Introduction to Web Programming

Lydia Jane G

January 9, 2024

Table of contents

Learning Objectives

- Today we are going to learn about
  - The Internet and the WWW
  - History of Internet
  - Web protocols
  - Web Organizing and Addressing
  - Internet Resources
  - URL
  - MIME Types
  - Plug-ins
  - Net meeting and Chat
  - Search Engines

Internet and WWW

Internet and WWW

Are they the same?

Internet and WWW

Are they the same?
Is there a difference?

Internet and WWW

Are they the same?
Is there a difference?
What is it?

How do you use Internet?

How do you use Internet?

- eMail

How do you use Internet?

- eMail
- IM, Skype

How do you use Internet?

- eMail
- IM, Skype
- Gaming

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software
- Streaming video and audio

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software
- Streaming video and audio
- Learning

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software
- Streaming video and audio
- Learning
- Remote Backup

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software
- Streaming video and audio
- Learning
- Remote Backup
- Collaboration - Participation (Wiki)

How do you use Internet?

- eMail
- IM, Skype
- Gaming
- Search Engines
- Blogging
- Social networks
- Software
- Streaming video and audio
- Learning
- Remote Backup
- Collaboration - Participation (Wiki)
- Video Conferencing

Internet

Internet

Internet is short for **INTERconnected NETwork**
Internet is a network of networks
A network may include:

Internet

Internet is short for **INTERconnected NETwork**

Internet is a network of networks

A network may include:

- Personal Computers

Internet

Internet is short for **INTERconnected NETwork**

Internet is a network of networks

A network may include:

- Personal Computers
- Servers

Internet

Internet is short for **INTERconnected NETwork**

Internet is a network of networks

A network may include:

- Personal Computers
- Servers
- Printers

A network is connected through a communication channel

Networks

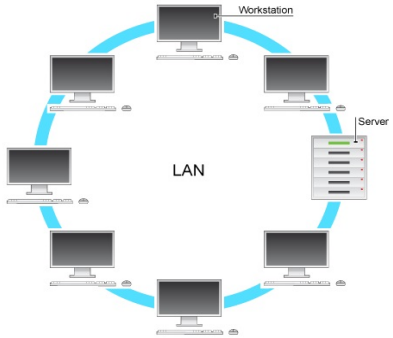


Figure: LAN

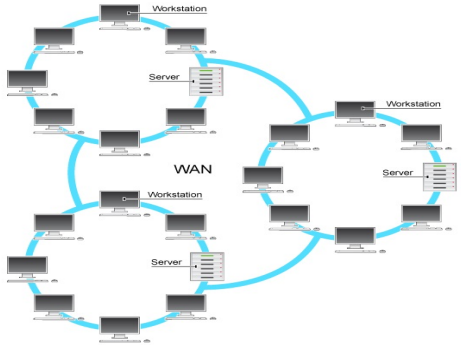


Figure: WAN

History -Internet

- Research by DoD US 1962
- ARPANET
- First long distance communication 1965
- Packet Switching - Leonard Kleinrock - 1970's
- TCP/IP protocol - Vint Cerf and Bob Kahn
- Ethernet -Xerox-early 1980's
- Personal Computers - IBM - 1981
- Private networks by organizations
- **INTERNET**

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's
- The WWW is a service that operates over the internet
- WWW combines the following 4 ideas:

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's
- The WWW is a service that operates over the internet
- WWW combines the following 4 ideas:
  - Hypertext

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's
- The WWW is a service that operates over the internet
- WWW combines the following 4 ideas:
  - Hypertext
  - URI/URL

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's
- The WWW is a service that operates over the internet
- WWW combines the following 4 ideas:
  - Hypertext
  - URI/URL
  - Client-Server model computing (Web browsers and servers)

History - WWW

- Often the phrases **'www'** or **'the web'** is used interchangeably with internet
- But they are **different** entities
- The web did not exist until the late 1980's
- The WWW is a service that operates over the internet
- WWW combines the following 4 ideas:
  - Hypertext
  - URI/URL
  - Client-Server model computing (Web browsers and servers)
  - Markup Language

WWW

WWW

- The web is the brain child of Tim Berners-Lee

WWW

- The web is the brain child of Tim Berners-Lee
- He created a set of technologies that allowed information on the internet to be linked through the use of links, or connections in documents

WWW

- The web is the brain child of Tim Berners-Lee
- He created a set of technologies that allowed information on the internet to be linked through the use of links, or connections in documents
- The language used to write these documents is HTML

WWW

- The web is the brain child of Tim Berners-Lee
- He created a set of technologies that allowed information on the internet to be linked through the use of links, or connections in documents
- The language used to write these documents is HTML
- All clients and servers in WWW speak the HTTP language

WWW

- The web is the brain child of Tim Berners-Lee
- He created a set of technologies that allowed information on the internet to be linked through the use of links, or connections in documents
- The language used to write these documents is HTML
- All clients and servers in WWW speak the HTTP language
- Mosaic Browser - First Browser with graphical interface - Marc Anderson

Browsers

- An application that provides a way to
  - Look at and
  - Interact

  with information on the World Wide Web
- It
  - Retrieves
  - Presents and
  - Traverses

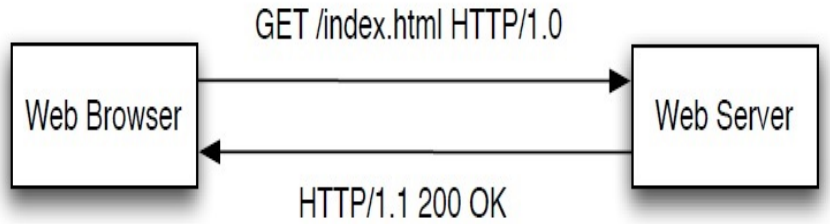
  information resources

Browsers

Which one do you use?

Web Servers

- The web server is responsible for:
  - Accepting requests for content described through URL
  - Checking whether access is permitted, requesting for authentication for any
  - Sending the content back to browser



Web Server

- A Web Server is the **machine** *and* the **software** that serves the content
- Some of the popular web servers are:
  - Apache Web Server
  - Microsoft IIS Server
  - Sun Java Server

Web Protocols

- Internet relies on a number of protocols to function properly
- A protocol is a standard for enabling the connection,communication and data transfer between two computers on a network
- The Internet Protocol Suite consists of a set of protocols. Some of them are:
  - TCP/IP
  - HTTP
  - HTTPS
  - FTP
  - UDP

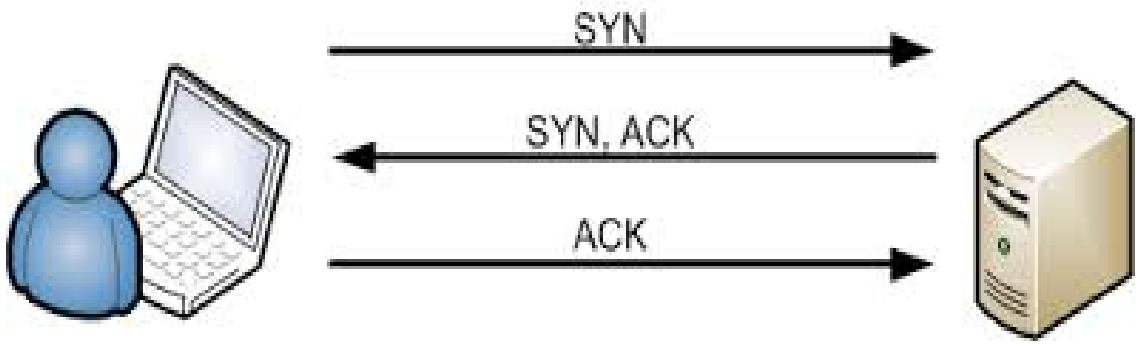
TCP/IP

- TCP/IP defines how electronic devices should be connected to the network
- They also define how data transfer should happen among them
- TCP/IP protocol is embedded in the TCP/IP software that is part of the OS
- TCP handles communication between applications
  - TCP uses a fixed connection
  - If one application wants to communicate with another via TCP,
    - It sends a communication request
    - This request must be sent to an exact address
    - After a 'handshake' between the applications a communication line opens

TCP/IP

TCP 3-WAY Handshake

- Host A sends a TCP SYNchronize packet to Host B
- Host B receives A's SYN
- Host B sends a SYNchronize-ACKnowledgement
- Host A receives B's SYN-ACK
- Host A sends ACKnowledge
- Host B receives ACK.
- TCP socket connection is ESTABLISHED.



TCP/IP

- IP handles communications between computers
- IP is responsible for routing each packet to its correct destination
- Communicating via IP is like sending a long letter as a large number of small postcards, each finding its own (often different) way to the receiver
- IP is a connection-less protocol
- Every computer has an IP Address!

TCP/IP

- TCP/IP is TCP and IP working together
- TCP takes care of the communication between application software and network software
- IP takes care of communication with other computers
- TCP is responsible for breaking down data to packets and assembling the packets when they arrive
- IP is responsible for routing the packets to the correct destination
- IPv4: 32-bit address
  192.168.1.1
- IPv6: 128-bit address
  fe80:0:0:0:462a:60ff:fef6:278a

HTTP

- HyperText Transfer Protocol
- HTTP takes care of communication between a web browser and a web server
- Before HTTP, FTP protocol was used
- HTTP uses the concepts of hyperlinks
- There are four messages within this protocol:
  - Connection: Establishes a connection between the client and the server
  - Request: Asks for a resource
  - Response: Delivers the resource
  - Close: Terminates the connection

HTTPS

- HTTP Secure
- Also called as HTTP over TLS, HTTP over SSL
- It is a protocol used for secure connection over internet
- Guarentees identity of server

FTP

- File Transfer Protocol
- Used to transfer files from one computer to another
- Authentication using a clear-text protocol
- Anonymous connection can be allowed
- FTPS - Secure FTP

UDP

- User Datagram Protocol
- Connectionless
- No handshake
- No guarantee of delivery

Addresses on the Web

- Each computer on the internet has a unique identification number called an IP address.
- TCP/IP uses four numbers to address a computer. The numbers are always between 0 and 255.
- IP addresses are normally written as four numbers separated by a period, like this: 192.68.20.50
- TCP/IP uses 32 bit addresses. One computer byte is 8 bits. So TCP/IP uses 4 computer bytes.
- A byte can hold 256 different values: – 00000000, 00000001, 00000010, 00000011, 00000100, 00000101, 00000110, 00000111, 00001000 .....and all the way up to 11111111
- The rapid growth of the Internet has led to a shortage of IP addresses. No one could have anticipated the Internet when the protocol was first devised
- The Internet Protocol (IPv6) will provide relief to this problem by lengthening the IP address from 32 bits to 128 bits

Domain Names

- Names are easier to remember than 12 digit numbers!
- Most browsers do not use IP addresses but they use domain names to locate web sites and individual pages
- A domain name is a unique name associated with a specific IP address by a program that runs on an Internet host computer.
- This program, which coordinates the IP addresses and domain names for all computers attached to it, is called DNS (Domain Name System) software.
- The host computer that runs this software is called a domain name server

Domain Names

- When you address a web site, like http://www.vit.ac.in, the name is translated to a number by a Domain Name Server (DNS)
- When a new domain is registered together with a TCP/IP address, DNS servers all over the world are updated with this information
- Domain names can include any number of parts separated by periods, however most domain names currently in use have only three or four parts.
- Domain names follow hierarchical model that you can follow from top to bottom if you read the name from the right to the left.
- For example, the domain name academics.vit.ac.in is the computer connected to the Internet, which is an academic unit of the VIT Universiy (vit), which is an academic institution (ac) based in India (in).
- No other computer on the Internet has the same domain name.

URL

- The IP address and the domain name each identify a particular computer on the Internet
- However, they do not indicate where a Web page's HTML document resides on that computer
- To identify a Web pages exact location, Web browsers rely on Uniform Resource Locator (URL)
- A full Web address is like: https://academics.vit.ac.in/faculty/fac_login.asp

URL

URL is a four-part addressing scheme that tells the Web browser:

- What transfer protocol to use for transporting the file
- The domain name of the computer on which the file resides
- The pathname of the folder or directory on the computer on which the file resides
- The name of the file

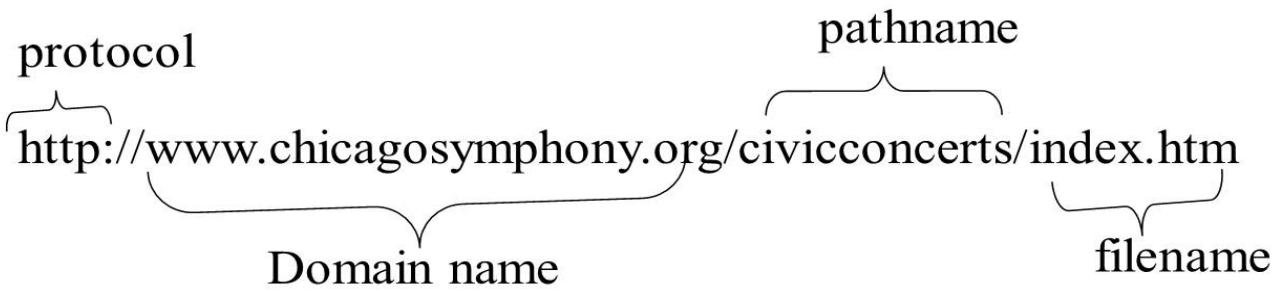
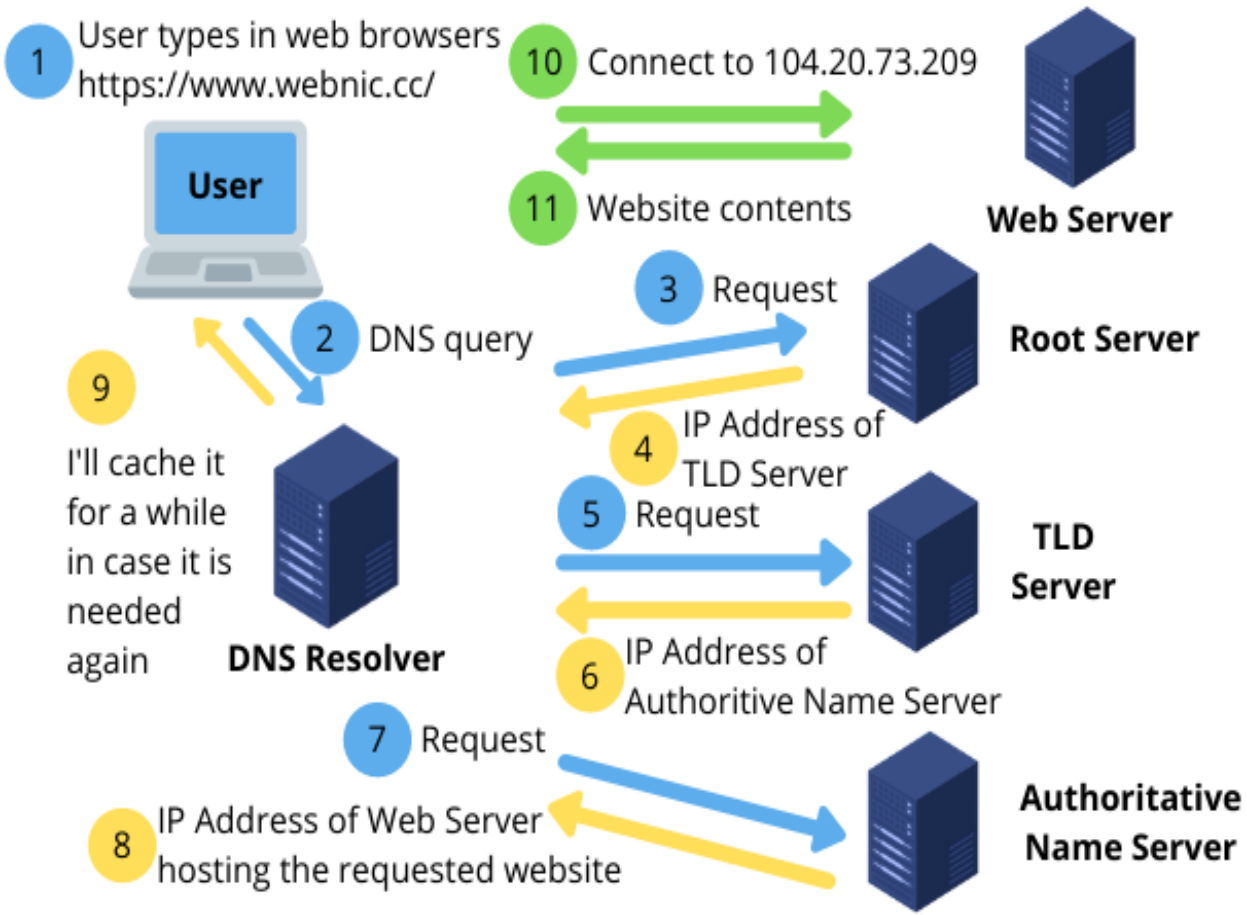
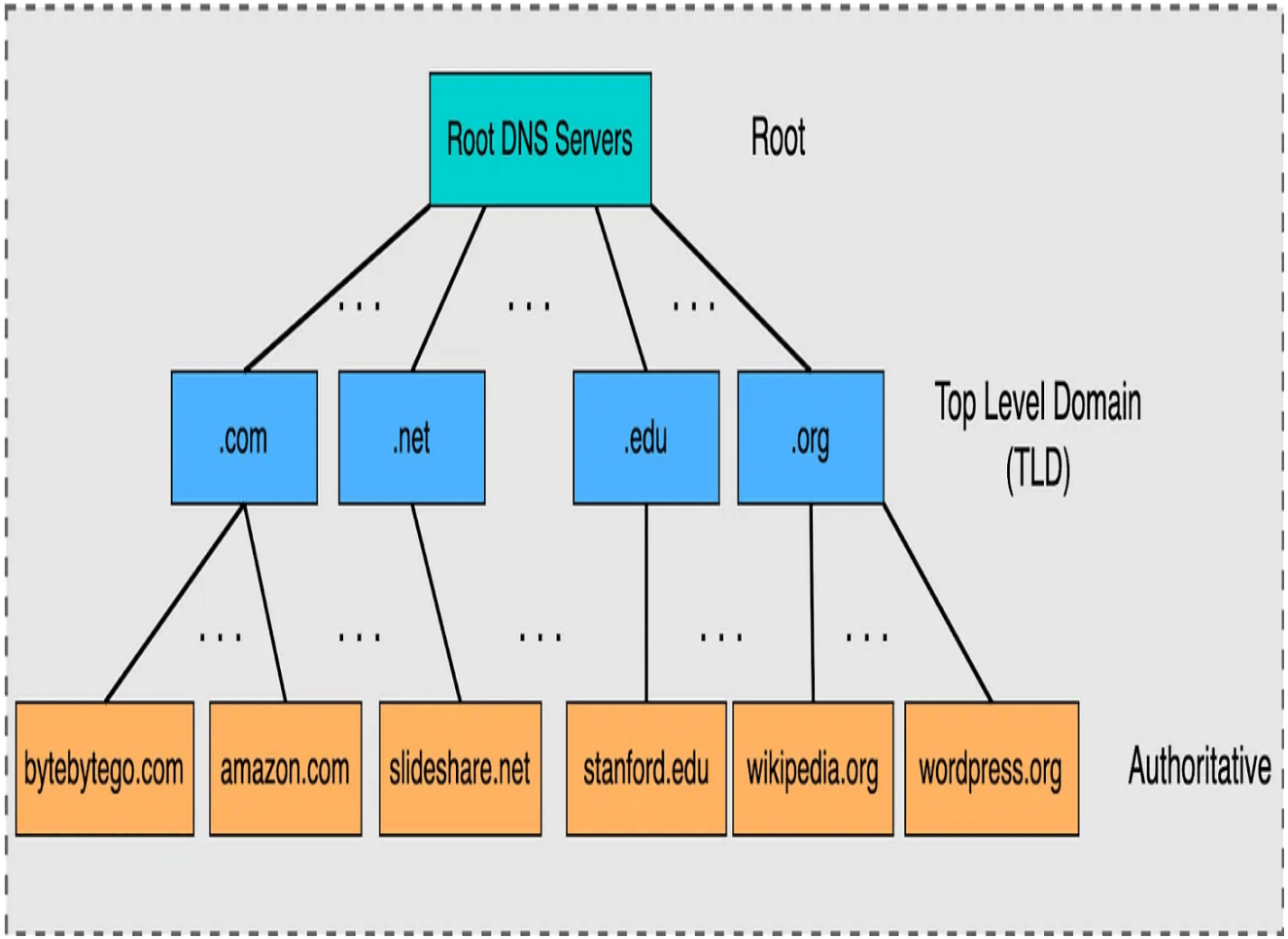


Figure:

How does the information get to me?



DNS Hierarchy



Root DNS Servers

- At the top of the DNS hierarchy are the root server
- The root servers are contacted when a server is not actually able to resolve a name
- Much of the infrastructure associated with the root servers is the responsibility of ICANN (Internet Corporation for Assigned Names and Numbers)
- There are 13 logical root servers around the world
- but each of these logical root servers is actually replicated, so corresponding to these 13 logical servers are actually close to a thousand physical servers around the world

Top-Level Domain (TLD) DNS servers

- Each of the servers in the TLD layer is responsible for resolving one of the addresses that have an ending like .com, .edu, .net, and .org
- All top-level country domains like uk, fr, ca, in, etc

Authoritative DNS servers

- Authoritative servers are the definitive source for domain name resolutions within their specific domain
- They store the actual name-to-IP address mappings for a given domain
- Domain owners or administrators configure their domain's records, but the actual infrastructure—the DNS servers—is often maintained and operated by DNS hosting providers or registrars like Cloudflare, Namecheap, GoDaddy, and others

Recursive DNS Servers (Resolver)

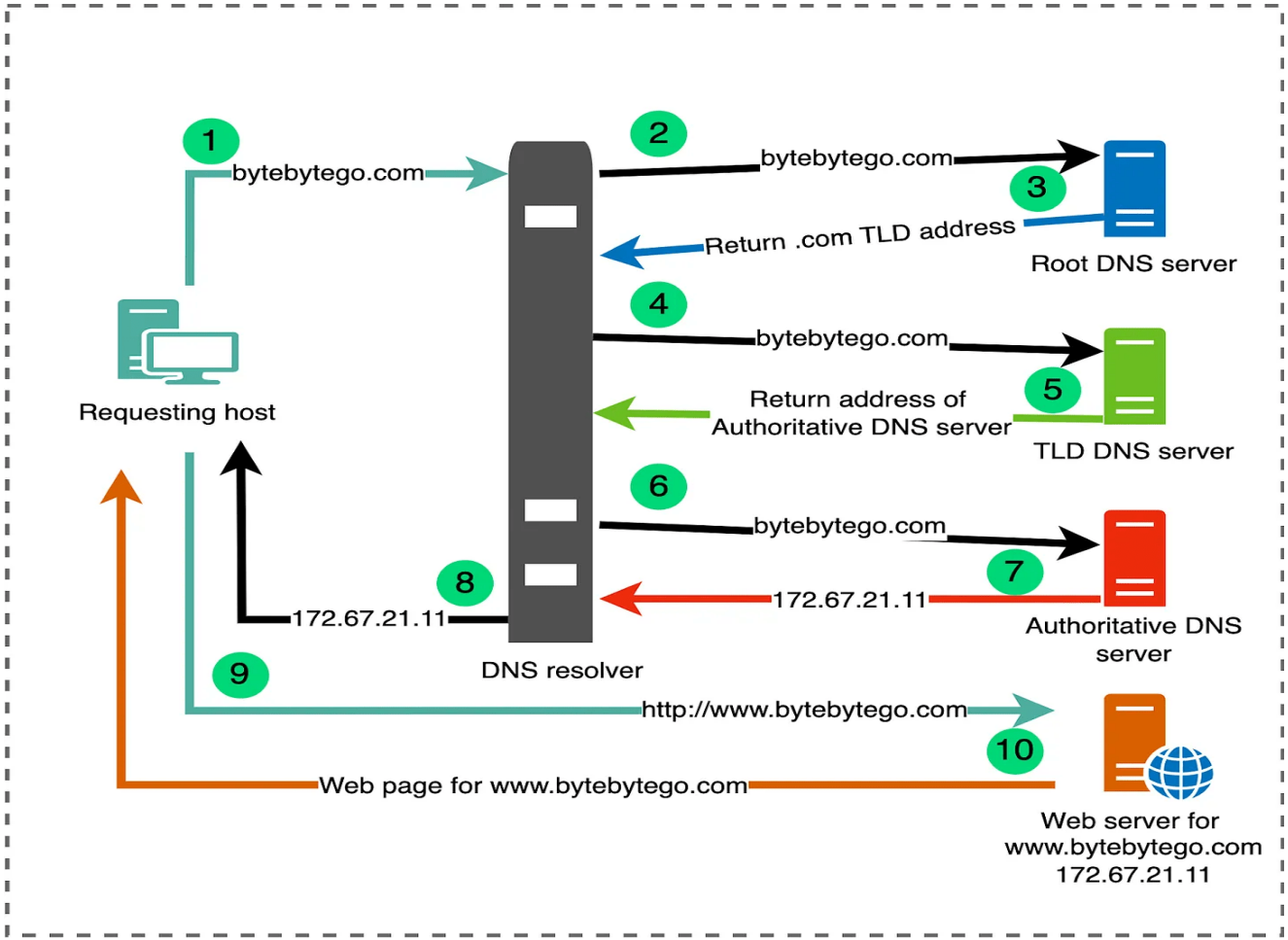
- Recursive servers handle DNS queries from client devices like computers and smartphones
- When a device wants to resolve a domain name, it contacts these servers. Acting on behalf of the client, recursive servers traverse the DNS hierarchy, consulting various DNS servers to determine the IP address associated with a domain name
- nce they obtain the answer, they return it to the client
- or efficiency, recursive servers often cache responses to avoid repeatedly querying the same information

How DNS Resolution Works

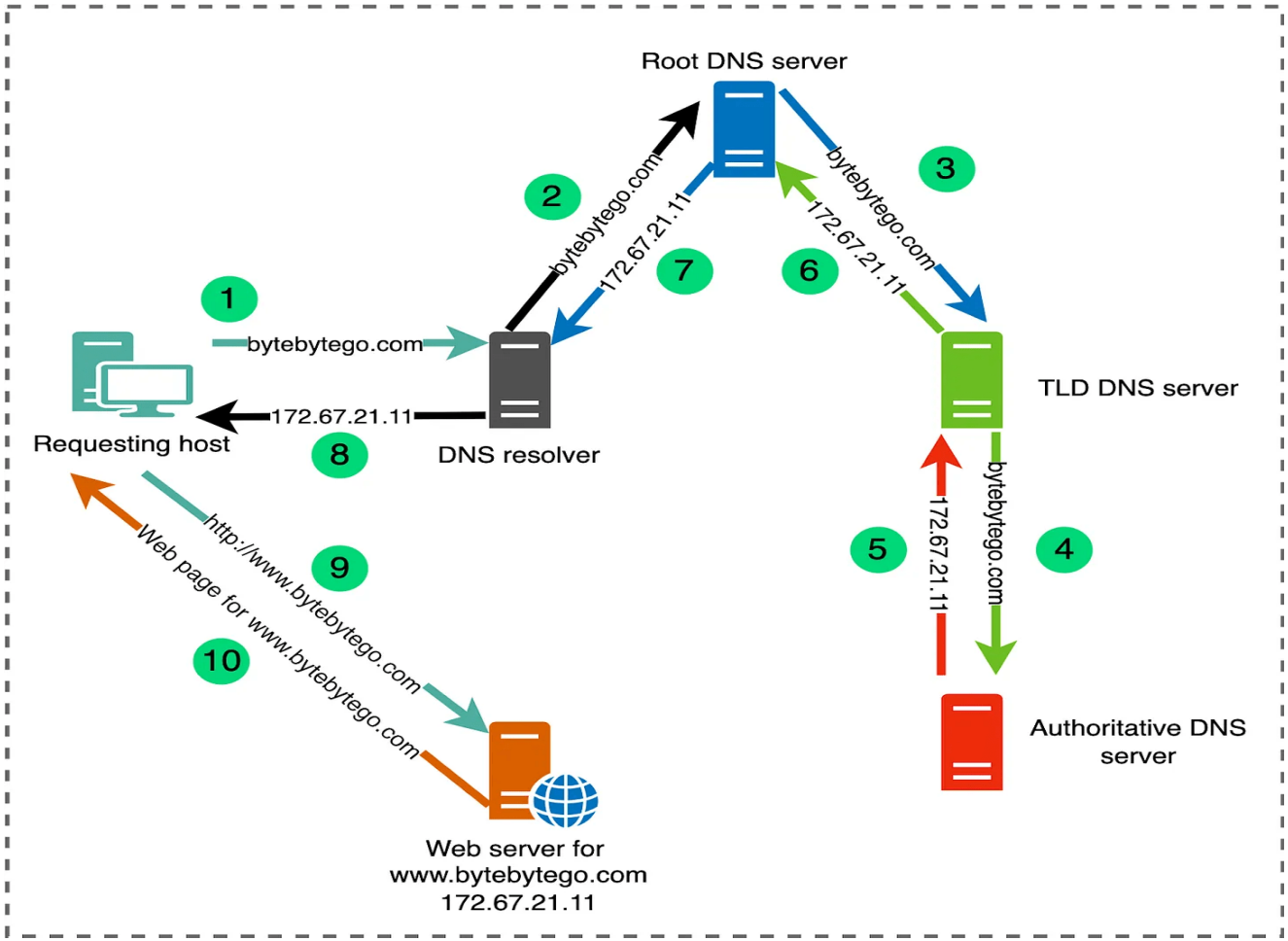
There are two main methods of query resolution in DNS:

- Iterative query resolution
- Recursive query resolution

Iterative Query Resolution



Recursive Query Resolution



Email

- Electronic Mail
- Transmission of messages over communication networks
- Email may contain text, mail, images and attachments
- Can be set to an individual or can be broadcast to a group
- First Ray Tomlinson in 1971.
- Several fields are required when sending an e-mail:
  - **To** field: where you type the e-mail address of the person who is the recipient of your message.
  - **From** field: contains your e-mail address.
  - **CC** field: allows you to send a copy of the message to another e-mail address, but is not mandatory.
  - **Subject Line**: although not required, should consist of a few words describing the e-mail's contents.
  - **Message Body**: is the location you type your main message. It often contains your signature at the bottom; similar to a hand-written letter.

Newsgroups

- Newsgroups are Internet discussion forums
- roups of users with common interests gather to talk about everything from software to movies to politics
- Unlike email messages, newsgroup messages can be read by anyone
- Are generally organized by subject

UUCP

- Unix-to-Unix Copy
- Suite of computer programs and protocols allowing remote execution of commands and transfer of files, email between computers

Finger

- An Internet utility (or program) that lets you find out information about an Internet user who has an e-mail address
- Person's real name and whether or not he or she is online at the present moment.
- Initially this utility used to work only for identifying Unix users
- Now it can now finger someone on the World Wide Web.
- Many e-mail programs now have a finger utility built into them

Chat

- Real time communication between users via computer
- When talking to someone in chat any typed text is received by other participants immediately

MIME

- Multi-purpose Internet Mail Extensions
- Also known as Internet Media Type
- It is a way of identifying files on the internet according to their nature and format
- For example, using the "Content-type" header value defined in a HTTP response, the browser can open the file with the proper extension/plugin

Plug-ins

- Also called as addin /addon or extension
- plug-in is a software component that adds a specific feature to an existing browser application
- When an application supports plug-ins, it enables customization
- The common examples are the plug-ins used in web browsers to add new features such as search-engines, or the ability to utilize a new file type such as a new video format
- Popular plug-ins are
  - Adode's Acrobat
  - Adobe Flash player
  - Real Player
  - Java Plugin

Net Meeting

- Multiperson chat
- group chat feature lets you talk with many friends at oncE
- You can chat with up to 100 people in a group discussion
- Any participant can invite others to join

Search Engines

- A program that searches for and identifies items in a database that correspond to keywords or characters specified by the user, used especially for finding particular sites on the World Wide Web.
- The search results are generally presented in a line of results often referred to as search engine results pages (SERPs)
- The information may be a mix of web pages, images, and other types of files
- Some popular search engines are:
  - Google
  - Yahoo!Search
  - MSN search
  - Bing

Prepare to get your hands Dirty!