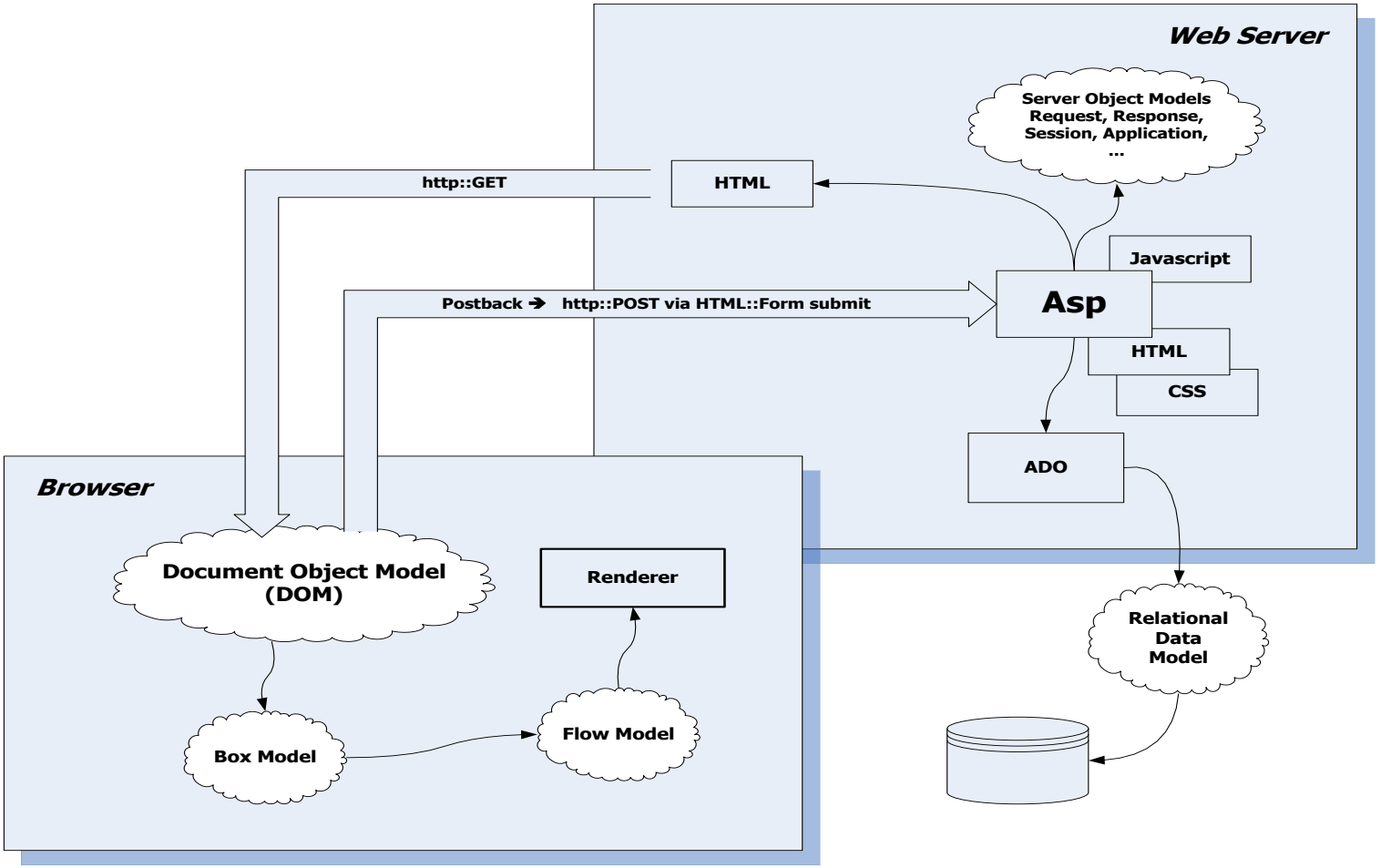
Programming the Web
Server-Side Code

- What is server-side code?
  - Software that runs on the server, not the client
  - Receives input from
    - URL parameters
    - HTML form data
    - Cookies
    - HTTP headers
  - Can access server-side databases, e-mail servers, files, web services, etc.
  - Dynamically builds a custom HTML response for a client

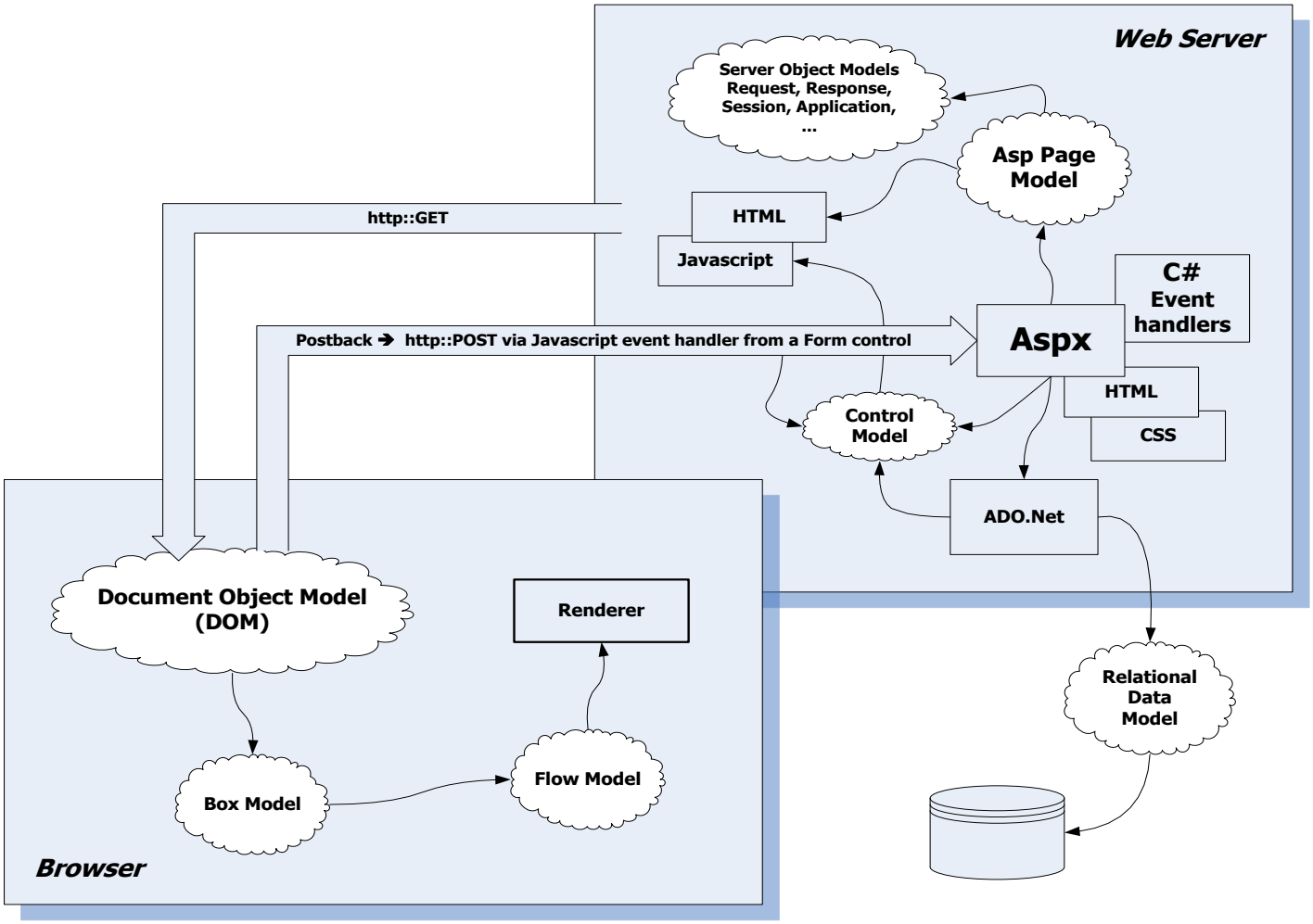
Traditional HTML Serving Model



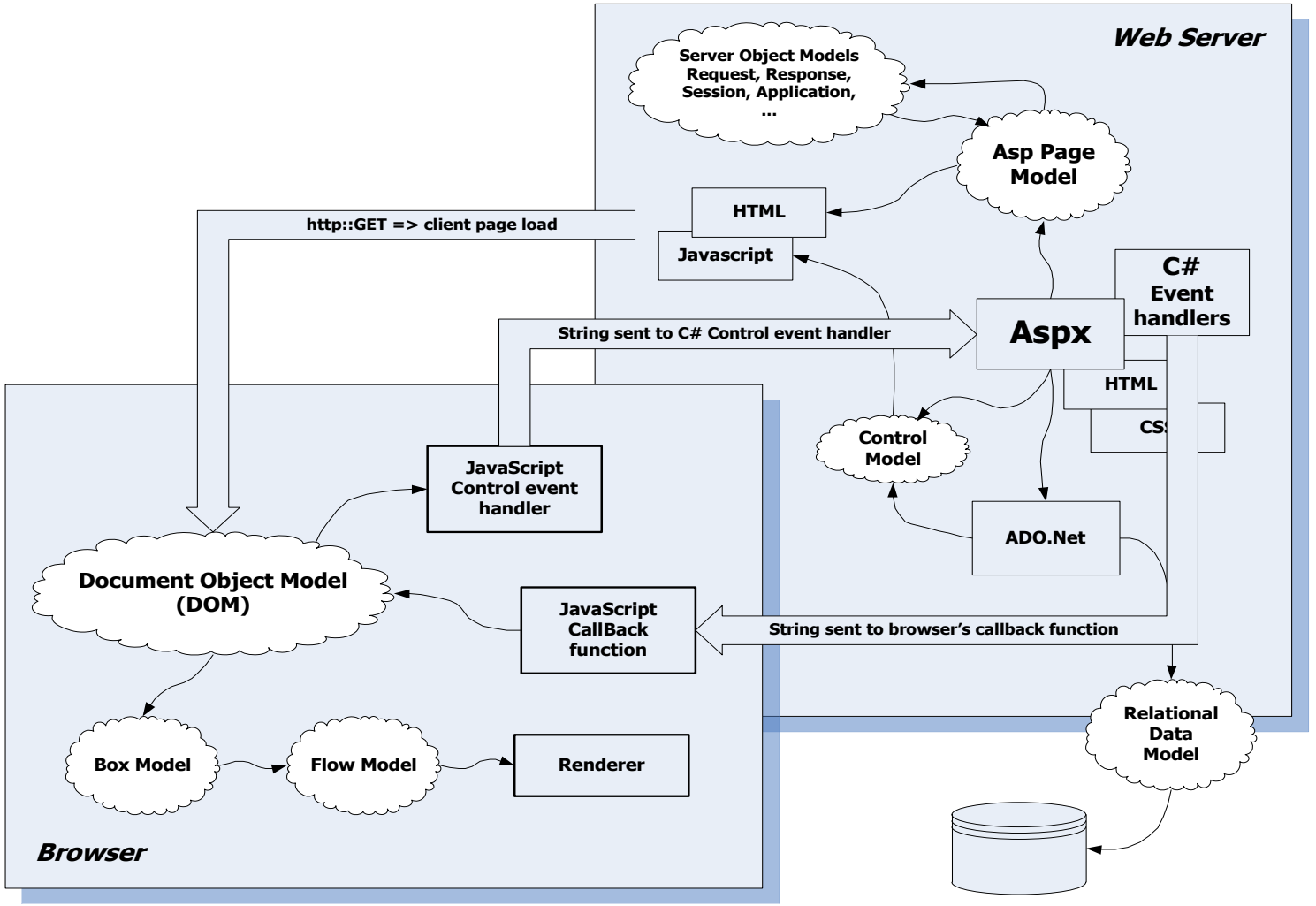
ASP Dynamic Serving Model



Asp.Net Serving Model

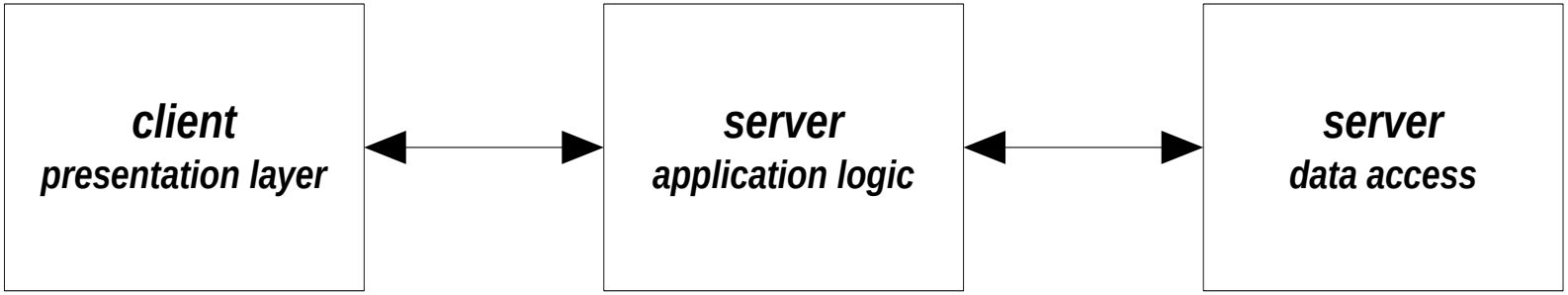


Asp.Net Ajax Serving Model



Three-Tier Architecture

- Client tier
  - Presentation layer
  - Client UI, client-side scripts, client specific application logic
- Server tier
  - Application logic, server-side scripts, form handling, data requests
- Data tier
  - Data storage and access



Some Examples

- Basic HTML pages
  - Example 1

Server Object Model

- Application object
  - Data sharing and locking across clients
- Request object
  - Extracts client data and cookies from HTTP request
- Response object
  - Send cookies or call Write method to place string in HTML output
- Server object
  - Provides utility methods
- Session object
  - If browser supports cookies, will maintain data between page loads, as long as session lasts

Security Issues

- Threats
  - Data integrity
    - Code that deletes or modifies data
  - Privacy
    - Code that copies confidential data and makes it available to others
  - Denial of service
    - Code that consumes all of CPU time or disk memory
  - Elevation of privilege
    - Code that attempts to gain administrative access

Protections

- Least-privilege rule:
  - Use the technology with the fewest capabilities that gets the job done.
- Digital signing
  - Who are you?
- Security zones
  - Trusted and untrusted sites
- Secure sockets layer (SSL)
- Transport layer security (TLS)
- Encryption

Extending the Web

Current Extensions

- Describe data with XML
- Extend HTML into XHTML
- Separate style from content with CSS
  - Cascading style sheets
  - Can be included from a file to give uniform style of pages and documents
- Document Object Model (DOM)
  - Defines a scripting interface

Areas of Exploration

- XML - Universal Data Services
- TVWeb - Merger of features
- MathML - Mathematical Markup Language
- RDF - Resource Description Framework
- Accessibility - For the handicapped
- SMIL - Synchronized Multimedia Integration Language
- Internationalization
- Speech

People in the Web

- Web development
  - Web server, HTTP
    - Tim Berners-Lee, Robert Cailiau
  - Mosaic web browser
    - Marc Andreessen
- Internet
  - TCP/IP protocol
    - Vinton Cerf, Robert Kahn
  - Internet flow control
    - Larry Roberts

References

- World Wide Web Consortium
  - Excellent Tutorial Papers, standards
- Source of several slides used here
  - Mark Sapposnek
- webdev.htm
  - Tutorials
  - Web developer's links
  - Web designer's links
  - Tech details links
- XHTML Black Book, Steven Holzner, Coriolis, 2000
  - Aging but comprehensive treatment of HTML, XHTML, JavaScript
- Web Developers Virtual Library
  - More tutorials

ENGINEERING@SYRACUSE