

FROM URL TO WEB PAGE

A Brief Overview

What Is a URL?

`http://sub.host.com:80/path?query=string#hash`

Uniform Resource Locator

- Wikipedia: “A **uniform resource locator**, abbreviated **URL**, also known as **web address**, is a specific character string that constitutes a reference to a resource.”

Scheme/Protocol

`http://sub.host.com:80/path?query=string#hash`

Indicates the type of resource requested

- Web pages use “http” or “https” (secure HTTP) scheme
- Other common schemes include “ftp” (file transfer protocol), “file”, and “mailto”
- Optional in web browsers. Will default to “http”

Host Name

`http://sub.host.com:80/path?query=string#hash`

- Indicates the server where the resource resides
- Increasingly specific from right-to-left

Top Level Domain (TLD)

`http://sub.host.com:80/path?query=string#hash`

Usually the right-most element of the host; however there are some exceptions for country TLDs where the TLD is the two right-most items (e.g., .co.uk for British sites)

- Various restrictions apply to acquiring some TLDs. Country-codes must be hosted in the country indicated. .mil and .gov are restricted to U.S. military and government domains, respectively.

Second-level Domain Name

`http://sub.host.com:80/path?query=string#hash`

Consists of the TLD and one item to its left.

- When you register (rent) a domain for a website, it is a second-level domain. For example, andyhill.us is registered to me.
- Often simply called a domain name

Subdomain

`http://sub.host.com:80/path?query=string#hash`

Technically, in the example above, host.com is a subdomain of .com; however in standard usage, subdomains refer to domains below the domain name.

- Hosts can have any number of subdomains; e.g., mail.example.com, shop.example.com, etc.
- There can also be multiple levels of subdomains (e.g., mail.shop.example.com)
- “www” is generally an alias for the domain name

Phishing

While there are many aspects of phishing attacks, a couple points to note based on what we've seen so far:

Paying attention to the host can reveal an impersonation of a host. E.g.,

1. secure.microsoft.com = domain owned by microsoft.com
2. secure.microsoft.phishing.com = not a microsoft domain
2. Any transaction involving financial or personal data should be using the https protocol. If it doesn't, it is probably not legitimate.
3. Official government web sites should have the .gov TLD if it does not it is suspect

Port

`http://sub.host.com:80/path?query=string#hash`

- According to Wikipedia: “In computer networking, a **port** is an application-specific or process-specific software construct serving as a communications endpoint in a computer's host operating system.”
- The port is often omitted in a URL. If this is the case, the default port for the scheme is used:
 - http: 80
 - https: 443

Path

`http://sub.host.com:80/path?query=string#hash`

- The path indicates the specific resource on the server being requested.
- The path can represent an actual path on the server, meaning a path of /a/b/c can represent a folder structure on the server
- The path can also represent a virtual path, meaning that there may not be an /a/b/c folder structure on the server, but the server maps this path to a specific resource (for example, a RESTful architecture).

Query String

`http://sub.host.com:80/path?query=string#hash`

- Comprises additional information passed to the path (for example, the search parameters passed to a search engine)

Starts with a question mark and is a series of name value pairs of the form

`name1=value1&name2=value2`

- Spaces are converted to plus signs. Other special characters are converted to a percent sign followed by the hexadecimal code for the character (e.g., ('+' => '%2B'))

Hash

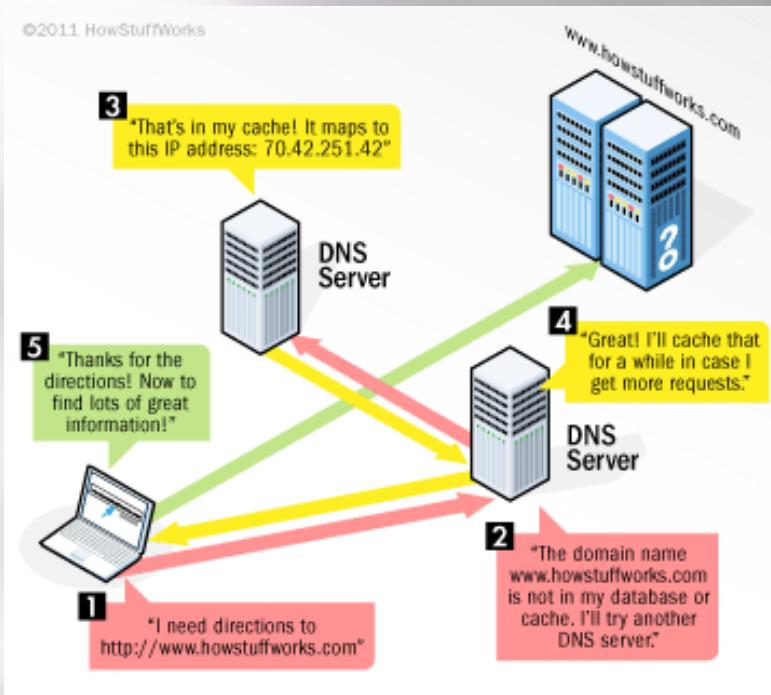
`http://sub.host.com:80/path?query=string#hash`

Indicates a location within the resource.

Traditionally, these have been used for tables of contents and “Return to Top” links

In modern applications, however, they are often used for navigation without reloading the entire resource. For example: <http://www.andyhill.us/news/>

Domain Name Service (DNS)



- The web host in the URL must be converted to an IP address
- Once this is accomplished, the request is sent to the server
- nslookup, tracert

Server Processing

- Server routes to correct resource based on scheme, port, and path

Additional information can be passed to the resource via GET (Query String) or POST (Form Data) parameters (see HTTP Request Methods)
- Various server-side processing can be employed: Java, ASP.NET, PHP, ColdFusion, Ruby on Rails, Perl, Python, etc.
- Regardless, if a web page is requested, what is returned is an HTML document; however, this document may request more resources

HTML Document

The HTML Document is received and processed by the browser. Essentially, this document can include:

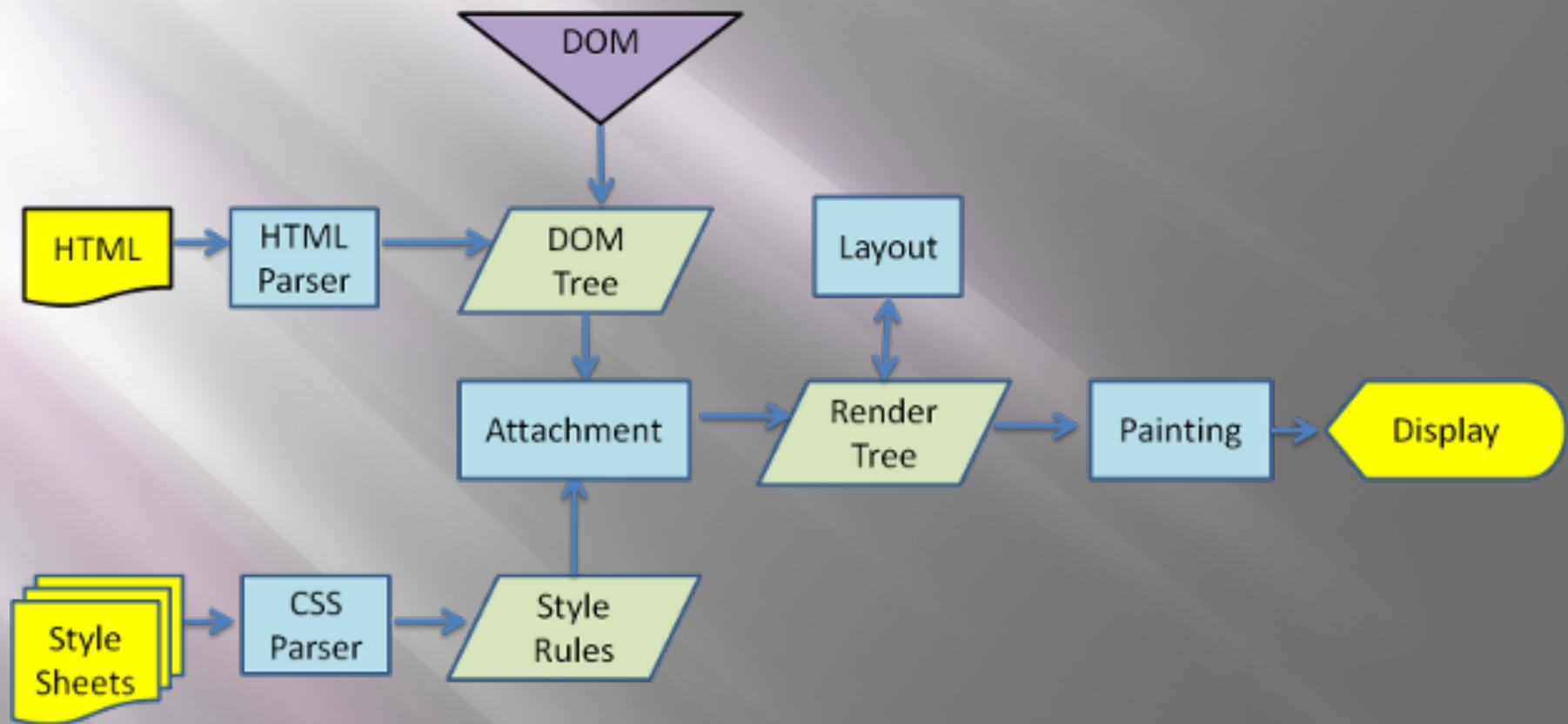
HTML Markup indicating the structure of the document and its content

JavaScript which defines the behavior of the elements in document

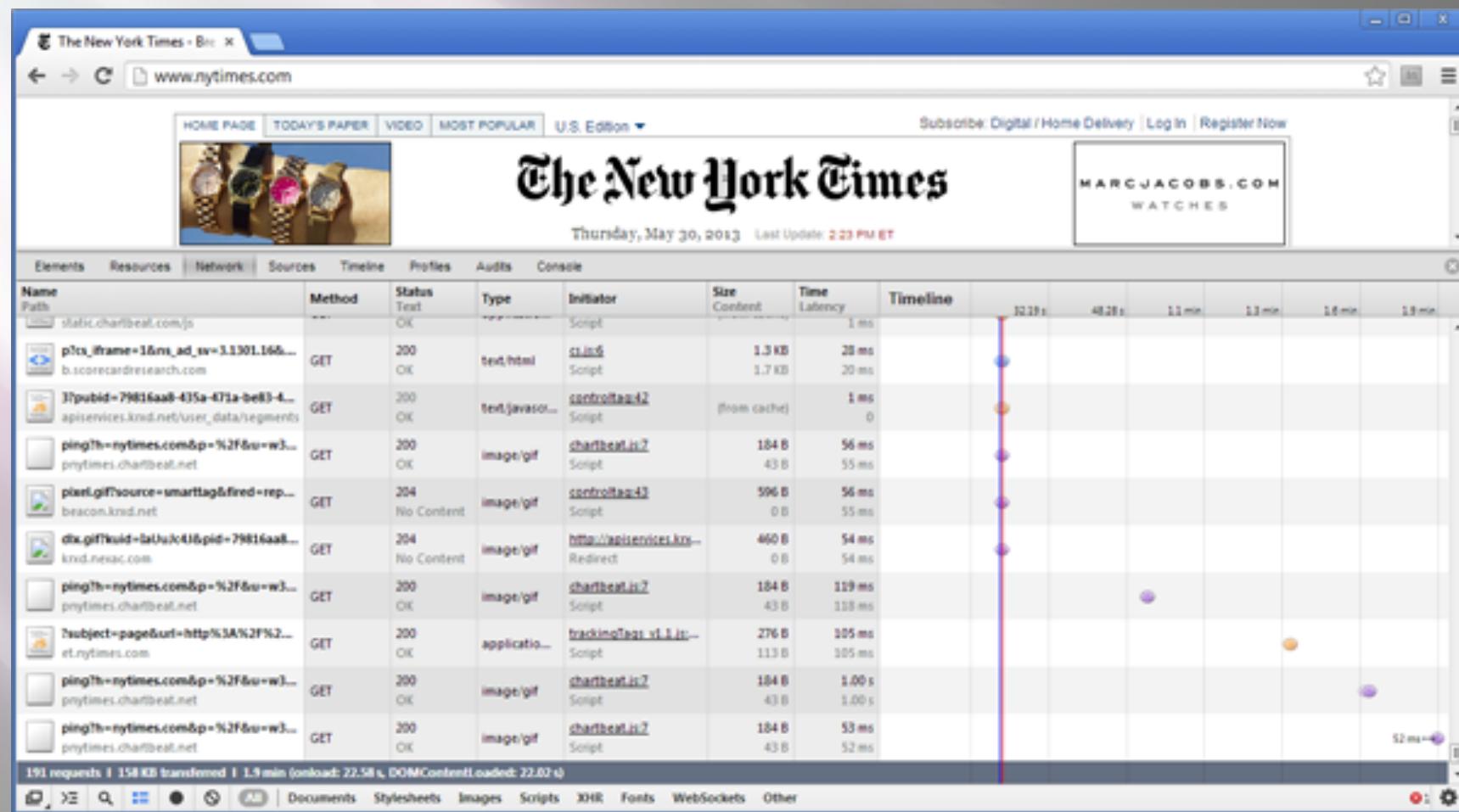
- Cascading Style Sheets (CSS) which define the presentation of the document
- Embedded objects, which can include images, Flash or Java applets, and external CSS and JavaScript files

Browser Processing

Webkit Main Flow



Timeline



HTML

```
<!DOCTYPE html>
<html>
<head>
  <meta charset="utf-8">
  <title>HTML Demo</title>
</head>
<body>
  <h1 id="main-header">HTML Demo</h1>
  
  <p>
    Lorem ipsum dolor sit amet, consectetur adipiscing
```

History

- Styling: text
- Behavior: <form action="get" onclick="validationFunction(this);">...</form>

CSS

```
body {  
    background: gray;  
    color: #FFF;  
    margin: 1em;  
    position: relative;  
}
```

```
#main-header {  
    text-align: center;  
    font-style: italic;
```

JavaScript

```
document.getElementById("main-header").onclick=function(e){  
    alert('header clicked');  
};  
  
Array.filter( document.getElementsByClassName('iu-icon'), function(elem)  
{  
    elem.onmouseover=function(e) {  
        this.src='/global/img/iu/blockiu_crimson.gif';  
    };  
    elem.onmouseout=function(e) {  
        this.src='/global/img/iu/IU.jpg';  
    };  
}  
);
```

jQuery

```
$(function() {  
    $('#main-header').click(function(e) {  
        alert('header clicked');  
    });  
    $('.iu-icon').on({  
        mouseover: function(e) {  
            $(this).attr('src', '/global/img/iu/blockiu_crimson.gif');  
        },  
        mouseout: function(e) {  
            $(this).attr('src', '/global/img/iu/IU.jpg');  
        }  
    });  
    $('p').draggable();  
});
```