1.2 IP addresses, domain names, and URLs

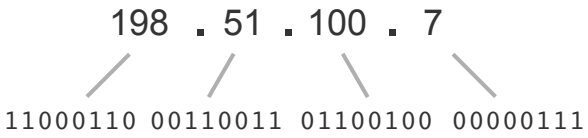
IP addresses

A computer communicates with another computer on the internet by sending packets back and forth. An internet ***packet*** contains To and From IP addresses, the information to communicate, and other configuration information.

An ***IP address*** (short for ***Internet Protocol*** address) is a computer's unique address on the internet (like a house's unique address in the world), usually represented numerically like 198.51.100.7. A typical IP address is 32 bits, divided into four 8-bit groups, each group often written as a decimal number.

PARTICIPATION ACTIVITY 1.2.1: IP addresses.

1 2 3 2x speed



0 = 00000000

255 = 11111111

The smallest possible number in an IP address is 0, and the largest is 255.

Captions

1. An IP address is usually written as four numbers separated by dots.
2. Each number represents 8 bits of the IP address, for a total of 32 bits.
3. The smallest possible number in an IP address is 0, and the largest is 255.

Feedback?

The original Internet Protocol, known as ***IPv4***, has 32-bit addresses. 32 bits can represent 2^{32} or about 4 billion unique addresses, originally believed to be more than enough, but 4 billion is no longer enough. A new version of the Internet Protocol, ***IPv6***, uses 128-bit addresses, capable of representing 2^{128} addresses. The number 2^{128} is 3.4×10^{38} or 340,000,000,000,000,000,000,000,000,000,000,000, which is hopefully enough for quite a while. IPv4 and IPv6 currently co-exist and likely will for a long time.

PARTICIPATION ACTIVITY 1.2.2: IP addresses.

1) What is the size of an IP address in IPv4?
   - (•) 32 bits
   - ( ) 128 bits

Correct

IPv4 provides 4 billion unique IP addresses. In 2010, 10 billion devices were connected to the internet with an estimate of 50 billion by 2020. IPv4 did not provide a sufficient number of unique addresses. (Source: Wall Street Journal)

2) IPv6 supports about how many IP addresses?
   - ( ) 100 billion
   - (•) 3.4×10^{38}

Correct

3.4×10^{38} is enormous, and hopefully enough to address increasing numbers of internet devices for a long time.

3) With the introduction of IPv6, devices on the internet no longer use IPv4.
   - ( ) True
   - (•) False

Correct

IPv4 and IPv6 are not interoperable, but mechanisms exist to allow communication between the two protocols.

Feedback?

Domain names and DNS

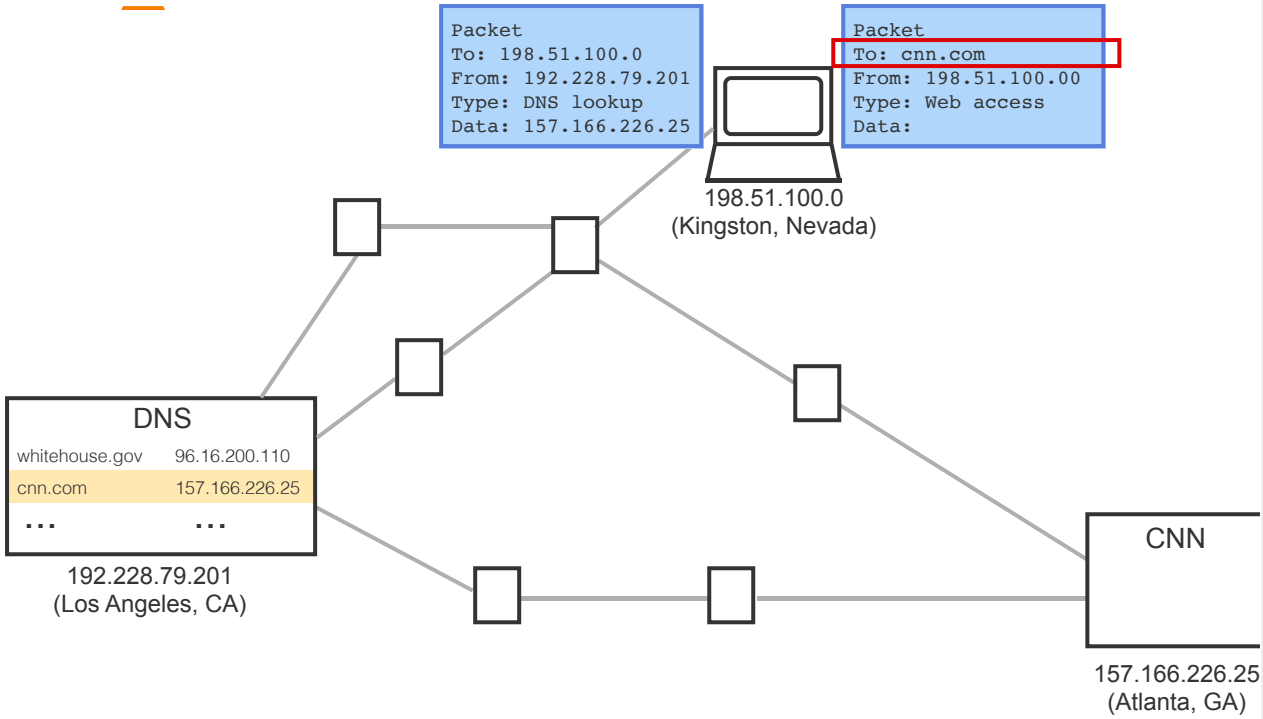
Some websites can be directly reached using the computer system's IP address. Ex: Google could say "Go to 216.58.193.206 to search the internet." But those numbers are hard to remember, and IP addresses can change, so domain names are commonly used. A ***domain name*** is a name for an IP address, such as the name wikipedia.org for the IP address 198.35.26.96; the name is easier to remember and type. Capitalization doesn't matter: Wikipedia.org, wikipedia.org, and WIKIPEDIA.ORG are treated the same.

When a computer sends a packet using a domain name over the internet, the first step is to contact a ***DNS server*** to convert the domain name to an IP address. ***DNS*** is short for Domain Name System.

Thirteen main DNS servers (called ***root servers***) exist in the world, and a computer's operating system or an ISP keeps a reference to the root servers' IP addresses. Ex: 198.41.0.4 (run by Verisign), 192.228.79.201 (run by USC), 199.7.91.13 (run by Univ. of Maryland), and 192.203.230.10 (run by NASA). The first step of sending an internet packet to a domain name is thus to lookup the IP address via a DNS server.

PARTICIPATION ACTIVITY 1.2.3: A DNS server first looks up a domain name's IP address.

1 2 3 2x speed



First, a packet is sent to a DNS server to lookup the IP address for cnn.com.

Captions

1. A computer wants to send a packet to cnn.com.
2. First, a packet is sent to a DNS server to lookup the IP address for cnn.com.
3. The packet contains the destination address. The computer can now communicate with CNN.

Feedback?

Registering a domain name

*Anyone may register an unused domain name with a **domain name registrar**. Most registrars charge a yearly fee for keeping the domain registered. Registration information is made publicly available from ICANN's Whois service.*

When a website is hosted with a web hosting company, the company will update the DNS servers so the website's domain name is associated with the IP address of the hosted website. The web hosting company may also manage the yearly registration of the domain name.

PARTICIPATION ACTIVITY 1.2.4: IP addresses and domain names.

1) Websites have *either* an IP address or a domain name, but not both.
   - ○ True
   - ◉ False

   Correct
   All websites have an IP address. A domain name provides an easy-to-remember name for an IP address.

2) cnn.com is a(n) _____.
   - ○ IP address
   - ◉ domain name
   - ○ DNS

   Correct
   A domain name provides an easy-to-remember name for a user to reach a website.

3) A DNS server's primary role is to convert _____.
   - ○ an IP address into 32 bits
   - ○ an IP address to a domain name
   - ◉ the domain name to an IP address

   Correct
   Given a domain name, a DNS server returns an IP address.

Feedback?

Try 1.2.1: IP addresses and domain names.

Several sites will lookup the IP address of a domain, such as the following.

- Click to open www.site24x7.com. Type "Wikipedia.com" or your favorite website name.

Given an IP address, some sites will indicate the geographic location of that IP address' computer. Try entering 171.67.215.200, which is Stanford.com's IP address, whose computers are in Palo Alto, California.

- Click to open www.iplocation.net.

Typing "What is my IP?" in a search engine can be used to determine one's IP address.

- Click to open Google.com.

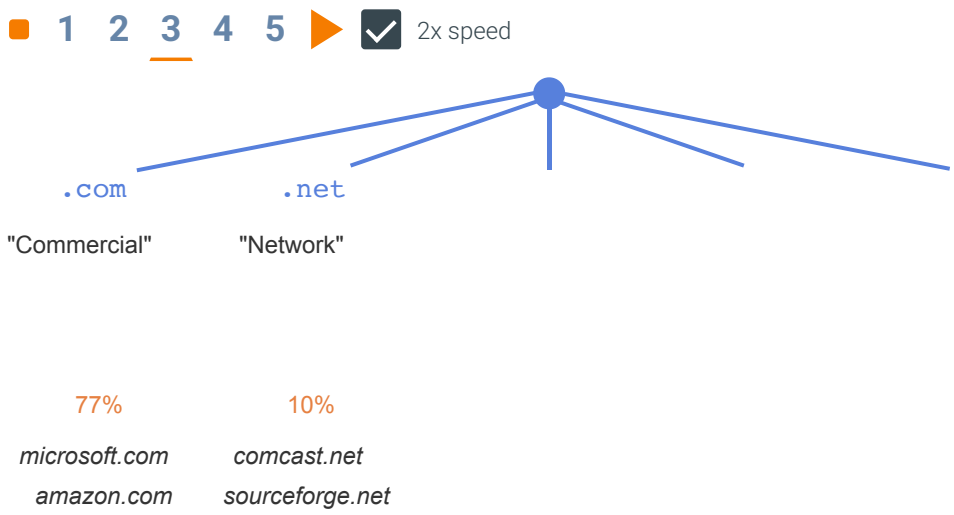
- Click to open Bing.com.

Feedback?

Domain name levels

Domain names are hierarchical. A domain name belongs to one of numerous ***top-level domains*** (***TLD***), such as .com, .net, .org, .edu, and .gov. Also, each country is assigned a unique two-letter ***country code top-level domain*** (***ccTLD***) like .uk (United Kingdom), .ru (Russia), and .de (Germany). IANA, the organization that manages TLDs, allows companies and organizations to create customized TLDs, like .church, .pizza, and .music.

PARTICIPATION ACTIVITY 1.2.5: Common top-level domains.



Next most popular is .net. Typically for networking/internet-focused companies.

Captions

1. Internet domain names are divided by top-level domains.
2. The most popular is .com. Most companies have a .com name.
3. Next most popular is .net. Typically for networking/internet-focused companies.
4. Next is .org, traditionally for non-profit organizations.
5. Other well-known top-level domains are .edu for colleges, and .gov for U.S. governments. Hundreds more exist.

Feedback?

Immediately after a top-level domain comes a ***second-level domain***, such as wikipedia in wikipedia.org. A second-level domain is commonly an organization's name as in Stanford.edu, or indicates the purpose of a website as in DoPython.org. Third-level and

further level domains refer to sub-computer systems local to an organization, as in cs.stanford.edu where the cs is for Stanford's Computer Science department. A common third-or-deeper-level domain is www, short for World Wide Web, usually referring to an organization's web server. Many organizations use www optionally, so stanford.edu by default goes to www.stanford.edu.

PARTICIPATION ACTIVITY 1.2.6: Domain hierarchy.

Given a website such as www.whitehouse.gov, match the domain to the corresponding domain-level hierarchy.

If unable to drag and drop, refresh the page.

.gov	Top-level domain. .gov typically indicates a U.S. federal, state, or local government webpage.	Correct
www	Third-level domain www refers to the organization's web server.	Correct
whitehouse	Second-level domain. whitehouse refers to the name of the organization.	Correct

Reset

Feedback?

Note

The top-down structure of a domain name is, somewhat counterintuitively, from right-to-left rather than left-to-right. Thus, cs.stanford.edu has a top-level domain of .edu, second-level stanford, and third-level cs. Common scams use this counterintuitive right-to-left feature to trick people into thinking a site

belongs to a particular company. Ex: bankofamerica.xxyz.com might trick a person into thinking the domain name is for Bank of America.

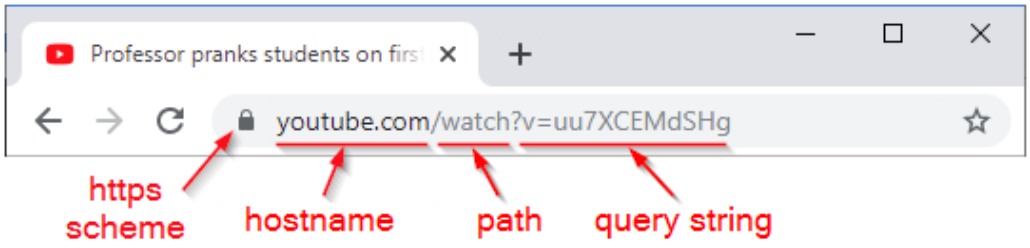
URLs

Domain names are most commonly seen in URLs. A ***URL*** (***Uniform Resource Locator***) is the location of a web resource on the web, such as http://www.cdc.gov/alcohol/faqs.htm. A ***web resource*** is any retrievable item, like an HTML file, image, video, CSS stylesheet, etc.

A URL is composed of several parts:

- ***Scheme*** - Characters at the beginning of a URL followed by a colon ":" or a colon and double slashes "://". Common URL schemes include `http`, `https`, `mailto`, and `file`. Ex: In https://www.cdc.gov/alcohol, the scheme is "https".
- ***Hostname*** - The complete domain name following the scheme in a URL. Ex: In https://www.cdc.gov/alcohol, the hostname is "www.cdc.gov".
- ***Path*** - The characters to the right of the hostname in a URL. Ex: In https://www.cdc.gov/alcohol, the path is "/alcohol".
- ***Query string*** - Optional characters to the right of the question mark (?) in a URL that provide data for the web server. In https://www.youtube.com/watch?v=uu7XCEMdSHg, the characters after the ? tells YouTube's server to play a video having code uu7XCEMdSHg.
- ***Fragment*** - Optional characters at the end of a URL that start with a hash character (#) and refer to a certain location within a webpage. In https://en.wikipedia.org/wiki/URL#History, the fragment "#History" refers to the webpage's History section.

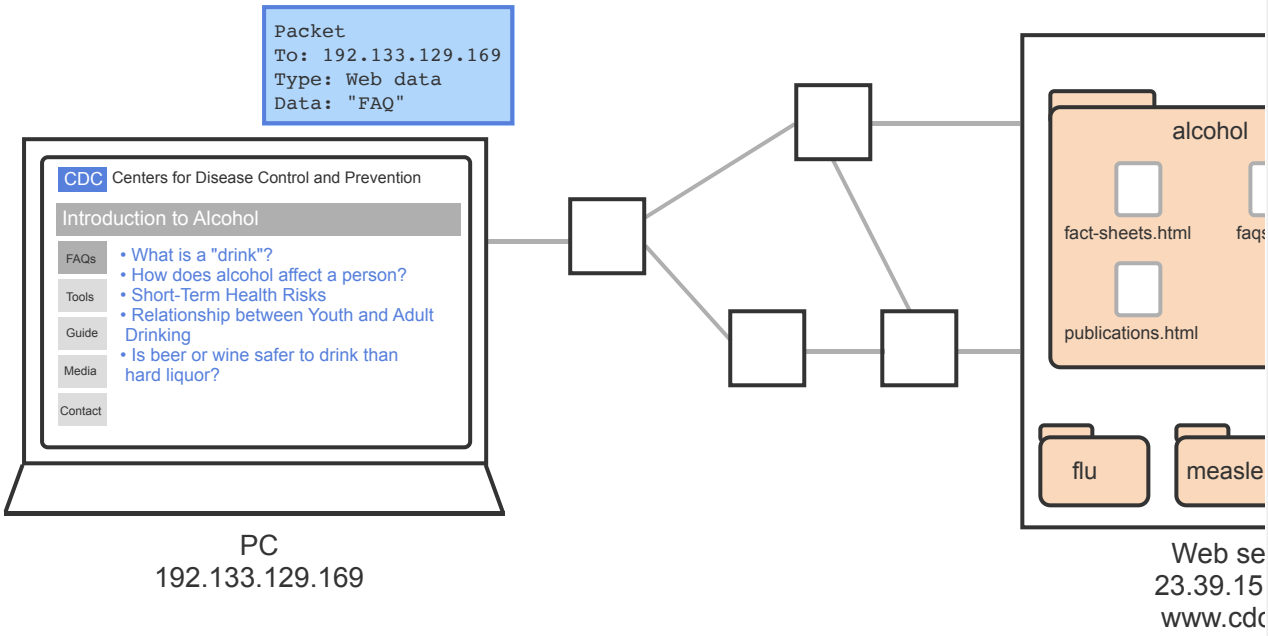
Figure 1.2.1: Chrome browser hides the scheme and "www" in hostname.



Feedback?

PARTICIPATION ACTIVITY 1.2.7: A URL can specify the location of subfolders and files.

1 2 3 2x speed



The faqs.html file is returned and displayed in the computer's browser.

Captions

1. A computer's browser sends a packet requesting information from a web server at IP address 23.39.150.112, with hostname www.cdc.gov.
2. The information is in a subfolder named alcohol, then in a file in that folder named faqs.html.
3. The faqs.html file is returned and displayed in the computer's browser.

Feedback?

PARTICIPATION ACTIVITY 1.2.8: URLs.

1) Is the following a valid URL?

https://www.whitehouse.gov/

- (•) Yes
- ( ) No

Correct

A valid URL must have a scheme with separator, hostname, and path. When a filename is not provided in the path, the web server accesses a default file.

2) Is the following a valid URL?

Correct

https://www.nasa.gov/cubequest/details/#.VOaO_VPF9U9

- (●) Yes
- ( ) No

The URL's path refers to the cubequest and details subfolders on www.nasa.gov. The fragment #.VOaO_VPF9U9 refers to a location within the webpage. The URL may or may not produce a valid webpage since some websites are frequently reorganized, but the URL is structured correctly.

3) What is the hostname of the following URL?

http://www.weather.com/summer/temperatures.html

- (●) www.weather.com
- ( ) weather.com

Correct

The hostname is the complete domain name, which is the characters after the scheme and before the path.

4) Which file does the following URL access?

http://www.weather.com/summer/temperatures.html

- ( ) summer
- (●) temperatures.html

Correct

The subfolder is summer, and the file is temperatures.html.

5) Which scheme should a URL use to provide a secured HTTP connection to a web server?

- ( ) http
- (●) https

Correct

The https scheme encrypts web traffic, allowing users to safely submit sensitive information like credit card numbers to a website.

6) Is www.nasa.gov a valid URL?

- ( ) Yes
- (●) No

Correct

The scheme and separator "https://" is missing. Most web browsers do not require users to type the scheme and will often hide the scheme from users. A web

browser will assume the user meant https://www.nasa.gov/ and go to NASA's webpage.

Feedback?

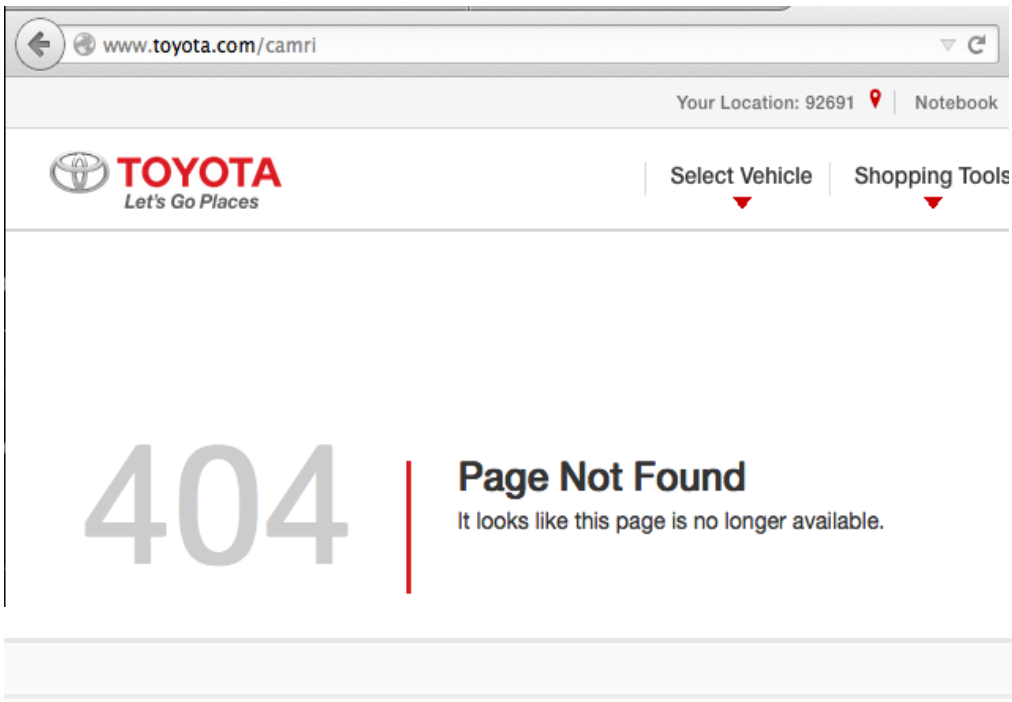
Webpage errors

If a domain name is not found by a DNS server (usually because the domain name is not registered), a page is displayed indicating such, as in "Sorry, the website www.xyz.blahblahblah cannot be found."

A domain name may be found (so a valid IP address exists for that name) but the web server may not respond, resulting in a message like "The website is not responding" or "Could not reach the website". Such non-response could be due to the web server being turned off or undergoing maintenance, or due to an essential router malfunctioning, for example.

If a web server is reached but the specific requested page isn't found, the server returns a ***404*** status code, which is a code number for page not found. Various other status codes exist. Many web servers return a page that includes the number "404".

Figure 1.2.2: An example webpage that returns a message with the code 404 shown.



Feedback?

Linkrot

The web is always changing with content being continually added and removed. When content is removed from the web, URLs that used to point to the content return a 404 status code. ***Linkrot*** *is the general name for a once valid link that now returns a 404 status code.*

The Internet Archive is one of many organizations around the world that fights linkrot by archiving the web for posterity. The Internet Archive's Wayback Machine provides access to billions of archived webpages and can often show users what a particular URL looked like at different times in history.

PARTICIPATION ACTIVITY 1.2.9: Webpage errors.

1) A 404 message indicates that a domain name is not recognized.
   - ○ True
   - ◉ False

Correct

A 404 means that the website was reached (the domain name is valid and the web server is working), but the web server could not find the requested page, often due to a misspelling or out-of-date page link.

2) If a web server is down (such as powered off), the server may return a page saying to please try again later.
   - ○ True
   - ◉ False

Correct

If a server is down, the server cannot return a page. Instead, a message may be displayed saying "Cannot reach the website" or similar.

Feedback?

Exploring further:

- URL Standard (WHATWG)
- List of Top-Level Domains (ICANN)

CHALLENGE ACTIVITY 1.2.1: IP addresses, domain names, and URLs.

530096.4000608.qx3zqy7

Jump to level 1

Which URL(s) are valid?

- [ ] (A) https:/www.nature.com/flowers/azalea.doc
- [ ] (B) www.climate.com/fall/temperatutes.pdf
- [ ] (C) http://www.nature.com/ flowers/daisy.gif
- [x] (D) http://www.weather.com/summer/wind.doc
- [ ] (E) https://www.zoo.com/ animals/penguin.png
- [ ] (F) http:/www.cities.com/chicago/cafes.htm

1 2 3 4

Check Next **Done**. Click any level to practice more. Completion is preserved.

✔ Expected: (D)

Invalid URLs:
(A) Missing one of the two double slashes "//" after the colon ":".
(B) Missing scheme and characters that follow: "http://".
(C) Has unnecessary space before "flowers/".
(E) Has unnecessary space before "animals/".
(F) Missing one of the two double slashes "//" after the colon ":".

Feedback?

How was this section? **Provide section feedback**