

UNIT 6

Internet and World Wide Web (WWW)

(6 Hours) — Operating Systems

PART A — THE INTERNET

6.1 Introduction to Internet and its Applications

The Internet is a vast, global network of interconnected computer networks that communicate using a common set of protocols, chiefly TCP/IP. It is not owned or controlled by any single organization; instead, it is a cooperative network of networks — ranging from small home networks to the backbone networks operated by major telecommunications carriers — all agreeing to exchange data using the same underlying rules.

Since its origins in the 1960s as ARPANET, a US research network, the Internet has grown into an essential part of modern life, supporting an enormous range of applications:

- Communication — email, instant messaging, video calls, and social media.
- Information Access — the World Wide Web, online encyclopedias, and digital libraries.
- E-Commerce — online shopping, banking, and digital payments.
- Entertainment — streaming video and music, online gaming.
- Cloud Computing — remote storage and computation delivered as a service.
- Education — e-learning platforms, online courses, and remote research collaboration.

6.2 Connecting to the Internet

A device connects to the Internet through an access technology provided by an Internet Service Provider (ISP), which links the local device or network to the wider Internet backbone. Common connection methods include:

Connection Type	Description
Dial-Up	Uses a telephone line and modem; slow, largely obsolete.
DSL (Digital Subscriber Line)	Uses existing telephone lines for always-on broadband access.
Cable Broadband	Uses cable television infrastructure for high-speed access.
Fibre Optic (FTTH)	Uses optical fibre for very high bandwidth and low latency.

Connection Type	Description
Wi-Fi / Wireless	Connects via radio signals to a nearby wireless access point or router.
Mobile Data (3G/4G/5G)	Connects via a cellular network using a mobile device or dongle.
Satellite	Uses satellite links; useful in remote areas lacking ground infrastructure.

6.3 Client/Server Technology

Client/server is an architecture in which tasks are divided between two types of programs: clients, which request services, and servers, which provide them. The client and server may run on the same machine but are typically on separate computers connected by a network. A single server can handle requests from many clients simultaneously.

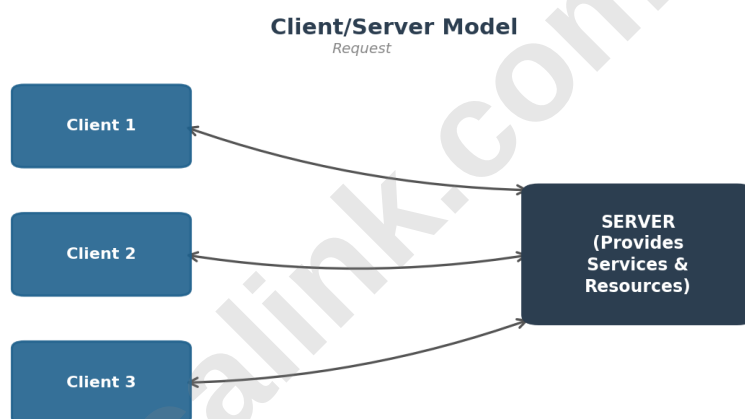


Figure 6.1 — The client/server model: multiple clients request services from a central server.

This model contrasts with peer-to-peer architecture, where each node can act as both client and server. Client/server architecture centralizes control and resources, making it easier to manage security, backups, and updates.

6.4 Internet as a Client/Server Technology

The Internet itself is built largely on the client/server model. When you browse a website, your browser acts as the client, sending a request to a remote web server, which processes it and returns the requested page. The same pattern applies to email (mail client and mail server), file transfer (FTP client and FTP server), and countless other Internet services. This design allows a single, powerful server to serve millions of clients around the world, while each client only needs enough capability to make requests and display results.

Example 6.1 — Everyday Client/Server Interaction

When you check your email on your phone, the Gmail app (client) sends a request to Google's mail servers, which retrieve your messages from storage and send them back to be displayed on your screen.

6.5 Email

Electronic mail (email) allows users to compose, send, and receive messages electronically over a network. An email address has the form `username@domain`, where the domain identifies the mail server responsible for that user's mailbox.

Sending and receiving email relies on a small set of protocols working together: SMTP is used to send outgoing mail from a client to a server (and between servers), while POP3 or IMAP is used by a client to retrieve mail from the server. IMAP additionally keeps mail synchronized across multiple devices, since messages remain on the server unless explicitly deleted.

Note: An email typically consists of a header (sender, recipient, subject, date) and a body (the message content), optionally with file attachments.

6.6 Video-Conferencing

Video-conferencing allows two or more people in different locations to hold a face-to-face conversation in real time using video and audio transmitted over the Internet. It requires a camera, microphone, and sufficient bandwidth at each participant's end, along with software or a service (e.g., Zoom, Google Meet, Microsoft Teams) that compresses and streams the audio-video data and keeps all participants synchronized.

Beyond personal communication, video-conferencing is widely used for remote business meetings, distance education, telemedicine consultations, and virtual job interviews, reducing the need for travel.

6.7 Internet Service Providers (ISP)

An Internet Service Provider is a company that supplies individuals and organizations with access to the Internet, along with related services. ISPs maintain the physical infrastructure — cables, routers, and switches — or lease capacity from larger carriers, and connect their customers' networks to the global Internet backbone.

Typical services offered by an ISP include:

- Internet access (broadband, fibre, mobile data).

- Email hosting and web hosting.
- Domain name registration.
- Technical support and network security services.

6.8 Domain Name Server (DNS)

Computers identify each other using numeric IP addresses, but humans find domain names such as `www.example.com` far easier to remember. The Domain Name System (DNS) is a distributed, hierarchical naming service that translates domain names into the IP addresses needed to route traffic — functioning much like a phone book for the Internet.

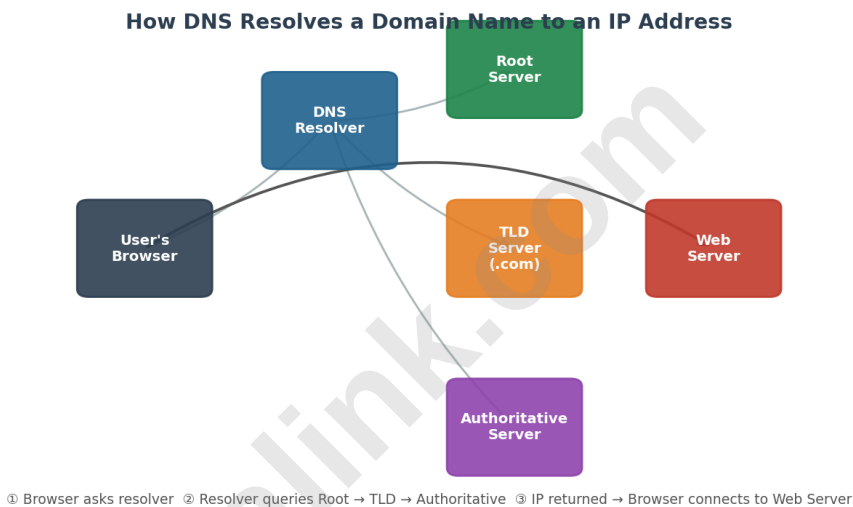


Figure 6.2 — The DNS resolution process, from browser query to web server connection.

DNS is organized as a hierarchy: the root servers know where to find the servers responsible for each Top-Level Domain (TLD) such as `.com` or `.org`; each TLD server, in turn, directs the query to the authoritative server that holds the actual IP address for a specific domain.

6.9 Internet Address (IP Address)

An IP address is a unique numeric label assigned to every device connected to a network that uses the Internet Protocol, allowing it to be located and communicated with. Two versions are in use today:

Version	Format	Example	Address Space
IPv4	32-bit, four decimal numbers separated by dots	192.168.1.10	~4.3 billion addresses

Version	Format	Example	Address Space
IPv6	128-bit, eight groups of hexadecimal digits	2001:0db8:85a3::8a2e:0370:7334	Practically unlimited

IPv6 was introduced primarily because the supply of IPv4 addresses could not keep pace with the growing number of Internet-connected devices.

6.10 Internet Protocols

A protocol is an agreed set of rules that governs how data is formatted and exchanged between systems. The Internet relies on a stack of specialized protocols, each responsible for a specific function.

Protocol	Full Form / Purpose
IP	Internet Protocol — handles logical addressing and routing of packets across networks.
TCP	Transmission Control Protocol — provides reliable, ordered, connection-based delivery of data.
HTTP	HyperText Transfer Protocol — transfers web pages between browsers and web servers.
FTP	File Transfer Protocol — transfers files between a client and a server.
SMTP	Simple Mail Transfer Protocol — sends outgoing email between mail servers.
POP	Post Office Protocol — downloads email from a mail server to a client, typically removing it from the server.
Telnet	A protocol for remote terminal access to another computer over a network (largely superseded by SSH for security reasons).
Gopher	An early protocol for organizing and retrieving documents over a network using hierarchical menus, a precursor to the Web.
WAIS	Wide Area Information Server — an early protocol/system for indexing and searching large text databases across a network.

Note: Telnet, Gopher, and WAIS are largely of historical interest today; they illustrate the Internet's evolution before HTTP and the Web became dominant.

6.11 Introduction to Intranet

An intranet is a private network, based on Internet technologies (TCP/IP, HTTP, web browsers), that is restricted to the members, employees, or authorized users of a single organization. It looks and behaves like a small, private version of the Internet, typically used to share internal news, documents, policies, and applications, protected from outside access by firewalls.

6.12 Internet vs. Intranet vs. Extranet

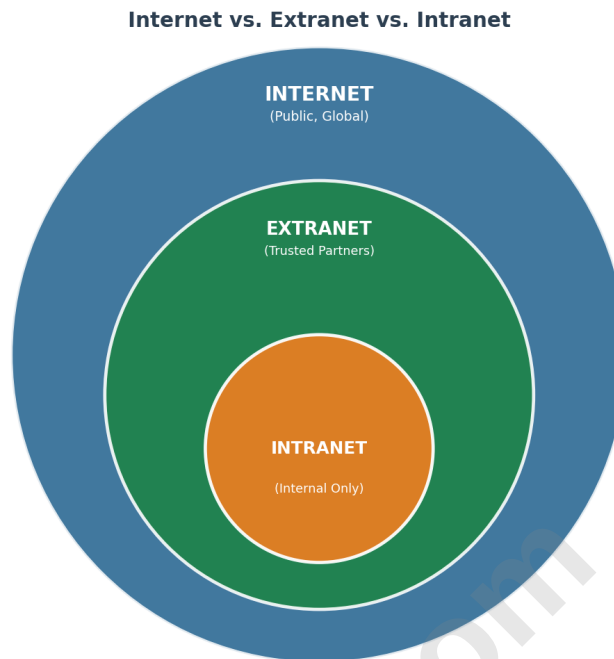


Figure 6.3 — The Intranet is private to an organization, the Extranet extends limited access to trusted partners, and the Internet is public.

Aspect	Internet	Intranet	Extranet
Access	Open to the public	Restricted to organization members	Restricted to the organization plus selected outside partners
Ownership	No single owner	Owned/controlled by one organization	Owned/controlled by one organization, shared selectively
Purpose	Global information sharing and services	Internal communication and resource sharing	Secure collaboration with vendors, suppliers, or clients

6.13 Advantages & Disadvantages of Intranet

Advantages	Disadvantages
Fast, secure sharing of internal information	Setup and maintenance can be costly
Improves collaboration and communication among employees	Requires ongoing IT management and security oversight
Centralizes documents, policies, and applications	Risk of information overload if poorly organized

Advantages	Disadvantages
Reduces paper use and communication costs	Limited to authorized users only; not accessible remotely without extra setup

— *End of Part A* —

bcalink.com

6.14 World Wide Web and Its Evolution

The World Wide Web (WWW) is a system of interlinked hypertext documents and other resources, accessed via the Internet using a web browser. It was invented by Tim Berners-Lee in 1989 at CERN as a way for researchers to share documents through hyperlinks, and was made publicly available in 1991. It is important to note that the Web is not the same as the Internet: the Internet is the underlying global network of networks, while the Web is one of many services (alongside email, FTP, and others) that runs on top of it.

Era	Characteristics
Web 1.0 (early-to-mid 1990s)	Static, read-only pages; limited interactivity; simple hyperlinked documents.
Web 2.0 (mid-2000s onward)	Interactive, user-generated content; social media, blogs, wikis, dynamic applications.
Web 3.0 (emerging)	Semantic, data-driven web; increasing use of AI, decentralization, and interconnected data.

6.15 Architecture of Web

The Web follows a client/server architecture in which a web browser (client) requests documents from a web server, which responds over the HTTP protocol.

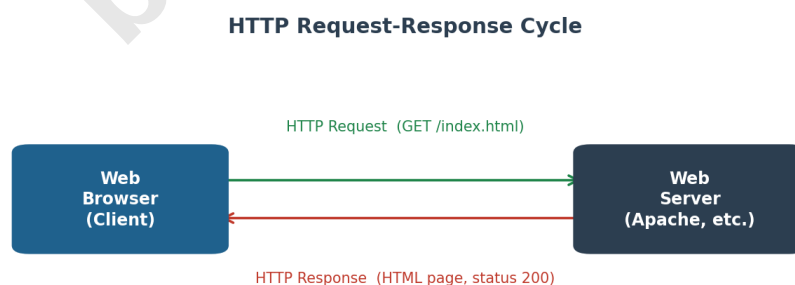


Figure 6.4 — The basic request-response cycle underlying the Web.

Three core building blocks make this possible: HTML (HyperText Markup Language) structures the content of a page; HTTP defines the rules for requesting and delivering that content; and URLs uniquely identify each resource on the Web so it can be requested. Modern web architecture often adds application and database layers behind the web server to generate pages dynamically rather than serving fixed files.

6.16 Uniform Resource Locator (URL)

A URL is the address used to locate a specific resource on the Web. It follows a standard structure:

Example 6.2 — Anatomy of a URL

`https://www.example.com:443/products/index.html?id=42#reviews`

`https` — protocol (scheme) used to access the resource

`www.example.com` — domain name (host) of the server

`443` — port number (often omitted; defaults to 80 for HTTP, 443 for HTTPS)

`/products/index.html` — path to the specific resource on the server

`?id=42` — query string, passing parameters to the server

`#reviews` — fragment identifier, pointing to a section within the page

6.17 Browsers

A web browser is application software that requests, interprets, and displays web resources, rendering HTML, CSS, and JavaScript into the pages users see and interact with.

Browser	Notes
Internet Explorer	Developed by Microsoft; bundled with Windows for many years; now retired in favour of Edge.
Netscape Navigator	One of the earliest popular browsers from the 1990s; its codebase later gave rise to the Mozilla project.
Opera	Known for early innovation in tabbed browsing, speed-optimization features, and a built-in VPN.
Firefox	Developed by the Mozilla Foundation; open-source, privacy-focused, descended from the Mozilla browser suite.
Chrome	Developed by Google; currently the most widely used browser, built on the open-source Chromium project.
Mozilla (Suite)	The original open-source application suite (browser, mail client, and more) from which Firefox was later spun off as a standalone browser.

6.18 Search Engine

A search engine is a web-based tool that helps users find information on the Web by indexing content from millions of pages and returning ranked results relevant to a user's query. Search engines generally operate in three stages:

- Crawling — automated programs (crawlers or spiders) follow links to discover pages across the Web.
- Indexing — the content of discovered pages is analyzed and stored in a searchable index.
- Ranking — when a user searches, the engine scores and orders matching pages by relevance and quality.

Note: Popular search engines include Google, Bing, and DuckDuckGo, each using its own ranking algorithms and criteria.

6.19 Web Servers

A web server is software (often running on dedicated hardware) that stores web content and delivers it to clients in response to HTTP requests.

6.19.1 Apache

The Apache HTTP Server is a free, open-source web server maintained by the Apache Software Foundation. It has been one of the most widely used web servers since the mid-1990s, valued for its stability, extensibility through modules, and cross-platform support.

6.19.2 Proxy Server

A proxy server acts as an intermediary between a client and the destination server, forwarding requests on the client's behalf. It can be used to cache frequently requested content for faster access, filter or block certain content, hide a client's IP address for privacy, and enforce organizational access policies.

Illustration — Why Use a Proxy Server

A school network routes all student web traffic through a proxy server. The proxy blocks access to inappropriate sites, caches popular educational resources to reduce load times, and logs usage for administrative review — all without each student's device needing individual configuration.

6.20 HTTP Protocol

HyperText Transfer Protocol (HTTP) is the application-layer protocol that governs how messages are formatted and transmitted between web browsers and web servers, and how servers should respond to various requests. HTTPS is the secure version of HTTP, adding encryption via TLS/SSL so that data cannot be read or tampered with in transit.

Method	Purpose
GET	Requests a specified resource, such as a web page.
POST	Submits data to the server, such as a completed form.
PUT	Updates or replaces an existing resource on the server.
DELETE	Removes a specified resource from the server.

6.21 FTP Protocol

File Transfer Protocol (FTP) is used to transfer files between a client and a server over a network. Unlike HTTP, which typically fetches individual web resources, FTP is designed specifically for uploading and downloading whole files and directories, and commonly uses separate connections for control commands and actual data transfer.

FTP supports both authenticated access, where users log in with a username and password, and anonymous access, which allows the public to download files without credentials. Because standard FTP transmits credentials and data in plain text, secure variants such as FTPS and SFTP are preferred where security matters.

Key Takeaways — Unit 6

- ✓ The Internet is a global network of networks; the Web is one service that runs on it.
- ✓ Client/server architecture underlies most Internet services, including the Web, email, and FTP.
- ✓ DNS translates human-readable domain names into the IP addresses computers use to communicate.
- ✓ Internet protocols (IP, TCP, HTTP, FTP, SMTP, POP) each handle a distinct part of getting data across the network.
- ✓ Intranets are private, organization-only networks; extranets extend limited access to trusted partners.
- ✓ A URL identifies a specific web resource, combining protocol, domain, path, and optional parameters.
- ✓ Web servers (e.g., Apache) and proxy servers work together to deliver, cache, and control access to web content.

— *End of Unit 6* —