

# The Internet and Web

## Internet

What is **Internet**?

# The Internet and Web

## Internet

What is **Internet**?

The **Internet** (*contraction of interconnected network*) is the global system of interconnected *computer networks* that use the *Internet protocol suite (TCP/IP)* to link devices worldwide.

It is a network of networks that consists of *private, public, academic, business, and government networks* of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies.

- *from Wikipedia*

# The Internet and Web

## Internet

The **Internet** carries a vast range of information resources and services, such as the *inter-linked hypertext documents* and *applications* of the **World Wide Web (WWW)**, *electronic mail*, *telephony*, and *file sharing*.

- *from WikipediA*

# The Internet and Web

## Internet

A computer communicates via the **Internet** by sending a **packet**, containing information like:

- an *IP address* for a destination computer,
- the *data size*, and
- the *data* itself (email message, web page, piece of video, etc.).

# The Internet and Web

## Internet

A computer communicates via the **Internet** by sending a **packet**, containing information like:

- an *IP address* for a destination computer,
- the *data size*, and
- the *data* itself (email message, web page, piece of video, etc.).

Packets travel over communication links

- wires (like "cable TV" wires),
- fiber optic cables,
- wireless satellite communications

through several other computers on a path to a destination.

# The Internet and Web

## Internet Protocol

The basic low-level rules for communicating data over the Internet are called the **Internet Protocol (IP)** for short).

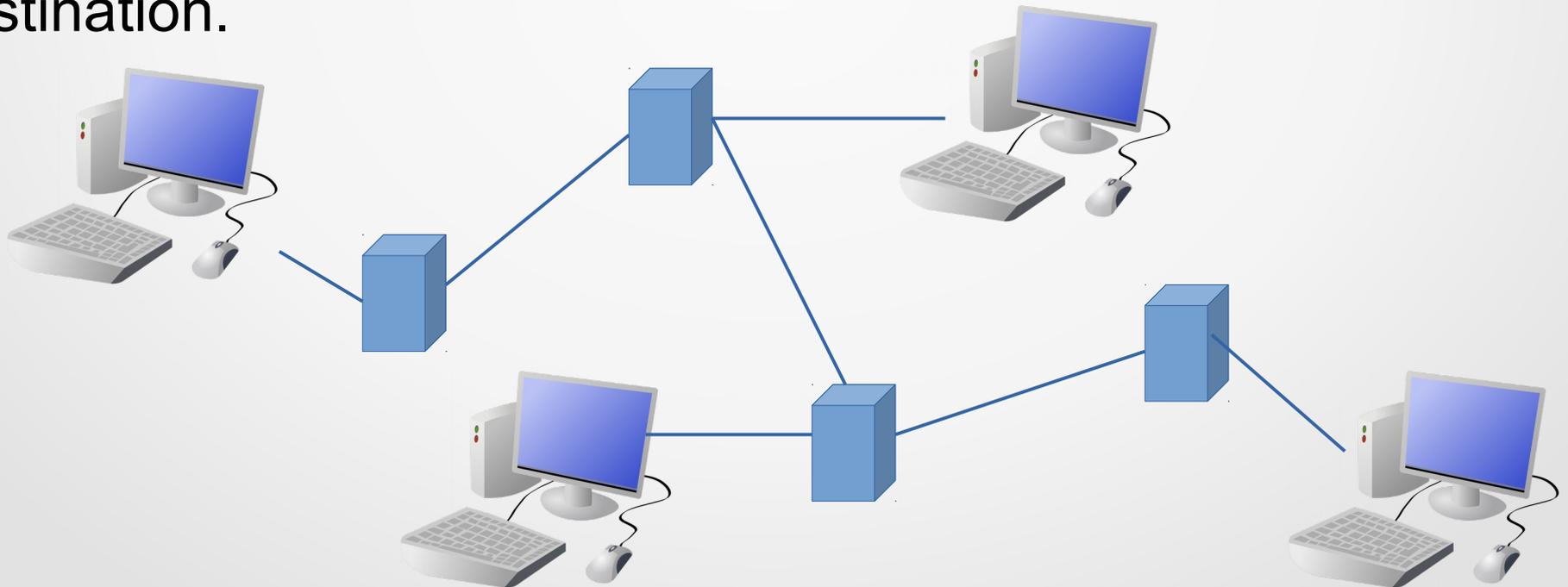
Networked computers, like humans, require protocols for communication.

**TCP/IP** stands for **Transmission Control Protocol (TCP)** and the **Internet Protocol (IP)**.

# The Internet and Web

## Routers and Internet service providers

A **router** is a specialized computer in a network that receives a packet from a computer/router and sends that packet to another computer/router on a path towards the packet's destination.



# The Internet and Web

## Routers and Internet service providers

**Internet Service Provider (ISP)** is an organization that provides services for accessing, using, or participating in the Internet.

**Internet service providers** may be organized in various forms, such as commercial, community-owned, non-profit, or otherwise privately owned.

*- from Wikipedia*

Usually the connection to the Internet is provided via satellite, cable lines, wi-fi, fiber optics, digital phone lines, etc.

# The Internet and Web

## IP addresses

An **Internet Protocol (IP)** address is a unique numerical address on the Internet of a connected device (desktop, laptop, smart phone, tablet, ...).

**Internet Protocol version 4 (IPv4)** defines an **IP** address as a **32-bit** number and looks like this one: **198.168.1.15**

*- 4 groups, 4 bytes each; written as decimal number*

**Internet Protocol version 4 (IPv6)** , developed in 1995, and standardized in 1998, final definition of the protocol was published in 2017, uses **128 bits** for the IP address. Looks like this one:

**2204:2000:1001:f1d1:e4a1:7763:6d1d:2193**

*- 8 groups, 4 bytes each, written as a hexadecimal number*

# The Internet and Web

## IPv4 vs IPv6

IPv4 uses 32-bits or 4 groups, 4 bytes each  
(4 groups × 8 bits = 32 bites )

How many different IP addresses can this provide?

0 1 0 0 \_ \_ ... \_ 1 1 (each place can be 0 or 1)

32 places , hence  $2 \times 2 \times 2 \times \dots \times 2 = 2^{32} = 4,294,967,296$

IPv6 uses 128 bits

How many different IP addresses can this provide?

0 1 0 0 \_ \_ ... \_ 1 1 (each place can be 0 or 1)

128 places , hence  $2 \times 2 \times 2 \times \dots \times 2 = 2^{128} =$

340,282,366,920,938,463,463,374,607,431,768,211,456

# The Internet and Web

## IP addresses

To find your local network IP address, do:

- 1) run `cmd` (right click on Windows icon → Run )
- 2) type  
> `ipconfig`

To find your global IP address:

Visit <https://whatismyipaddress.com/>

**Warning:** multiple adds, do not allow to memorize your location (there will be a pop-up window asking for it)

# The Internet and Web

## Static and dynamic IPs

Some computers have a **static IP address**, meaning the address doesn't change.

### Example:

whitehouse.gov's IP address is always 96.16.200.110.

Some computers have a **dynamic IP address**, which changes frequently.

A PC, laptop, smartphone, etc., that connects to the Internet via WiFi or a modem is typically assigned a new dynamic IP address each time the computer is turned on or is newly connected to the Internet.

# The Internet and Web

## Domain names and DNS servers

Although all websites can be reached via IP address, it is much more convenient to use names, that are called **domain names**.

### Example:

www.google.com

www.nytimes.com

www.bcc.cuny.edu

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

# The Internet and Web

In-class activity, see items 1 - 3 in the handout

# The Internet and Web

## Home Networking and Cellular Networks

read about them in ZyBooks (sections 7.3 – 7.4)

# The Internet and Web

## The web

Computers of the 1940s and 1950s were mostly solitary machines.

In the 1960s computer networks evolved to allow computers to communicate information with each other.

In the Internet's early days, much communication was for *email messages*, *discussion boards*, and *file transfers*.

**File transfer protocol (FTP)** was an early way for transferring files (word document, image file, ...). A user would login to another computer, and then type commands to copy specific files among the computers.

# The Internet and Web

## The web

In the early 1990s, a European researcher developed a more convenient way for computers to communicate files, known as [the web](#), involving three things:

- Text files, known as [HTML files](#), containing links to other text files.
- A program, known as a [browser](#), for viewing HTML files.
- A set of rules, known as the [HTTP protocol](#), for transferring HTML files among computers.

# The Internet and Web

## HTML files and HTTP

**HTML** stands for **hyper-text markup language**;

hyper-text means the text has links to other text (and today to images, videos, and more).

**HTTP** stands for **hyper-text transfer protocol**.

# The Internet and Web

## World Wide Web

The **web** was originally called the **World Wide Web**, hence the **www** in many website names.

The **web** was developed around 1989-1991 by **Tim Berners-Lee**, working at a Swiss research institute named CERN, to help physicists from around the world share data. He originally considered the name "**Information Mesh**."

Today, most but not all **Internet** traffic is **web** traffic.

**Note:** **Internet** is the interconnection of computers communicating using a set of rules. **Web** is just one particular use of the Internet.

# The Internet and Web

## Web search engines

As of January 2019, there are 1,518,207,412 websites in the World.

<https://www.millforbusiness.com/how-many-websites-are-there/>

A **web search** seeks web pages that best match user-typed words.

A **search engine** is a website that does web searches.

In 2018, popular search engines included [google.com](https://www.google.com), [bing.com](https://www.bing.com), [yahoo.com](https://www.yahoo.com), and [ask.com](https://www.ask.com).

For more, visit

<https://www.reliablesoft.net/top-10-search-engines-in-the-world/>

# The Internet and Web

## Search matching rules

For particular search words such as 'What are the rules of american football?', a search engine may find millions of matching pages. A search engine determines the "best" matching pages using web-search matching criteria like:

- How often the main words, like 'rules', 'american', and 'football', occur on a page.

Singular/plural doesn't matter, and common filler words like 'what', 'are', 'the', and 'of' are ignored.

- How high up the main words appear on a page.
- How often adjacent search words, like 'american football', also appear as adjacent on the page.
- How recently the page was published.

# The Internet and Web

## Search matching rules

For particular search words such as 'What are the rules of american football?', a search engine may find millions of matching pages. A search engine determines the "best" matching pages using web-search matching criteria like:

- How many pictures and videos the page contains.
- How geographically related is the page to the searcher's location.
- How related is the page to previous searches that the searcher has recently conducted.
- How "important" is the page itself, based on how many other pages have links to that page, or how often that page or its parent website is visited.

# The Internet and Web

## Search matching rules

Search engines differ in part based on what matching criteria the engine uses and how the engine combines those criteria.

Search engine companies keep their criteria a secret to reduce people "gaming the system" by artificially moving pages higher on the matching list.

# The Internet and Web

## Search advertising

Search is big business.

Companies pay to have search-specific advertisements appear near search results.

Such ads are effective because people doing searches are often ready to pay for something, as when searching for "online college math courses".

# The Internet and Web

## Default search engine

Usually every Internet browser has a **default search engine** which is used for search, so you can type in a search words directly in the browser's search box or URL box, rather than by going to a search engine site like google.com.

Most browsers have a financial arrangement with a search engine company to which the browser steers users.

# The Internet and Web

## Default search engine

### Examples:

- The **Chrome** browser uses **google.com** as search engine (both are made by Google).
- The **Internet Explorer** browser **bing.com** search engine (both are made by Microsoft)
- The **Firefox** browser makes deals with specific search engines. In 2015, Firefox made **yahoo.com** the default search engine (previously google.com was the default). Estimated revenue from that deal is over \$100 million.
- The **Safari** browser uses **google.com** as search engine

A user can change the browser's default search engine via the browser's settings/preferences.

# The Internet and Web

## Web crawlers

Search results must be returned within seconds.

Accessing millions of pages on the web during a search would take hours.

For speed, **search engines** pre-visit web pages and create an index of words pointing to specific pages.

**Example:** the word "cake" in the index may point to specific pages that are good matches. The word "chocolate" may point to other specific pages. Upon a search for "chocolate cake", those already-known pages are quickly combined into new search results.

# The Internet and Web

## Web crawlers

A **web crawler** is a program that pre-visits web pages to build an index.

It goes to known web pages, and then follows those pages' links to other pages, following those pages' links, etc., thus "**crawling**" around the web like a spider crawls around a spider web, updating the index information.

Crawlers work non-stop building and updating these indices and do contribute to Internet traffic.

# The Internet and Web

## Web search tips

read section 7.7 in zyBooks

# The Internet and Web

## Domain names and URLs

Let's recall that a **domain name** is a name for an **IP address**, such as the name **wikipedia.org** for the IP address **208.80.154.224** or **2620::861:ed1a:0:0:0:1**

The name is easier to remember and type.

Capitalization doesn't matter:

**Wikipedia.org**,

**wikipedia.org**, and

**WIKIPEDIA.ORG** are treated the same.

# The Internet and Web

## Domain names and URLs

Domain names and IP addresses are managed by ICANN (Internet Corporation for Assigned Names and Numbers), a non-profit organization formed in 1998 and under contract from the U.S. Dept. of Commerce.

A domain name registrar is a company authorized to reserve domain names.

In 2015, the largest domain name registrars included GoDaddy.com, Enom.com, and NetworkSolutions.com.

A typical domain name costs about \$10 per year.

# The Internet and Web

## Top level domains

Domain names are hierarchical.

A domain name belongs to one of numerous top-level domains, such as **.com**, **.net**, **.org**, **.edu**, and **.gov**.

Because many popular domains are already reserved, numerous additional top-level domains have evolved in recent years, such as **.biz** (for "business"), **.info**, **.co** (similar to **.com**), etc.

Also, each country is assigned a unique two-letter top-level domain, such as **.uk** (United Kingdom), **.ru** (Russia), **.de** (Germany), **.us** (for USA).

# The Internet and Web

## Second-level domain

Immediately after a top-level domain comes a second-level domain.

**Example:**  wikipedia.org

A second-level domain is commonly an organization's name.

# The Internet and Web

## Third-level and further domains

Third-level and further level domains refer to sub-computer systems local to an organization

**Example:**  `cs . stanford . edu`

the `cs` stands 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`.

# The Internet and Web

## Domain names and scams

Most top-level domains are not restricted.

**Example:** `.org` is typically for non-profit organizations, however any person or company can reserve a `.org` name.

`.edu` and `.gov` domain names are restricted; proof of being a college or a U.S. government entity is required.

# The Internet and Web

## Domain names and scams

Most top-level domains are not restricted.

**Example:** [.org](#) is typically for non-profit organizations, however any person or company can reserve a [.org](#) name.

[.edu](#) and [.gov](#) domain names are restricted; proof of being a college or a U.S. government entity is required.

**Example:** you get an e-mail from [IRS.gov.a.b.org](#) with a message saying that you owe \$5,000 and unless you pay it off immediately by following a link provided in the e-mail, there will be a case filed against you. Would you follow that link?

# The Internet and Web

## Domain names and scams

Most top-level domains are not restricted.

**Example:** [.org](#) is typically for non-profit organizations, however any person or company can reserve a [.org](#) name.

[.edu](#) and [.gov](#) domain names are restricted; proof of being a college or a U.S. government entity is required.

**Example:** you get an e-mail from [IRS.gov.a.b.org](#) with a message saying that you owe \$5,000 and unless you pay it off immediately by following a link provided in the e-mail, there will be a case filed against you. Would you follow that link? **No!**

# The Internet and Web

## Domain names and scams

**Cybersquatting** is the practice of registering a domain name in the hopes of selling later for a profit, like registering **coke.com** before the Coca-Cola Company had done so.

Numerous countries have passed laws to prevent **cybersquatting**.

**Example:** A U.S. law prohibits registering a trademarked name or a confusingly similar name with the bad faith intent of profit.

# The Internet and Web

## URLs

Domain names are most commonly seen in URLs.

A URL (Uniform Resource Locator) is the location of a file on the web, such as <http://www.cdc.gov/alcohol/faqs.htm>.

That URL first refers to the web server at domain [www.cdc.gov](http://www.cdc.gov), then a subfolder named [alcohol](http://www.cdc.gov/alcohol), and then a file in that folder named [faqs.htm](http://www.cdc.gov/alcohol/faqs.htm).

For a URL without a filename, like <http://www.cdc.gov> or <http://www.cdc.gov/alcohol>, the web server accesses a default file, like <http://www.cdc.gov/index.htm>. Some URLs have special characters, as in <https://www.youtube.com/watch?v=w7YCEMdSHg>

# The Internet and Web

## URLs

The [http](#) part of the [URL](#) refers to [hyper-text transfer protocol](#) used for communicating web pages.

Another, [https](#) ("s" for "[secure](#)"), is similar to [http](#) but uses encryption for secure transfers such as passwords or credit card numbers.

For a [URL](#) without a filename, like [http://www.cdc.gov](#) or [http://www.cdc.gov/alcohol](#), the web server accesses a default file, like [http://www.cdc.gov/index.htm](#). Some URLs have special characters, as in [https://www.youtube.com/watch?v=uu7XCEMdSHg](#).

# The Internet and Web

## URLs

URLs sometimes have special characters for giving the web server more information.

**Example:** <https://www.youtube.com/watch?v=uu7XCEMdSHg>

the ? tells YouTube's server to play a video having code uu7XCEMdSHg (technically, the part after the ? provides a "parameter")

# The Internet and Web

## HTML

**HTML** (**hyper-text markup language**) is a textual language for creating web pages.

**HTML** files are usually saved with a **.html** or **.htm** file extension, as in **index.html**.

An **HTML** file contains normal text surrounded by **tags** that indicate formatting, links, or other items.

A web browser like Chrome, Firefox, Safari, or IE reads an **HTML** file and displays the page.

# The Internet and Web

## HTML

```
<!DOCTYPE html>
<html>
<head>
  <title>My Awesome Page</title>
</head>
<body>
  <h1> This is a Header </h1>
  <p> Here is a paragraph, with <i>some italicized words</i>
    and <b>some bolded words</b> too.
  </p>
</body>
</html>
```

### **This is a Header**

Here is a paragraph, with *some italicized words* and **some bolded words** too.

# The Internet and Web

## HTML

**HTML formatting** involves an opening tag like `<b>` and a matching closing tag like `</b>`, with text in between.

Tags	Purpose	Example HTML
h1, h2, h3	headers. h1 is largest.	<code>&lt;h2&gt; Weather today &lt;/h2&gt;</code>
p	paragraph	<code>&lt;p&gt; 32 F, windy &lt;/p&gt;</code>
b, i, u	bold, italic, underline	<code>&lt;p&gt; This is &lt;u&gt;underlined&lt;/u&gt; and &lt;i&gt;italicized&lt;/i&gt;. &lt;/p&gt;</code>
sub, sup	Subscript and superscript	<code>&lt;p&gt; March 6&lt;sup&gt;th&lt;/sup&gt;, but x&lt;sub&gt;2&lt;/sub&gt;. &lt;/p&gt;</code>

# The Internet and Web

## HTML

A [link](#) on a web page is text that can be clicked to jump to another web page.

```
<a href = "http://www.natna.info/CSI11">CSI 11</a>
```

"a" stands for "anchor"

# The Internet and Web

## HTML with CSS

If we wish to add some styling to our web-page, then **CSS** (**Cascading Style Sheets**) can help us with it.

**CSS** is a language that can be used to style a web page via changes in colors, sizes, spacing, fonts, and more.

Styling can occur within style tags in the HTML file's **header** part. Each style rule indicates the **element** to be styled.

# The Internet and Web

## HTML with CSS

```
<!DOCTYPE html>
<html>
<head>

</head>
<body>
  <h1> This is a Header </h1>
  <p> Here is a paragraph</p>
</body>
</html>
```

```
<!DOCTYPE html>
<html>
<head>
  <style>
    h1 { color : blue; }
    p  { color : green; }
    body { background-color : lightgray; }
  </style>
</head>
<h1>This is a Header</h1>
<p>Here is a paragraph.</p>
<body>
</html>
```

**This is a Header**

Here is a paragraph

**This is a Header**

Here is a paragraph.

# The Internet and Web

## CSS properties

```
<style>
  h1    {
        color      : blue;
        font-style  : italic;
        font-size   : 50px;
        font-family : Helvetica;
        }
  body  {
        background-color : lightgray;
        margin           : 60px;
        }
</style>
```

# The Internet and Web

## CSS properties: color

A **CSS color** can be a pre-defined name as in `color:blue`,  
or  
an **rgb** value as in `color:rgb(50, 100, 255)`.

**rgb** is short for **red, green, blue**

all colors can be formed by combining those three colors.  
Values range from 0 (none) to 255 (bright).

zyBooks has a nice table in Section 7.10 showing the basic colors.

# The Internet and Web

Proceed to the in-class activity, items 4-5 in the handout