

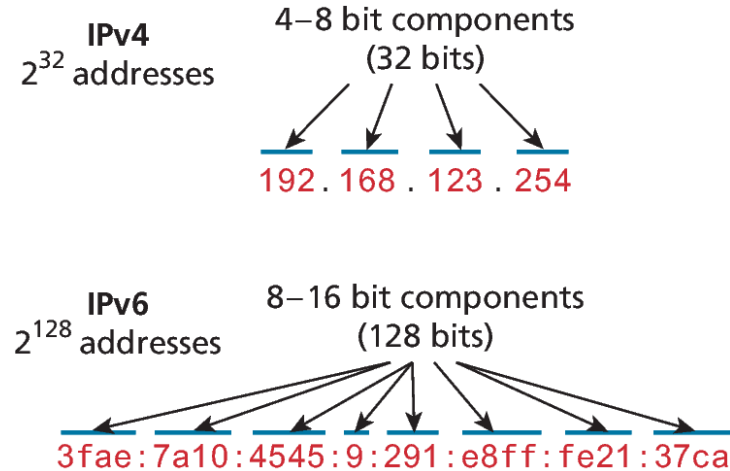
DNS, URLs

Sridhar Alagar

IP Addresses

There are two types of IP addresses: IPv4 and IPv6.

- In **IPv4**, four 8-bit integers separated by . encode the address.
- **IPv6** uses eight 16-bit integers and has over a billion billion times the number in IPv4

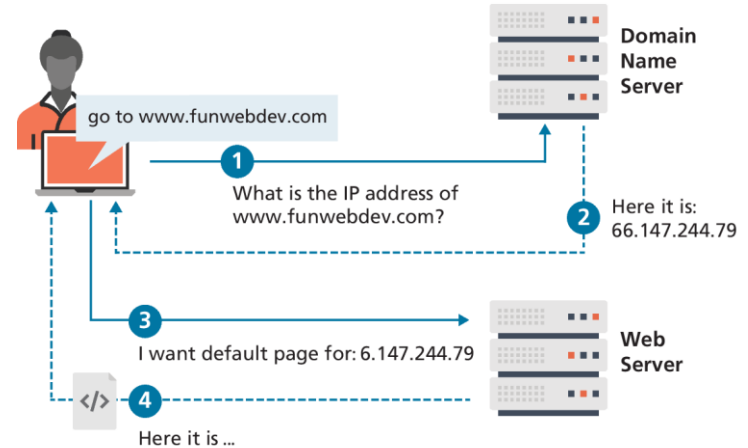


Domain Name System

- As elegant as IP addresses may be, human beings do not enjoy having to recall long strings of numbers.
- Even as far back as the days of ARPANET, researchers assigned **domain names** to IP addresses
- In those early days, the number of Internet hosts was small, so a list of a domains and associated IP addresses could be downloaded as needed as a **hosts** file
- As the number of computers on the Internet grew, this **hosts** file had to be replaced with a better, more scalable, and distributed system. This system is called the **Domain Name System (DNS)**

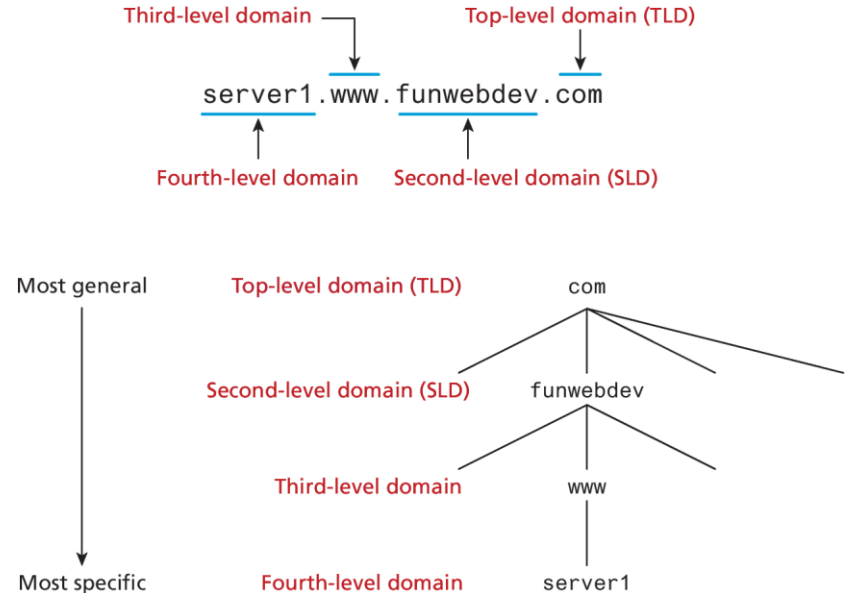
DNS Overview

- The DNS system maps/resolves domain names to IP addresses.
- By separating the domain name of a server from its IP address, a site can move to a different host without changing its name.
- Since the entire request-response cycle can take less than a second, it is easy to forget that DNS requests are happening in all.
- The actual process is more complex

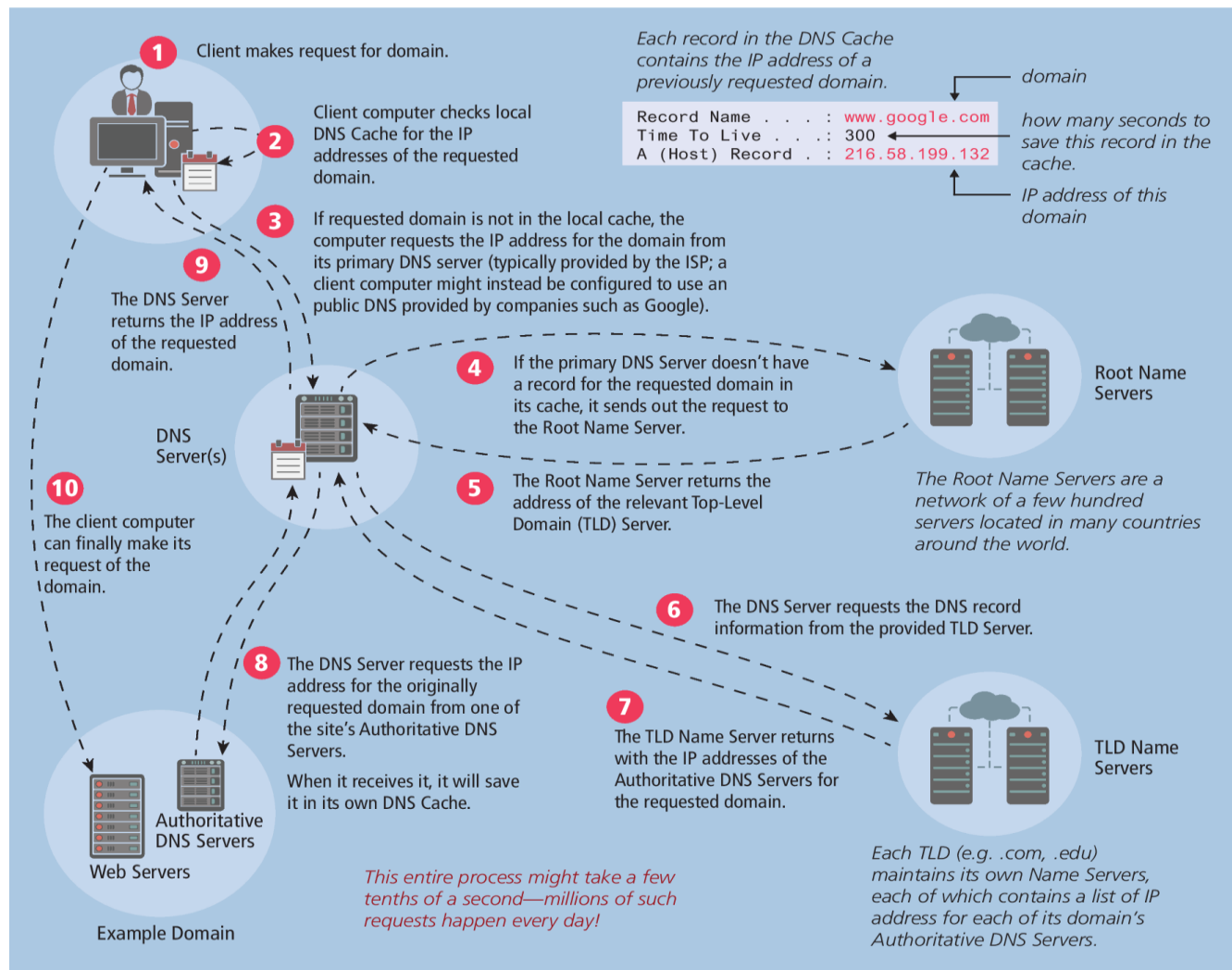


Name Levels

- A domain name can be broken down into several parts, which describe a hierarchy.
- All domain names have at least a **top-level domain (TLD)** name and a **second-level domain (SLD)** name



Address Resolution



Uniform Resource Locators

Uniform Resource Locators (URL) allow clients to request particular resources (files) from the server.

URLs consist of two required components:

1. the protocol used to connect and
2. the domain (or IP address) to connect to.

`http://www.funwebdev.com/index.php?page=17#article`

Protocol Domain Path Query String Fragment

Uniform Resource Locators (optional)

Optional components of the URL are:

- the **path** (which identifies a file or directory to access on that server),
- the **port** to connect to,
- a **query string**, and
- a **fragment** identifier

`http://www.funwebdev.com/index.php?page=17#article`

Protocol Domain Path Query String Fragment

Port (URL)

- A **port** is a type of software connection point used by the underlying TCP/IP protocol and the connecting computer.
- Although the port attribute is not commonly used in production sites, it can be used to route requests to a test server, to perform a stress test, or even to circumvent Internet filters.
- If no port is specified, the protocol determines which port to use. For instance, port 80 is the default port for web-related HTTP requests.
- Syntax is to add a colon after the domain, then specify an integer port number. <http://funwebdev.com:8080/> would connect on port 8080

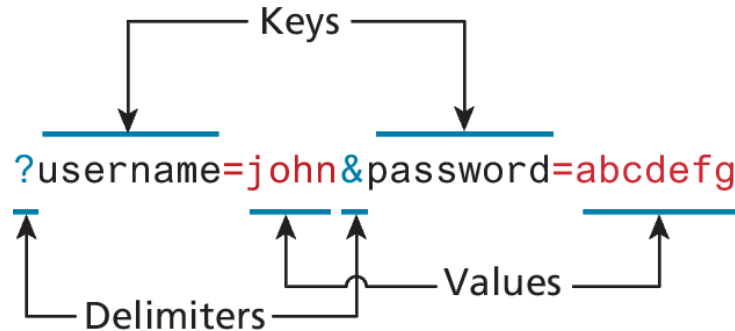
Path (URL)

- The **path** is like the one is used in a file system.
- The root of a web server corresponds to a folder somewhere on that server.
 - Linux servers: that path is **/var/www/html/** or something similar (for Windows it is often **/inetpub/wwwroot/**).
- The path is optional. However, when requesting a folder or the top-level page of a domain, the web server will decide which file to send you.

Query String (URL)

A critical way of passing information, such as user form input, from the client to the server.

- In URLs, they are encoded as key-value pairs delimited by & symbols and preceded by the ? Symbol
- A typical query string for passing name and password:



Fragment (URL)

- The last part of a URL is the optional fragment.
- This is used as a way of requesting a portion of a page.
- Browsers will see the fragment in the URL (denoted by #), seek out the fragment tag anchor in the HTML, and scroll the website down to it.
- “back to top” links are a common use of fragments.

Different types of link

Full URL: `2009 News`

Absolute URL: ``

same as <http://www.xyz.com/stock/quote.html>

Relative URL (intra-site links): ``

same as <http://www.xyz.com/news/2008/March.html>

Define an anchor point (a position that can be referenced with # notation):

Go to a different place in the same page: ``

URL Encoding

What if you want to include a punctuation character in a query value?

`http://www.stats.com/companyInfo?name=C&H Sugar`

Any character in a URL other than A-Z, a-z, 0-9, or any of `-_.` must be represented as `%xx`, where `xx` is the hexadecimal value of the character:

`http://www.stats.com/companyInfo?name=C%26H%20Sugar`

Name Levels (Top Level)

The rightmost portion of the domain name (to the right of the rightmost period) is called **the top-level domain**.

For the top level of a domain, we are limited to two broad categories, plus a third reserved for other use.

1. Generic top-level domain (gTLD)
2. Country code top-level domain (ccTLD)
3. **.arpa** (used for reverse DNS lookups)

Generic top-level domain (gTLD)

Generic top-level domains (gTLD) include the famous .com and .org. There are 3 subtypes of gTLD.

- **Unrestricted.** TLDs include .com, .net, .org, and .info.
- **Sponsored.** TLDs including .gov, .mil, .edu, and others
- **New.** Starting in June 2012, ICANN invited companies to launch new TLDs in order to provide more choice. Since then over 1000 new TLD have been created including .art, .cash, .cool, .jobs, .tax and so on

Country code top-level domain

Country code top-level domain (ccTLD) are under the control of the countries which they represent, which is why each is administered differently.

- United Kingdom - **co.uk**; Canada - **.ca**
- Other countries have peculiar extensions with commercial viability (such as **.tv** for Tuvalu) and have begun allowing unrestricted use to generate revenue.
- **Internationalized top-level domain name (IDN)** allows domains to use non-ascii characters and has been deployed since 2009. There are over 9 million IDN domains