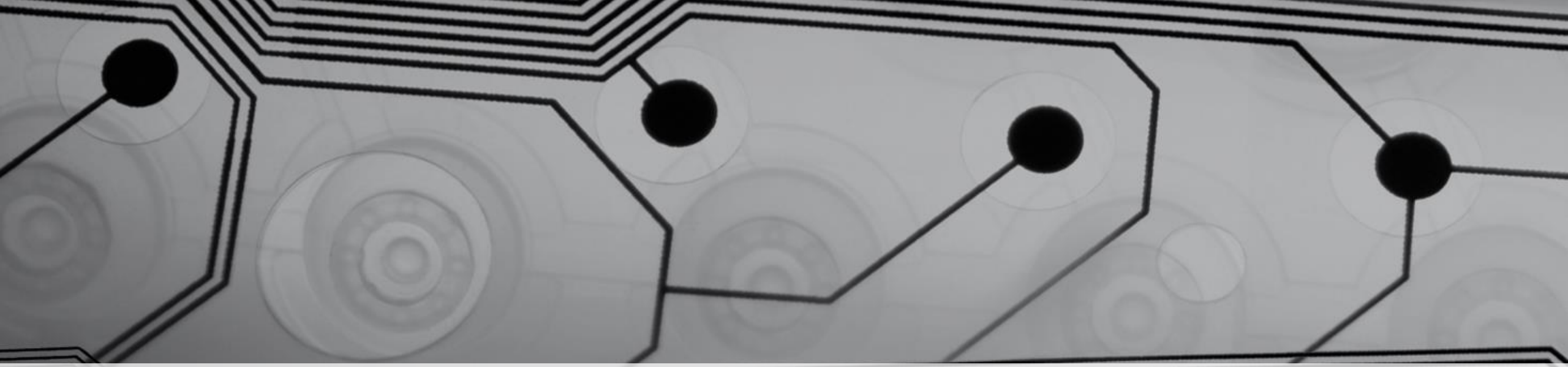


Overview of the Internet, WWW, and DNS

2020/21 COMP3322A Modern Technologies of WWW

Contents

- What is the Internet?
- What is WWW?
- Client-Server Communication
- A brief introduction to DNS

Learning Outcome

- LO1 - [Web technologies] be able to master the **key technologies** about the World Wide Web.

What is the Internet?

- Users see it as **services and applications**
  - Web and e-commerce
  - Email, texting, instant messenger
  - Social networking and blogs
  - Music and video download (and upload)
  - Voice and video teleconferencing



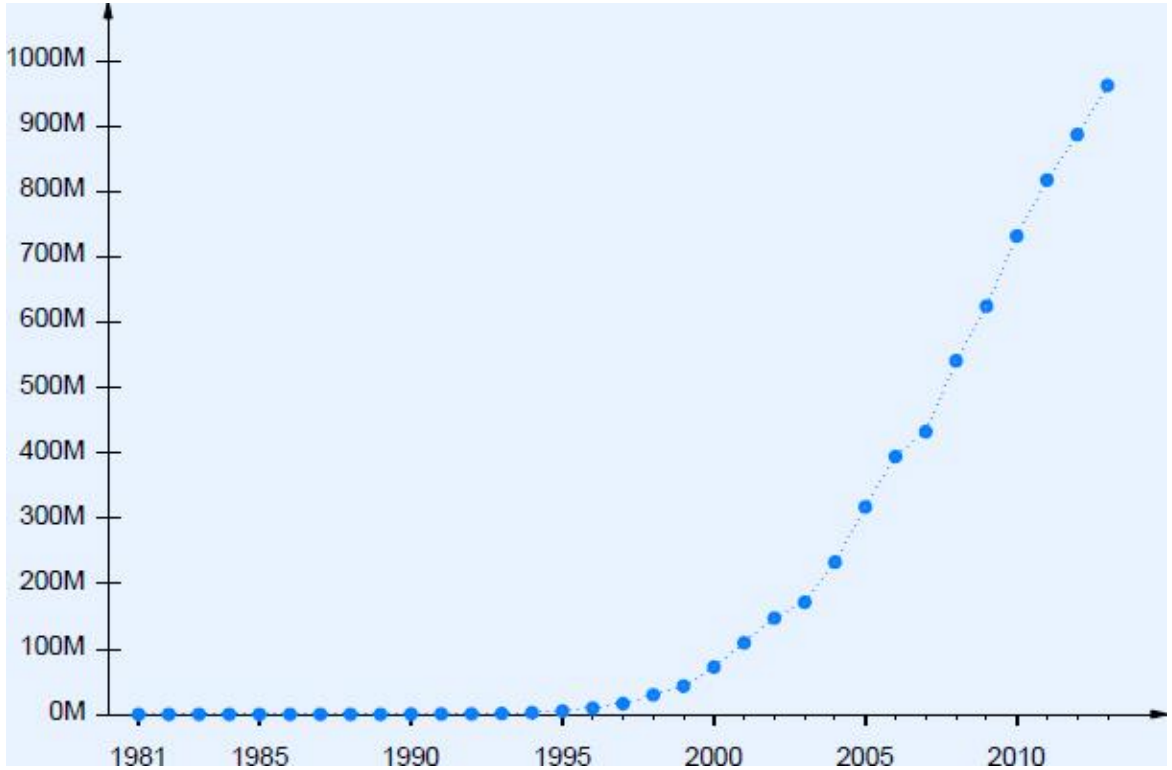





- Networking professionals see it as **infrastructure**
  - Platform on which above services run
  - A huge and complicated network and **is growing rapidly**

Size of the Internet

- Derived by walking the Domain Name System
  - Only report devices with IP addresses
- However, nowadays many devices do not have a fixed IP address
  - Connect behind a NAT box (e.g., wireless router)
- Actual size is difficult to measure



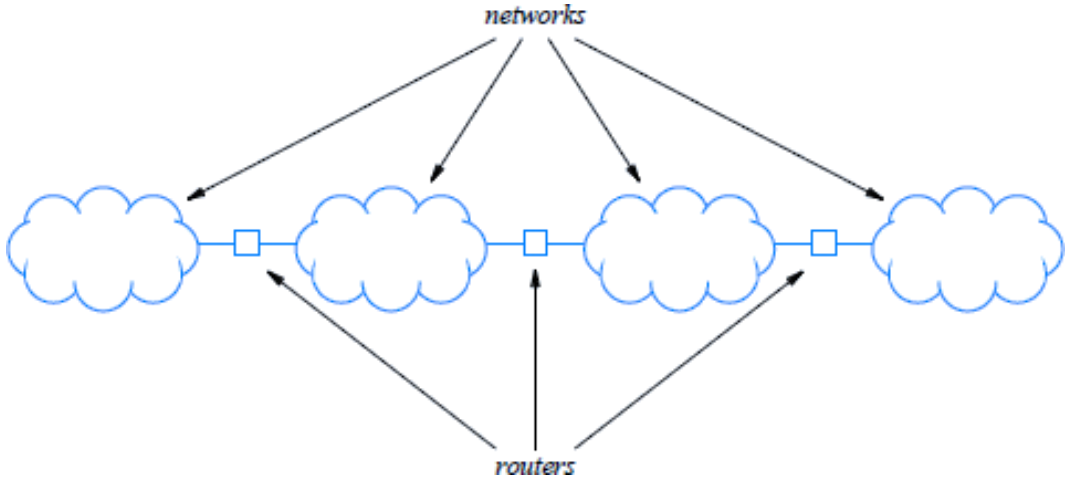
Number of computers on the Internet each year

What is the Internet?

- Internet is a network that accommodates extant services plus new services that will be invented
  - Only provides the **communication service**; does not care about the data contents
- Internet is designed to accommodate any network technology, allowing each technology to be used where appropriate
  - Does not restrict or dictate underlying network technologies

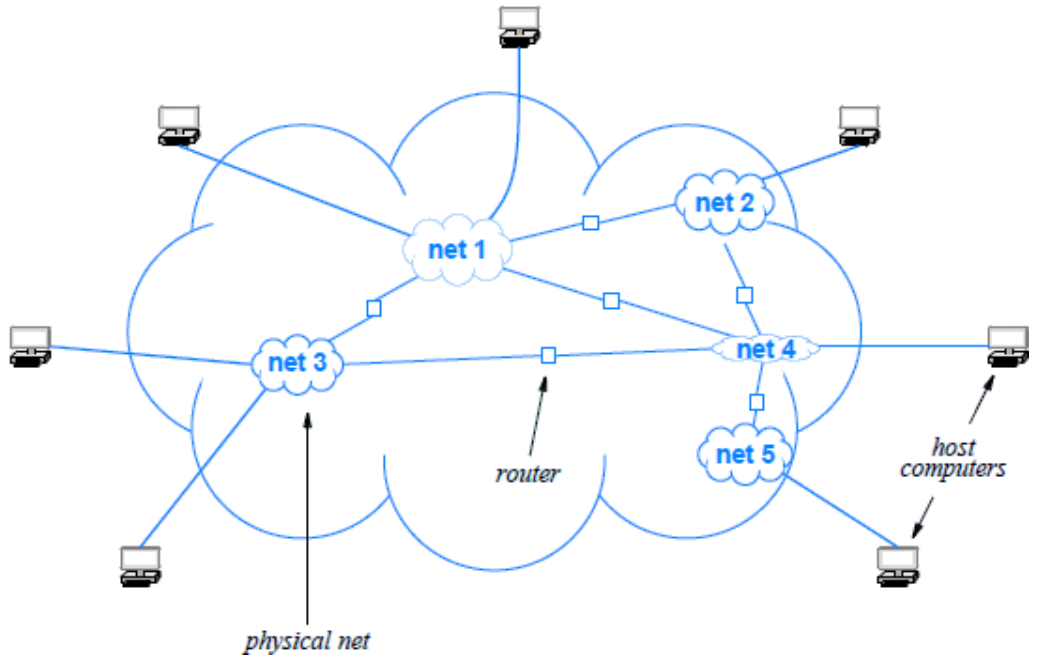
Internet Architecture

- Follows **a network of networks** approach
- Allows arbitrary networks to be included, which could also be a network of networks
- Uses **IP routers** to interconnect individual networks



Internet Architecture

- Computers attached to the Internet are known as host computers.
- Each host attaches to an access network, which in turn, attaches to a big network.
- To a host, Internet appears to be one giant network.



What is WWW?

- The World Wide Web (WWW or web) is just one of the ways information can be shared over the Internet.
- It allows documents to be linked to one another using **hypertext links**, thus forming a huge “web” of connected information.
- The invention of the WWW is usually attributed to the British Tim Berners-Lee, who, along with the Belgian Robert Cailliau, published a proposal in 1990 for a hypertext system while both were working at CERN in Switzerland.

Evolution of WWW

- The Web invented by Tim was a Read-only Web
  - It was static and somewhat mono-directional.
- Web 2.0 – Since mid-2000s
  - With reading as well as writing, the web could become bi-directional.
  - It represented the move toward a more social, collaborative, interactive and responsive web.
- Semantic Web, Artificially Intelligence Web – Web 3.0 ??

Core Features of the Web

1. A **URL** to uniquely identify a resource on the WWW.
2. At two ends, there are two programs interacting by means of the **HTTP protocol**, which is in the form of **client-server communication**.
   - The browser program which makes HTTP requests and that can display the HTML it receives.
   - The web server software program that responds to HTTP requests.
3. The HTTP protocol to describe how requests and responses operate.
4. **HTML** and **CSS** to publish documents.

URL Components

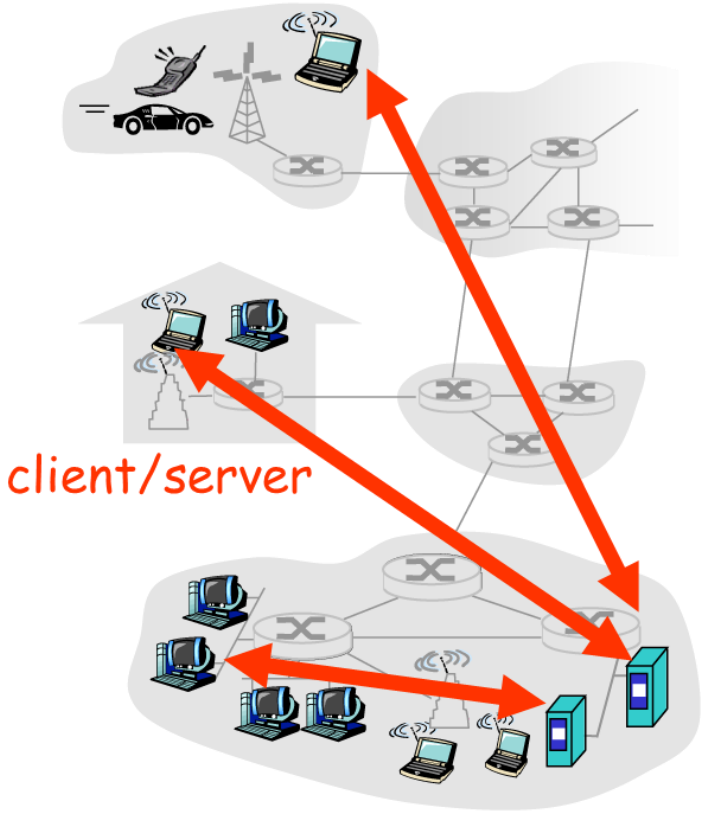
- In order to allow clients to request particular resources from the server, **a naming mechanism** is required so that the client knows **how to ask the server** for the file.
- For the web that naming mechanism is the **Uniform Resource Locator** (**URL**).

URI – Uniform Resource Identifier

- A **URI** is a compact sequence of characters that identifies an abstract or physical resource.
- **URL** is one type of URI that **identifies** the resource by its **location** on the network.
  - A URL begins with the name of the protocol to be used for accessing the resource and then contains sufficient information to point to how it can be obtained.
- The other type of URI is the **URN**, which identifies the resource by name or namespace.
  - Provides a way of uniquely naming a resource **without specifying** an access protocol or mechanism, and **without specifying** a particular location.
  - e.g. the International Standard Book Number (ISBN) for uniquely identifying a book is a URN, “urn:isbn:0451450523”

Client-Server Communication

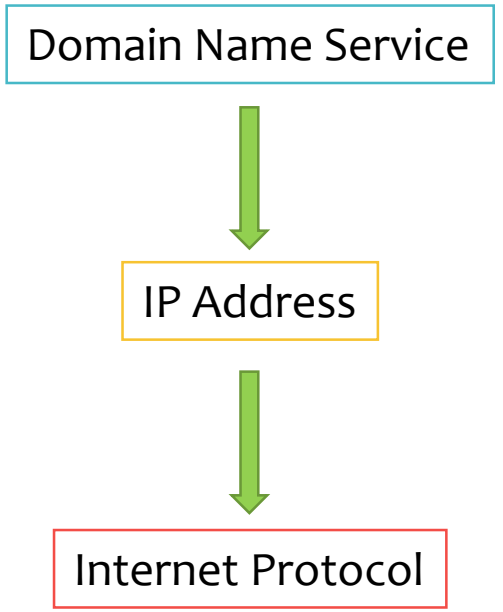
- Used by applications to establish communication
- One application program acts as a **server**
  - Always starts execution first; usually **always online**
  - Running in a host machine **known by** other peers
  - **Awaits contact** from any peer
- The other application program becomes a **client**
  - **Initiates contact** to the server
  - **Requests** for services and/or information
  - Its location is not important



- Important concept: once communication has been established, data can flow **in either direction** between a client and server

Host to Host communication

- For hosts to communicate with each other, they need:
  - Some method to identify the peer
    - For client-server communication, we need the **domain name** of the server.
  - Some method to locate the peer
    - We need the **IP address** of the server
  - Some method to deliver messages to the peer
    - We need to find the **network path** between the client and server



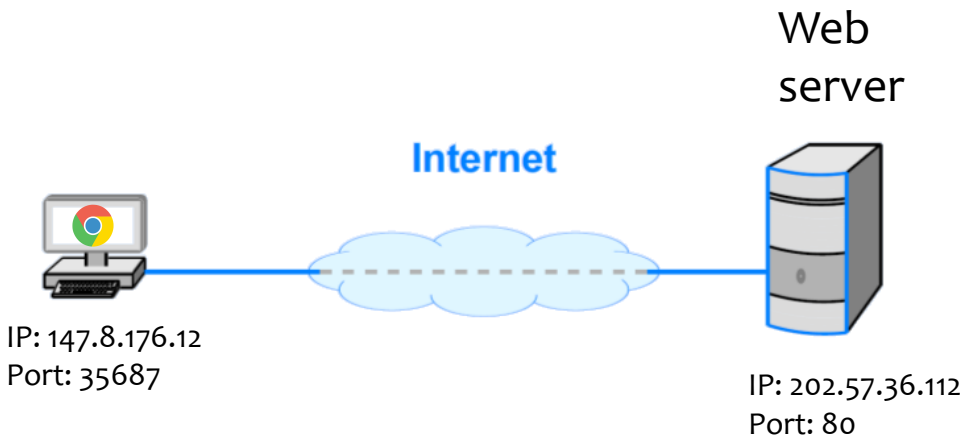
End to End Communication

- Many server applications can be running on a single host machine. How to **find** the server program?
- Each communication service is "labelled" by a **port number** P.
  - e.g., HTTP - 80, Telnet - 23, SSH - 22, FTP - 21 (20), . . .
  - This port # (service) **must be known** by all clients.
- When a server is up running, it informs the system that it is using port P.
  - **No other** application **running on the same machine** can use the **same port** P.
- By the **unique** IP address and port number, the network can locate those server programs.

Illustration of Steps Taken by Client and Server

Client Side

- Agree a priori on a port number, N
- Start after server is already running
- Obtain server name from user
- Use DNS to translate name to IP address
- Contact server using IP address and port N
- Interact with server and then exit



Server Side

- Agree a priori on a port number, N
- Start before any of the clients
- Register port N with the local system
- Wait for contact from a client
- Interact with client until client finishes
- Wait for contact from the next client…

Port Numbers

- **Valid Range**: 0 to 65535 (of size 16 bits)
- Divide into three ranges:
  - **Well-known ports**: 0 to 1023
    - Reserved ports
    - Can only be used by
      - system processes
      - privileged users
  - Registered ports: 1024 to 49151
    - Available for ordinary user processes
    - Also allow to register with **IANA** (Internet Assigned Number Authority)
  - Dynamic and/or Private ports: 49152 to 65535

IP Address

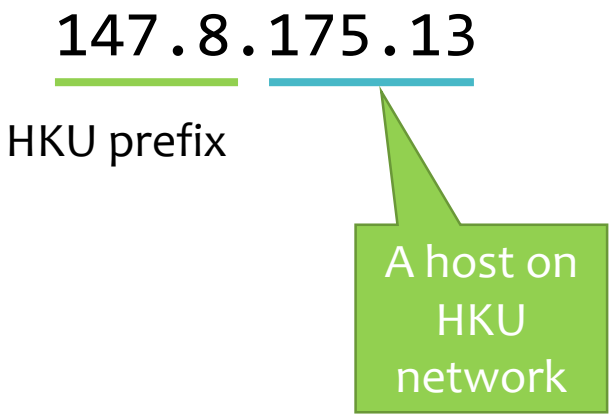
- Every host connected to the Internet runs the Internet Protocol and is assigned with an IP address, which is a **numeric code** that serves two purposes:
  - **Identity**
    - **Uniquely** identify **a network interface** of a host machine on the Internet.
  - **Locator**
    - Has information to **locate the physical network** in which the host is attached.

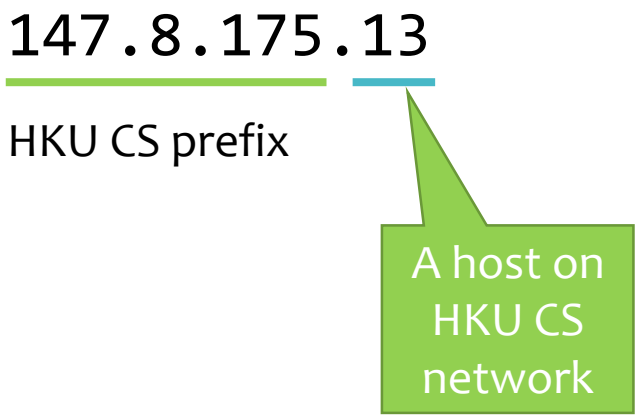
IPv4 Addressing

- It is a **32-bit** identifier and is assigned to a network interface of a host/router.
- Interfaces – network ports/cards
  - Routers have multiple ports
    - that means multiple interfaces
  - Host machines can have multiple interfaces too, e.g., Ethernet and WiFi
- Typically presented in Dotted-Decimal Notation
  - Divides IPv4 address into **octets** of eight bits each
    - E.g., IP address of `11011111 00000001 00000001 00000001` is `233.1.1.1` in dotted-decimal notation

IPv4 Addressing Scheme

- Technically, the 32-bit IP address is divided into two parts
  - **Prefix part** which identifies **physical network** (**locator**)
  - **Suffix part** which identifies **a host** on that physical network (**identity**)





IPv6 Addressing

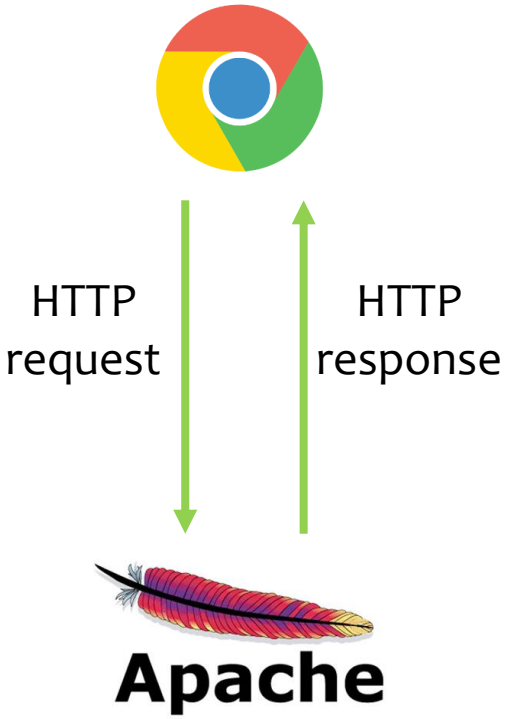
- All centrally allocated IPv4 addresses have been **exhausted** since 2011
  - IPv4 is not dead yet
  - We are in the process of migration from IPv4 to IPv6
- A dramatically large address space: from 32-bit to **128-bit**
  - there are 3.4×10^{38} (340,282,366,920,938,463,374,607,431,768,211,456) addresses
- Express as **8 groups of 16-bit** (4 hex digits) with colons between groups
  - FE80 : 0000 : 0000 : 0000 : 0001 : 0800 : 23E7 : F5DB
- Many IPv6 addresses contain long strings of zeroes. To simplify it, successive zeros can be replaced by two colons
  - FE80 :: 0001 : 0800 : 23E7 : F5DB or even to
  - FE80 :: 1 : 800 : 23E7 : F5DB

“localhost”

- In computer networking, localhost is a hostname which means **this computer**.
- It is used for accessing the network services that are running **locally**.
  - No network traffic is generated to the network.
- The name localhost normally resolves to:
  - IPv4 loopback address **127.0.0.1**
  - IPv6 loopback address **::1**
- For example, a locally installed website may be accessed from a web browser by the URL http://localhost or http://127.0.0.1 to display its home page.

Protocols

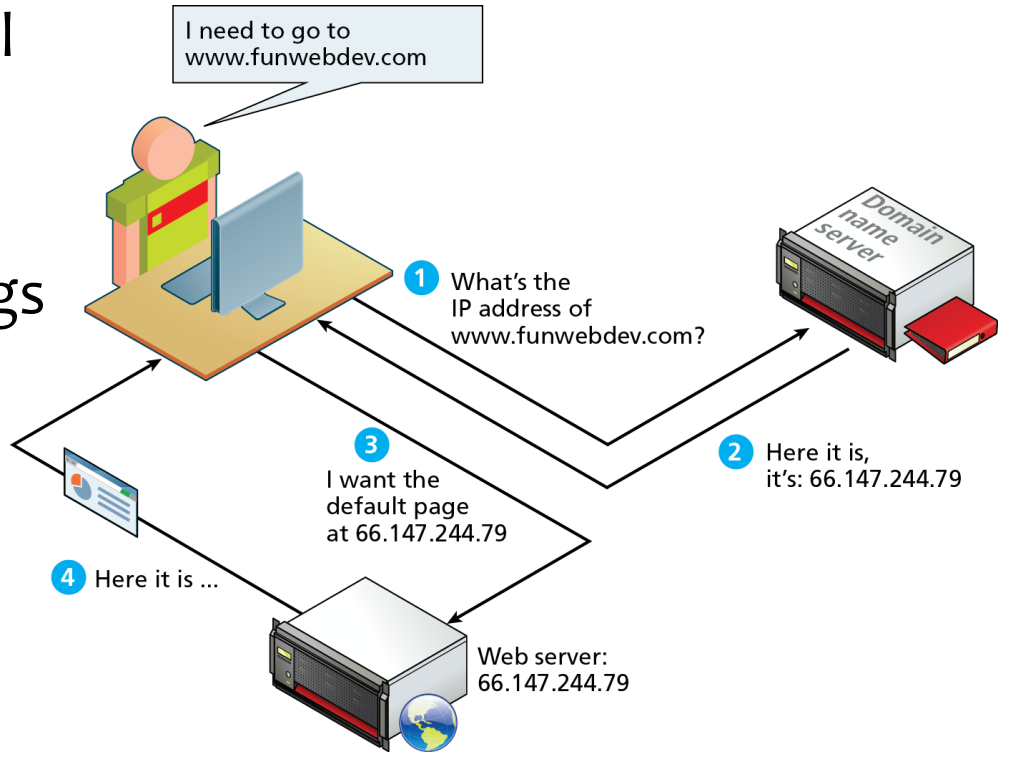
- The internet exists today because of a suite of interrelated communications protocols.
  - E.g., Ethernet, WiFi, ARP, ICMP, DHCP, NAT, RIP, IP, TCP, UDP, SCTP, DNS, HTTP, SMTP, FTP, SSH, telnet, Skype, Whatsapp, . . .
- A **protocol** is a set of rules that define how the partners in the communication are interacting.
- A communication protocol defines:
  - The **format** of the message.
    - e.g., where is the control information? which part carries the contents?
  - The **order** of the messages sent and received between two ends.
  - The **actions** taken on message transmission and receipt.



Domain Name System

Domain Name System (D N S)

- A network service which is **critical** to the normal operation of the Internet.
- Runs at the **application layer**.
- As elegant as IP addresses may be, human beings do not enjoy having to recall long strings of numbers.
- DNS **translates** human-readable **domain names** into the **binary addresses** used by the Internet Protocol.
- Example
  - Computer www.cs.hku.hk
  - Has the IP address 147.8.179.15



Services Provided by DNS

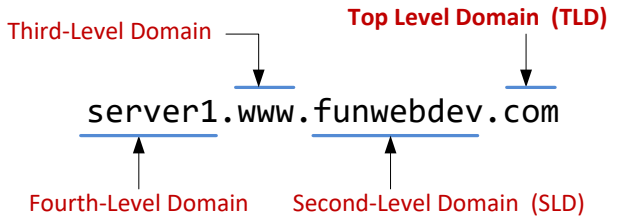
- Translates a given **domain name** to **IP address**.
- Reverse looks up the **domain name** from an **IP address**.
- Looks up the information of the **Email server** and **DNS server** of a **domain**.
- Load distribution
  - Maps one **domain name** to **a set of IP addresses**, e.g., replicated Web servers
    - Different users may select different machine (of different IP address) for requesting services

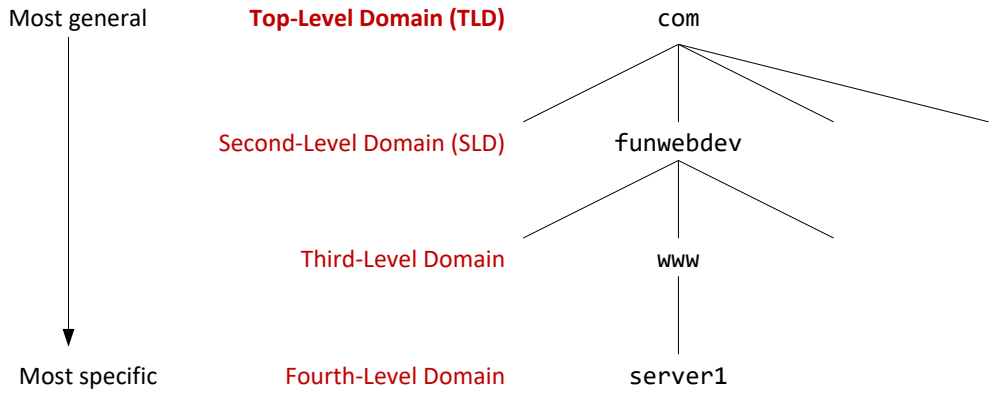
Address Resolution

- Taken place **every time** the browser tries to have the **first** contact with the Web server given by the URL.
- Where is the DNS server? How can it know all the translation mappings of the whole Internet? And how can it get the updates?
- Ideally, the DNS server should be **near to** its client applications.
- No single DNS server manages all the mapping info of the Internet.
- **Solution:** The use of a **distributed hierarchical database**.

Domain Names and Domain Levels

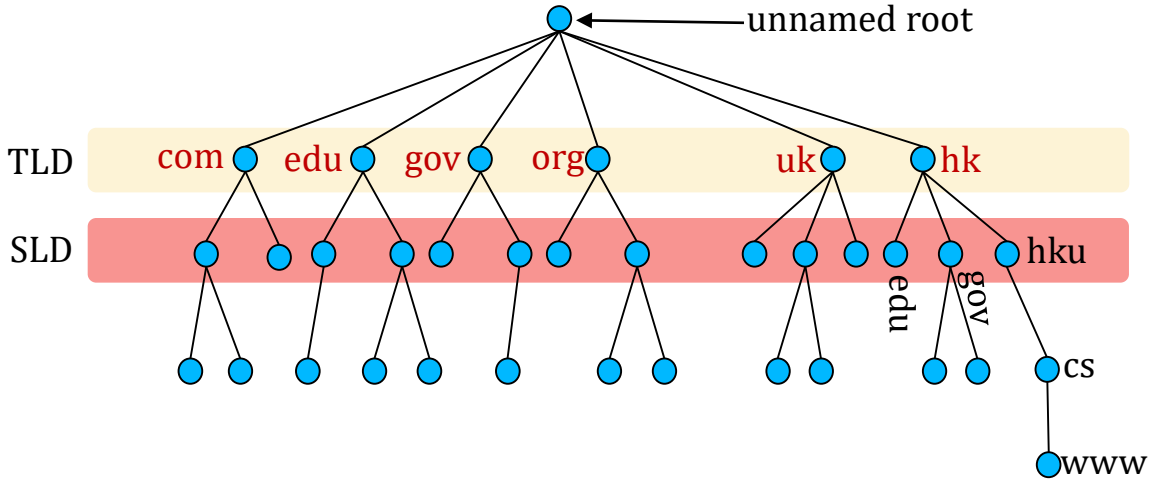
- Domain names are **hierarchical**.
- Domain name is a **sequence of labels** with dots in between, and it defines a realm of administrative autonomy of Internet hosts.
- Most specific label is on the left and most significant label is on the right.
- Rightmost label known as a **top-level domain** (**TLD**)



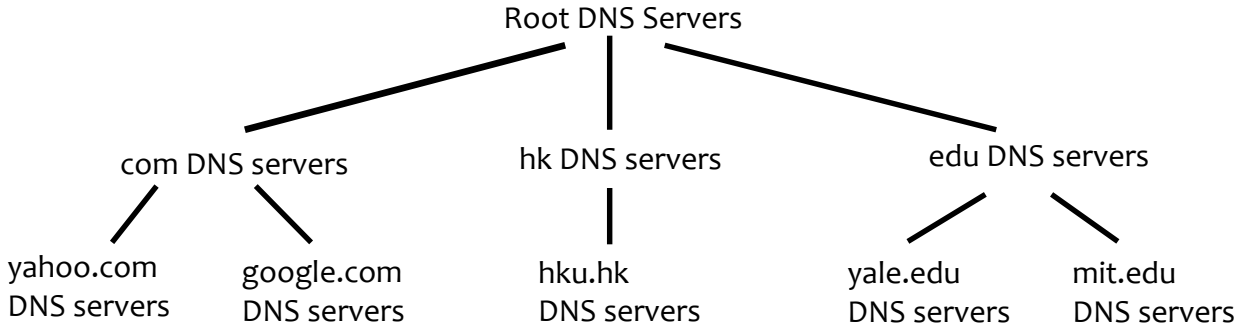


Domain Name Tree

- All TLDs are linked to the DNS root node, which does not have a name.



- The concept of distributed hierarchical database

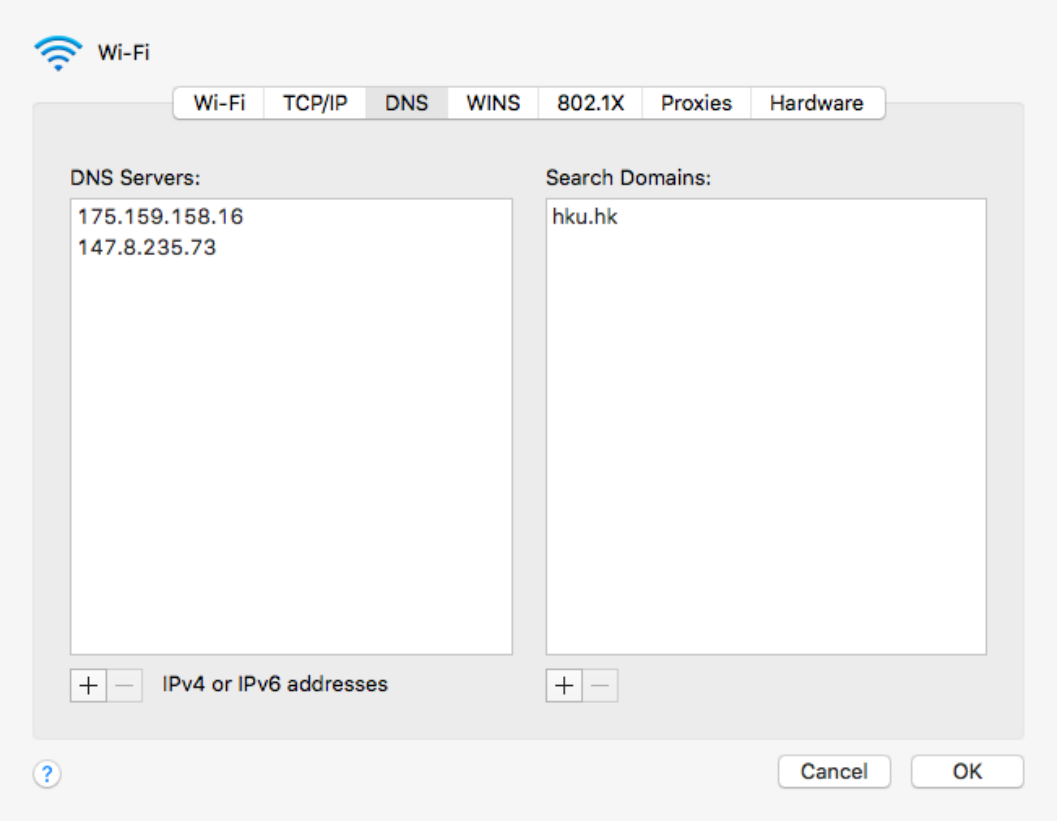


Distributed Hierarchical Database

- There are four levels of DNS servers:
  1. Root Name Servers
  2. TLD Servers
  3. Authoritative DNS Servers
  4. Local Name Servers (Default name servers)

Local Name Servers

- Each domain (i.e., ISP, company, university, . . .) has to provide **one or more local DNS** server for supporting internal requests.
- All client applications (in DNS term, we called them the **resolvers**) are configured with the address of the local DNS server(s).



Local Name Servers

- All client applications **first contact** the local DNS server to lookup the mappings.
- If local DNS server has the information,
  - it returns the answers to the resolvers immediately.
- If it does not have the information,
  - it sends the requests **to the root** name server.

Authoritative DNS servers

- Each organization should have at least one authoritative DNS server if they offer publicly accessible hosts.
- These servers provide authoritative hostname to IP mappings (and vice versa) for organization's computers (e.g., Web, mail).
  - They can serve as local name servers too.
- Example: Only CS department has the correct and up2date information of the mapping of our CS departmental servers/machines (e.g., ns1.cs.hku.hk)

TLD Servers

- Each top-level domain has its own set of name servers that are responsible for providing information under its domain subtree.
  - Examples: .hk subtree, .com subtree, .org subtree
- They know the authoritative servers of a particular subdomain in its subtree.
  - Example: .hk TLD servers has the info about who are the authoritative DNS servers of .hku.hk

Types of TLDs

Generic top-level domains (gTLD)

Domain Name	Assigned To
aero	Air transport industry
arpa	Infrastructure domain
biz	Businesses
com	Commercial organizations
edu	Educational institutions
gov	United States Government
info	Information
jobs	Human resource managers
mil	United States military
mobi	Mobile content providers
museum	Museums
net	Major network support centers
org	Non-commercial organizations
travel	Travel and tourism

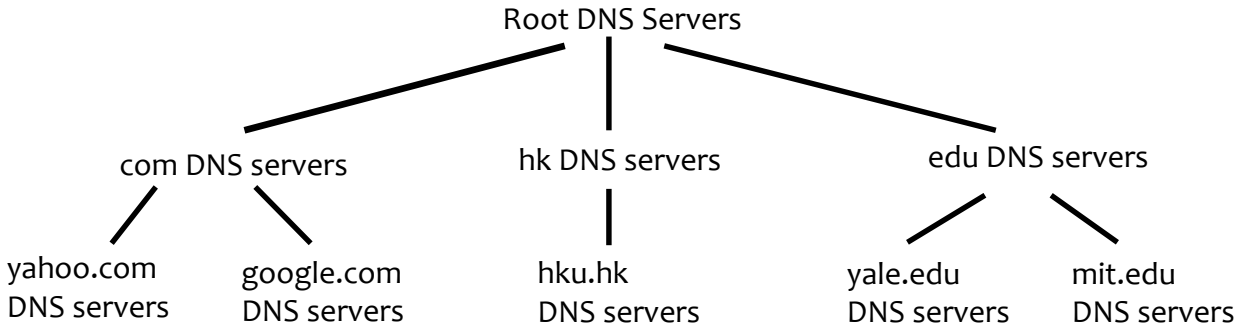
Country code top-level domain (ccTLD)

Domain Name	Assigned To
country code	A sovereign nation

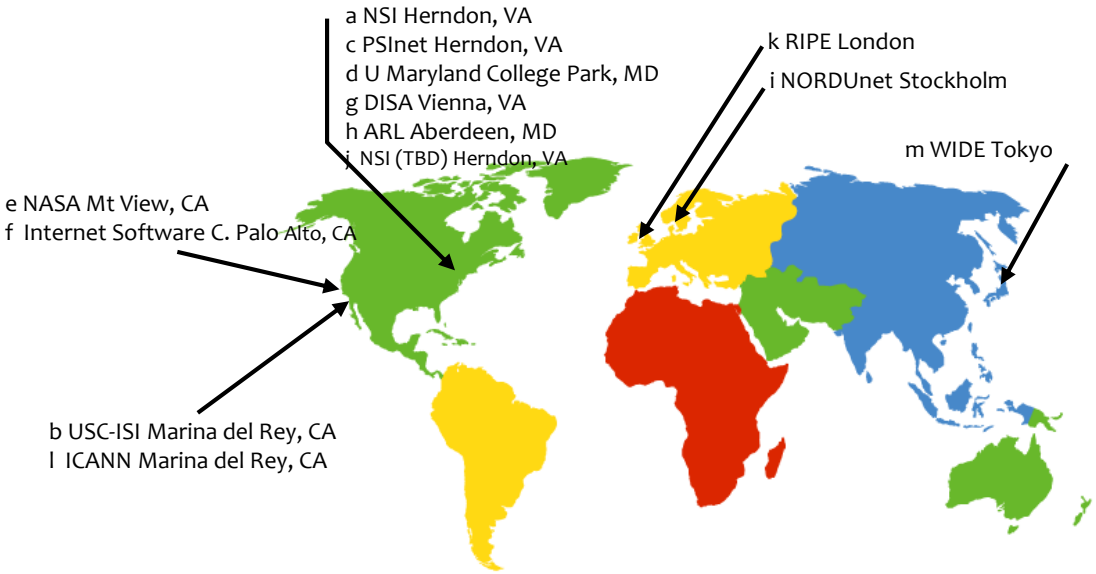
In 2014, ICANN decided to allow many new TLDs

Root Name Servers

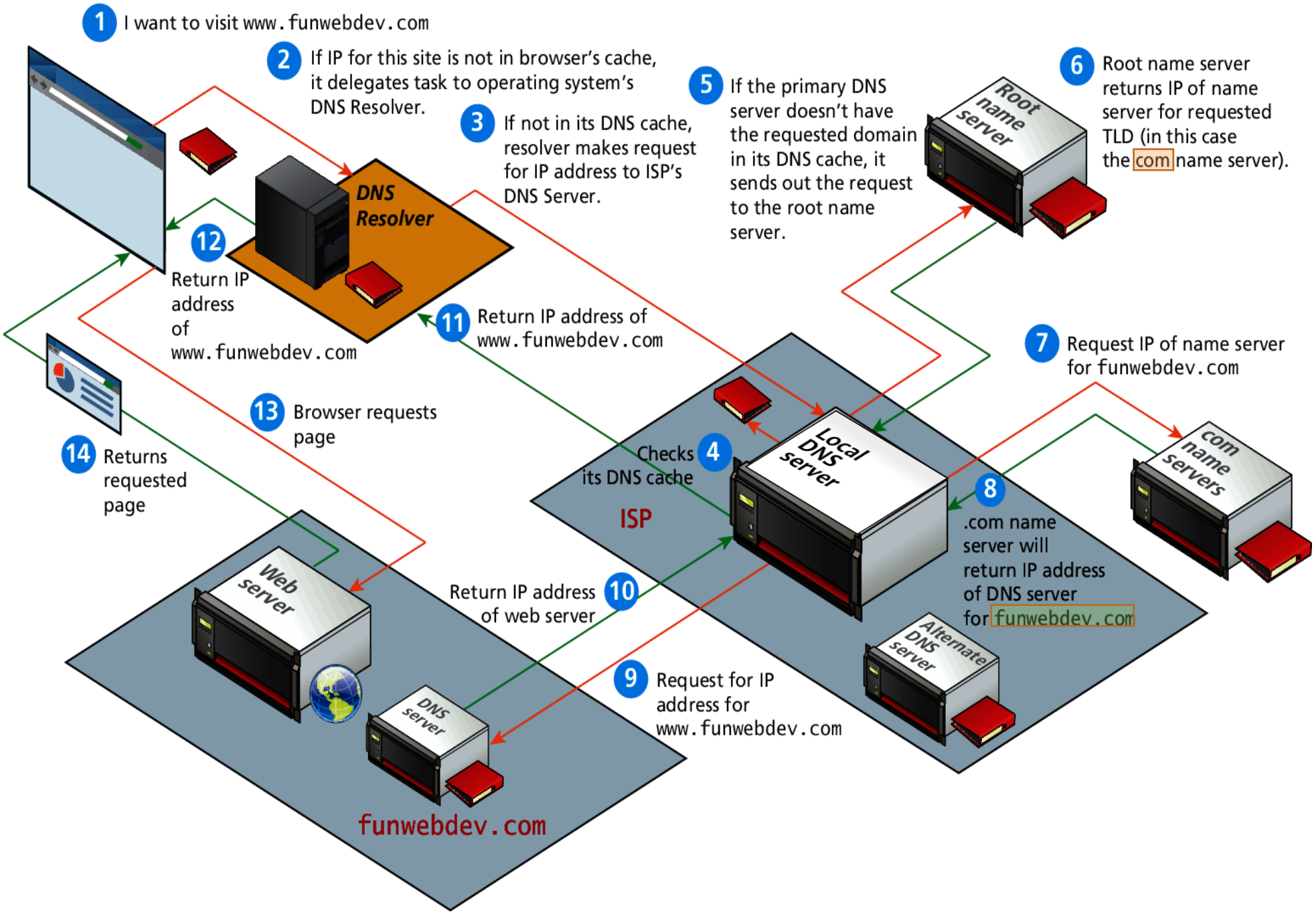
- Have information on **all TLD** name servers.
- **Known** (publicly) by *all local name servers*.
- The root servers response DNS queries from local name servers with the correct TLD name servers.



- There are 13 root name servers worldwide.



Domain name address resolution process



Reading

- MDN web docs
  - How does the Internet work?
    - https://developer.mozilla.org/en-US/docs/Learn/Common_questions/How_does_the_Internet_work
  - How the Web works?
    - https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works
  - What is a domain name?
    - https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_domain_name

Other References

- About DNS
  - http://www.tcpipguide.com/free/t_TCPIPDomainNameSystemDNS.htm
- An Introduction to Computer Networks
  - Chapter 1, An Overview of Networks
    - http://intronetworks.cs.luc.edu/current/html/intro.html