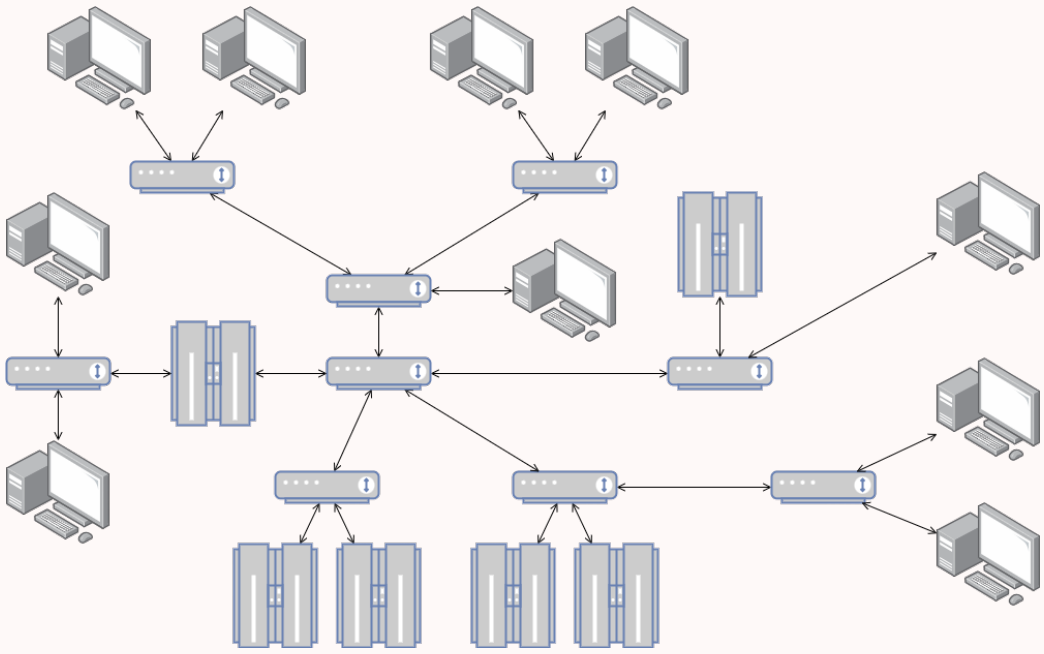
The Web

André Restivo

Web vs Internet

The Internet

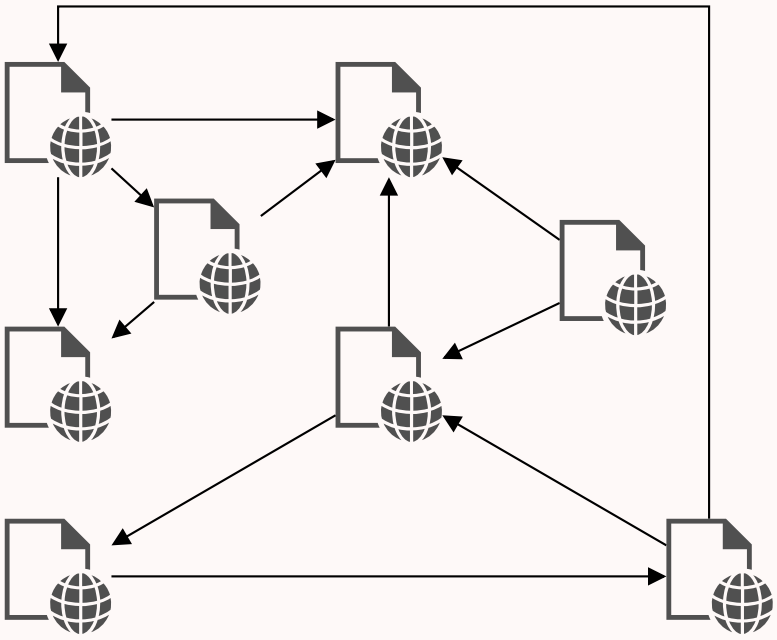
A global system of interconnected **computer networks** using the standard *Internet Protocol Suite* to link several **billion** devices worldwide.



The Web

A system of interlinked **hypertext documents** accessed via the **Internet** using a **Browser**.

Also known as the **World Wide Web** or **WWW**.



Web Origins

The Origins of the WWW

Tim Berners-Lee invented the WWW at CERN (1989).

Three constituents: **HTML** + **URL** + **HTTP**:

- **URL** is an notation for locating resources on servers.
- **HTTP** is a high-level protocol for file transfers.
- **HTML** is an SGML language for hypertext.

World Wide Web Consortium (W3C)

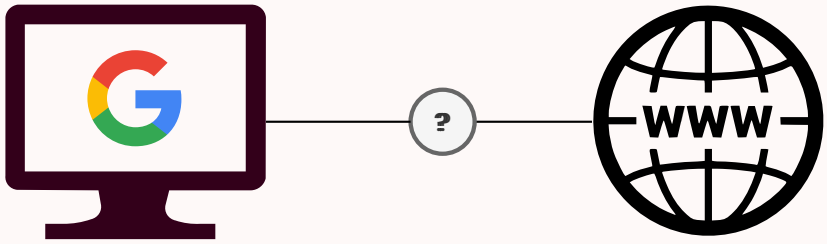
- Develops HTML, CSS, and most Web technologies.
- Founded in 1994.
- Has 380 companies and organizations as members.
- Is directed by Tim Berners-Lee.
- Located at MIT (US), Inria (France), Keiko (Japan).
- http://www.w3.org/

How does it work

From the browser to the server and back

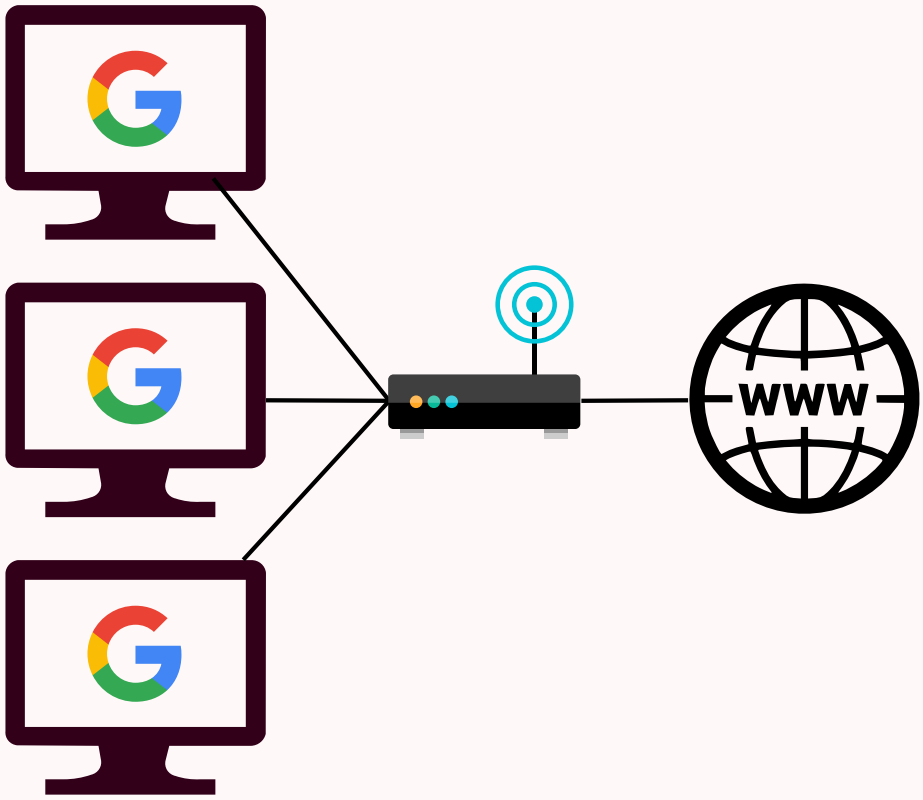
How does the web work?

What happens when you type http://www.google.com/ in the address bar of your browser?



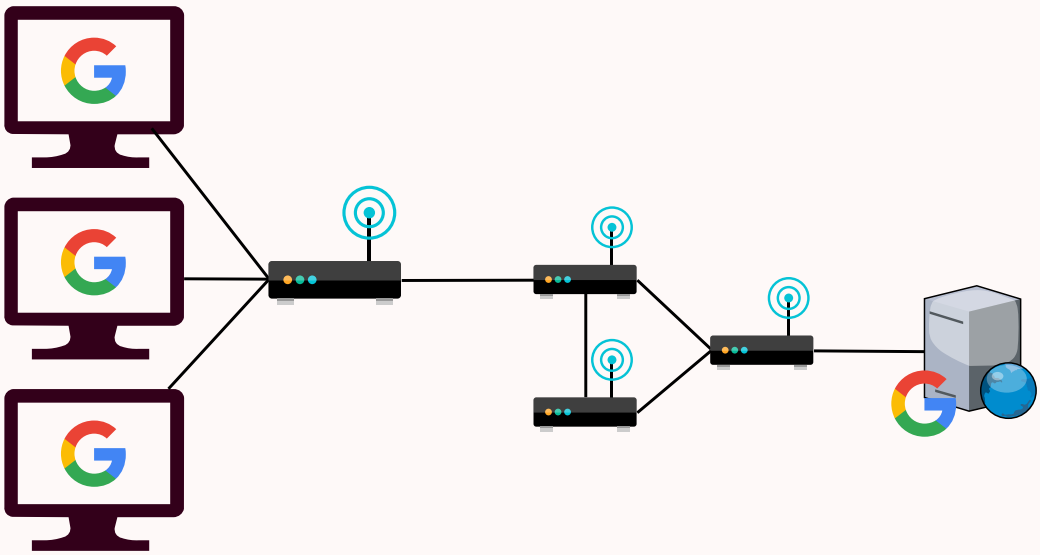
Routers

Computers are usually connected using other devices (such as routers).



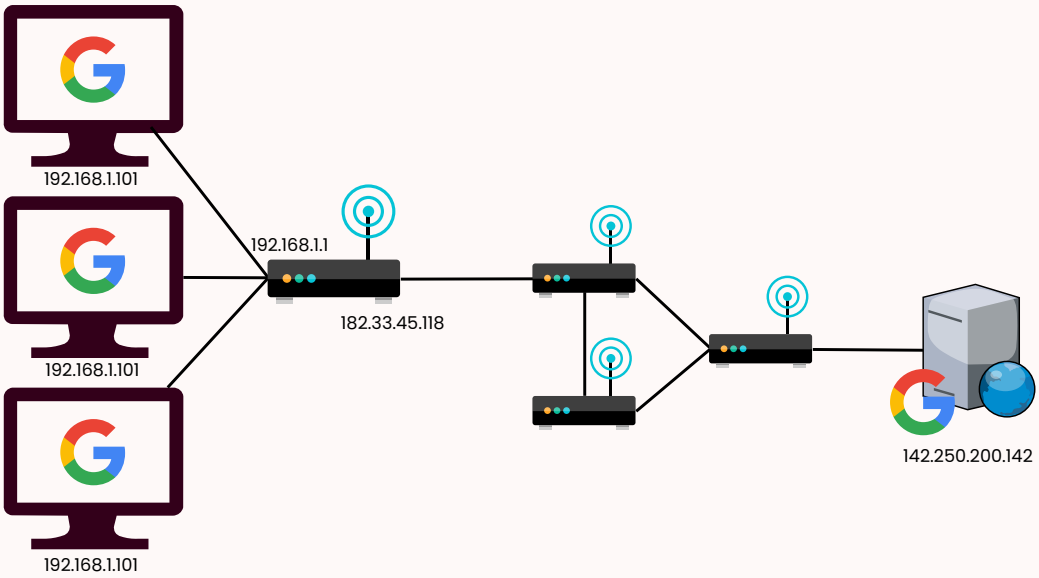
Internet Infrastructure

The Internet is a **redundant** "*network of networks*" that connects millions of hardware **devices** from laptops to servers.



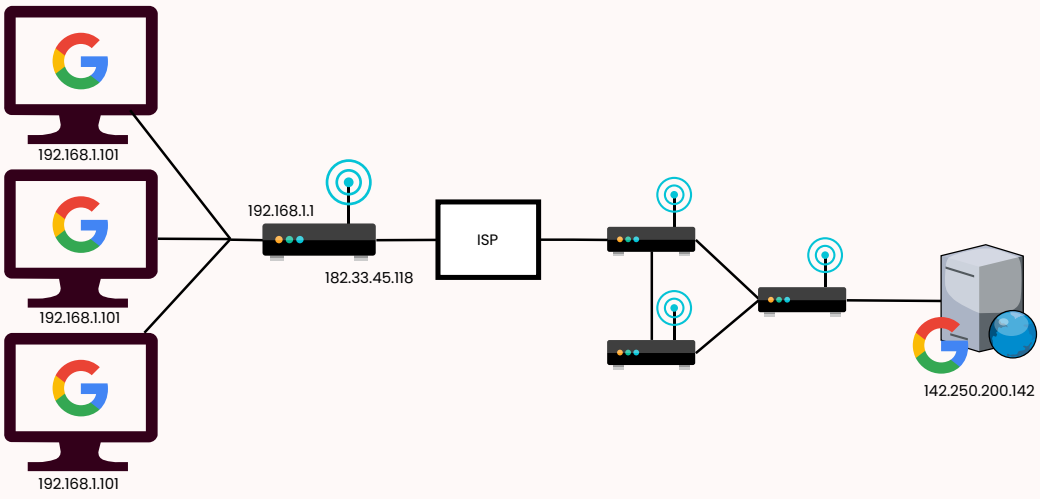
IP Addresses

- Each connected device has at least one IP (Internet Protocol) address.
- Given an address, routers can calculate where they should send information to reach the desired device.



Internet Service Providers

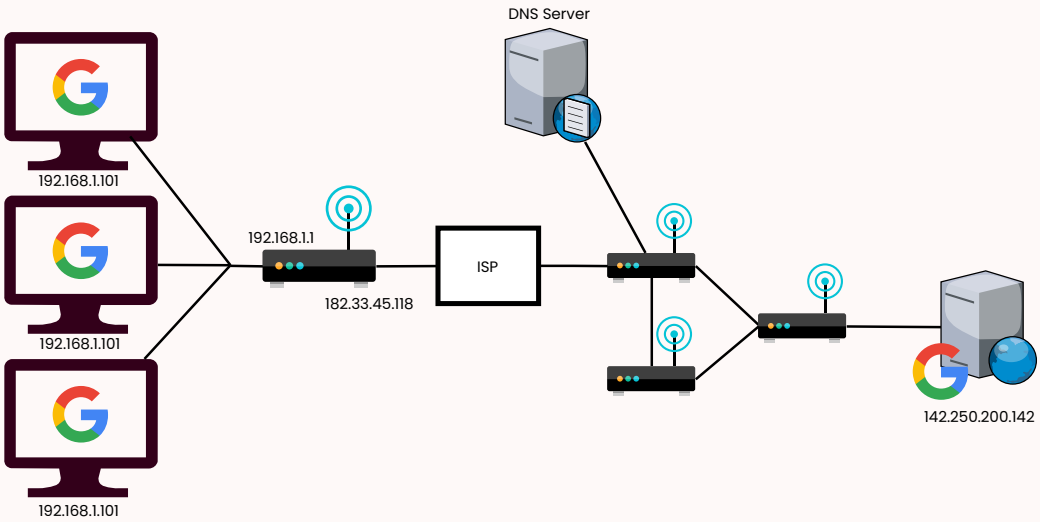
- ISPs are the **organizations** that connect **users** to the **Internet**.
- Most of the time, the **external** IP address on each router is assigned by the ISP using **DHCP**.



Name Resolution

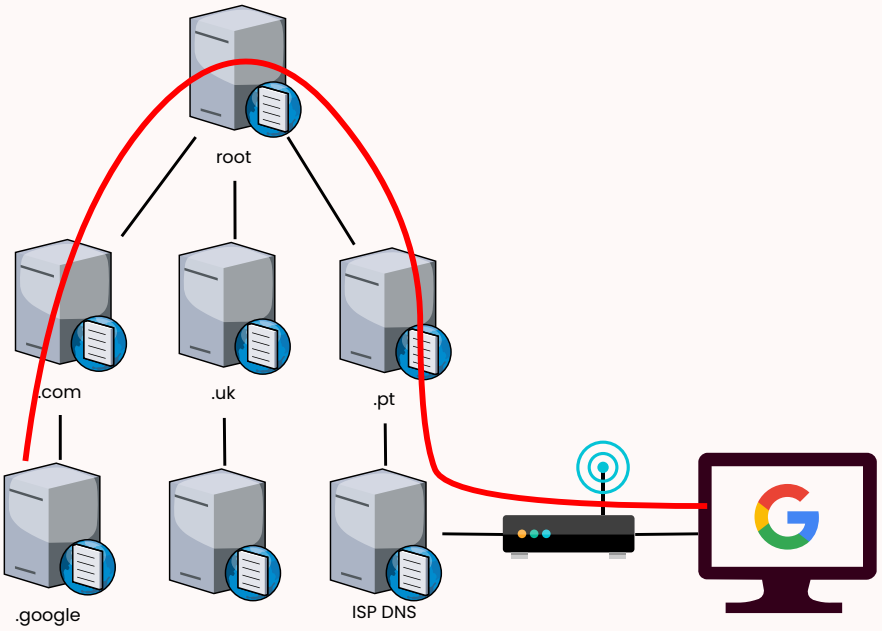
How do we go from www.google.com to 173.194.34.224?

The **Domain Name System (DNS)** is a hierarchical distributed naming system for computers connected to the Internet.



DNS Hierarchy

- If client caches (*i.e.*, computer, router) don't know the IP address, a **DNS query** must be made to the **DNS server** assigned by the ISP (via DHCP).
- DNS requests **escalate** the hierarchy until a DNS server contains a record for the desired name.
- If the root zone DNS does not have the record, the request goes down until it reaches the responsible zone DNS.



URL

Uniform Resource Locators

- A **Uniform Resource Locator (URL)** is a character string that constitutes a reference to an Internet resource.
- It always starts with a **scheme name** followed by a colon and two slashes.
- In the case of the HTTP scheme, it is followed by a **server name** (or an IP address) and, optionally, a **port number**, the **path** of the resource to be fetched, a **query** string, and a **fragment** identifier.
- Before the server name, it is also possible to add a **username** and a **password**.
- Other common schemes: https, file, ftp, smtp, ...

Uniform Resource Locator Examples

```
http://www.google.com/
http://username:password@www.example.com/path/ima
http://www.example.com:80/path?query_string#fragm
```

- The port is 80 by default.
- The query string allows one to pass parameters to the resource.
- The fragment id indicates a specific point on the resource.

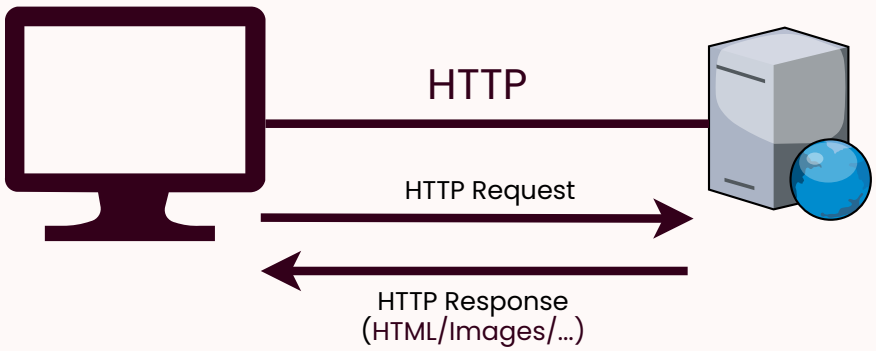
HTTP

Hypertext Transfer Protocol

- The **Hypertext Transfer Protocol (HTTP)** is a protocol that mediates the flow of information between a client computer (generally in the form of a browser) and a web server.
- When a particular URL is introduced into the browser location bar, the browser creates an HTTP connection to the desired server and requests the resource represented by the URL.
- It is the server's responsibility to return that resource to the browser via the same connection (or produce an error).
- The browser then presents the resource to the user.

Resources

- Resources can be of various types.
- The most common are HTML pages, but they can also be images, style sheets, PDF files, ...
- The browser is responsible for presenting them in the most convenient way to the user.



A Typical Scenario

- A web server that receives **requests** for HTML pages.
- HTML pages are **partially generated** from data in a **database**.
- The HTML reaches the browser that requests **additional content** (CSS, JavaScript, images, ...).
- Page is **rendered** in the browser.