Web Server Design

Lecture 7 – Content Negotiation

Old Dominion University

Department of Computer Science

CS 431/531 Fall 2019

Sawood Alam <salam@cs.odu.edu>

2019-10-10

Revisiting Terminology from Lecture 1

Important Web Architecture Concepts

URIs
`http://www.cs.odu.edu/~mln/`

Identify →

Resources

Represent →

Representations*

```
<html>
<head>
<title>
Home:: Michael L. Nelson, Old Dominion University
</title>
<link rel="stylesheet" type="text/css"
href="mln.css"/>
<script type="text/javascript"
src="mln.js"></script>
…
```
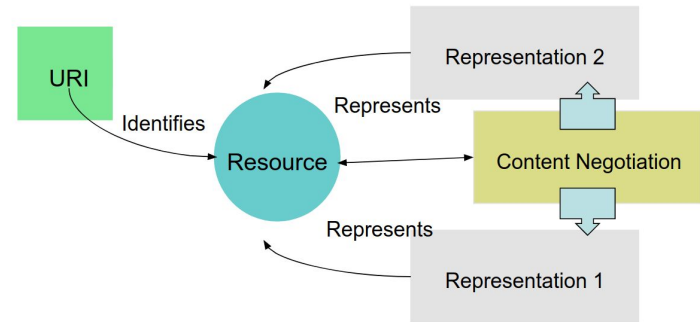
As defined by the Web Architecture
http://www.w3.org/TR/webarch/
*= "message" or "message body" in RFC 7231,
"entity"/"entity-body" in RFC-2616

Resources can have multiple, simultaneous *representations*

URI

Identifies →

Resource

Represents →

Representation 2

Content Negotiation

Represents →

Representation 1

Content Negotiation

RFC 7231, Section 3.4

- "Proactive" ("Server-side" in RFC 2616)
  - Server picks best representation
    - Agent can pass in "hints" via Accept.* headers
    - See Apache algorithm at: http://httpd.apache.org/docs/current/content-negotiation.html
- "Reactive" ("Agent-side" in RFC 2616)
  - Server sends a list to the agent and the agent picks from representation
- Transparent Negotiation
  - Combination of server-side and agent-side performed by caches, proxies, etc.
    - Mentioned in passing in RFC 2616; detailed in RFC 2295
      - https://tools.ietf.org/html/rfc2295

Generic vs. Specific Resources

Dimensions of genericity

When we discuss electronic resources, an interesting fact is that a small number of dimensions of genericity emerge.

Time	A resource may vary with time. For example, "The Wall Street Journal" varies with time. Each issue is a time-specific resource, which does not change with time. Most home pages on the Web change with time, in a less periodic way.
Language	When a document is translated, it is useful to be able to refer to it either in the generic, or to a particular specific translation.
Content-Type	A given resource may have mny ways in which it can be represented on the wire, using different `Content-types` (in HTTP terms). As an example, an image may be represented in PNG or JFIF format.
Target medium	A given resource may be targetted specifically to a specific medium, such as a printer, being displayed on laptop screen, being displayed on a cellphone, or being projected onto a large screen for an audience. (This is currenltly available for selecting CSS stylesheets, but is not done at the HTTP content negotiation level)

The fact that there are such a small number of dimensions currently apparent sugests that Web software should handle them individually in its interface with the user, even though the architecure should handle them as a single concpet.

https://www.w3.org/DesignIssues/Generic

"Cool URIs Don't Change"

What makes a cool URI?
A cool URI is one which does not change.
What sorts of URI change?
URIs don't change: people change them.

There are no reasons at all in theory for people to change URIs (or stop maintaining documents), but millions of reasons in practice.

In theory, the domain name space owner owns the domain name space and therefore all URIs in it. Except insolvency, nothing prevents the domain name owner from keeping the name. And in theory the URI space under your domain name is totally under your control, so you can make it as stable as you like. Pretty much the only good reason for a document to disappear from the Web is that the company which owned the domain name went out of business or can no longer afford to keep the server running. Then why are there so many dangling links in the world? Part of it is just lack of forethought. Here are some reasons you hear out there:

https://www.w3.org/Provider/Style/URI

"how-we-do-it-now"

There is a crazy notion that pages produced by scripts have to be located in a "cgibin" or "cgi" area. This is exposing the mechanism of how you run your server. You change the mechanism (even keeping the content the same) and whoops - all your URIs change.

For example, take the National Science Foundation:

NSF Online Documents
http://www.nsf.gov/cgi-bin/pubsys/browser/oldbrowse.pl

The main page for starting to look for documents, is clearly not going to be something to trust to being there in a few years. "cgi-bin" and "oldbrowse" and ".pl" all point to bits of how-we-do-it-now. By contrast, if you use the page to find a document, you get first an equally bad

Report of Working Group on Cryptology and Coding Theory
http://www.nsf.gov/cgi-bin/getpub?nsf9814

For the document's index page, but the html document itself by contrast is very much better:

http://www.nsf.gov/pubs/1998/nsf9814/nsf9814.htm

Looking at this one, the "pubs/1998" header is going to give any future archive service a good clue that the old 1998 document classification scheme is in progress. Though in 2098 the document numbers might look different, I can imagine this URI still being valid, and the NSF or whatever carries on the archive not being at all embarrassed about it.

https://www.w3.org/Provider/Style/URI

"what to leave out?"

Everything! After the creation date, putting any information in the name is asking for trouble one way or another.

- **Authors name**- authorship can change with new versions. People quit organizations and hand things on.
- **Subject**. This is tricky. It always looks good at the time but changes surprisingly fast. I discuss this more below.
- **Status**- Directories like "old" and "draft" and so on, not to mention "latest" and "cool" appear all over file systems. Documents change status - or there would be no point in producing drafts. The latest version of a document needs a persistent identifier whatever its status is. Keep the status out of the name.
- **Access**. At W3C we divide the site into "Team access", "Member access" and "Public access". It sounds good, but of course documents start off as team ideas, are discussed with members, and then go public. A shame indeed if every time some document is opened to wider discussion all the old links to it fail! We are switching to a simple date code now.
- **File name extension**. This is a very common one. "cgi", even ".html" is something which will change. You may not be using HTML for that page in 20 years time, but you might want today's links to it to still be valid. The canonical way of making links to the W3C site doesn't use the extension. (how?)
- **Software mechanisms**. Look for "cgi", "exec" and other give-away "look what software we are using" bits in URIs. Anyone want to commit to using perl cgi scripts all their lives? Nope? Cut out the .pl. Read the server manual on how to do it.
- **Disk name** - Gimme a break! But I've seen it.

So a better example from our site is simply

> http://www.w3.org/1998/12/01/chairs

a report of the minutes of a meeting of W3C chair people.

CN is how!

https://www.w3.org/Provider/Style/URI

HTTP Solipsism and Content Negotiation

- CN has a bad reputation, in part because some people have difficulty believing in things they can't see

  - https://stackoverflow.com/questions/44720631/is-http-content-negotiation-being-used-by-browsers-and-servers-in-practice

  - https://stackoverflow.com/questions/44735653/why-would-http-content-negotiation-be-preferred-to-explicit-parameters-in-an-api

- And there is a small performance cost

  - https://httpd.apache.org/docs/current/misc/perf-tuning.html
  - "If at all possible, avoid content negotiation if you're really interested in every last ounce of performance. In practice the benefits of negotiation outweigh the performance penalties."
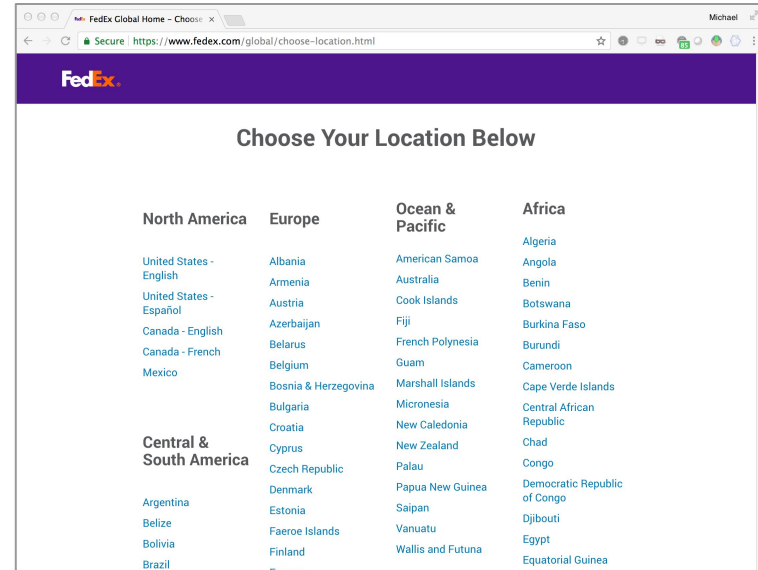
And "client-side" (aka reactive) CN is the norm for languages & file types…
But CN in some dimensions happens *all the time* in the wild…

Turning on Content Negotiation in Apache

- In Apache, content negotiation is turned off by default, and is turned on via:
  - Type-map file (*.var*)
  - *Options +Multiviews* directive in httpd.conf or .htaccess file
    - http://httpd.apache.org/docs/current/content-negotiation.html
- In our servers, content negotiation will be on by default

How it Works

- If a direct match for the requested URI is found, then the entity is returned
  - If the request is for "foo.txt" and you have "foo.txt", then return "foo.txt"
- *If a 404 would be result for the current request*, AND content *negotiation is available for this resource*, then content negotiation begins
  - If the request is for "foo", then the server considers the user agent's preferences and searches for the "best" *available* representation for "foo"

Request Headers & Status Codes

- Request headers
  - Accept
  - Accept-Charset
  - Accept-Encoding
  - Accept-Language
  - Negotiate (from RFC 2295)
- Response headers
  - Content-Location
  - Vary
  - TCN (from RFC 2295)
  - Alternates (from RFC 2295)

- Status codes
  - 300 Multiple Choices
  - 406 Not Acceptable

Test Directory

```
$ cd a3-test
$ ls
fairlane.gif    index.html.de   index.html.ja.jis      type-map.example
fairlane.jpeg   index.html.en   index.html.ko.euc-kr   vt-uva.html.gz
fairlane.png    index.html.es   index.html.ru.koi8-r   vt-uva.html.Z
$ cat .htaccess
Options All +MultiViews
```

Note: No "index.html"

Also note: The following examples no longer work on the departmental accounts.
Thanks, Nginx.

User-Agent (UA) passes no preferences, server chooses

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 04:04:22 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: fairlane.txt
Vary: negotiate,accept
TCN: choice
Last-Modified: Mon, 13 Mar 2006 04:00:53 GMT
ETag: "2288-c1-4414ee75;4414ee7a"          ←        Note: structured ETag
Accept-Ranges: bytes
Content-Length: 193
Connection: close
Content-Type: text/plain

Connection closed by foreign host.
```

This representation has its own URI: http://www.cs.odu.edu/~mln/teaching/cs595-s06/a3-test/fairlane.txt
But most (all?) UAs will display: http://www.cs.odu.edu/~mln/teaching/cs595-s06/a3-test/fairlane

Structured Entity Tag
RFC 2295, section 9.2

9.2 Structured entity tags

A structured entity tag consists of a normal entity tag of which the opaque string is extended with a semicolon followed by the text (without the surrounding quotes) of a variant list validator:

```
 normal       | variant list |   structured
entity tag    | validator    |   entity tag
------------+--------------+--------------
  "etag"      |    "vlv"     |   "etag;vlv"
W/"etag"      |    "vlv"     | W/"etag;vlv"
```

ETag: `"2288-c1-4414ee75;4414ee7a"`
`2288-c1-4414ee75` is for `fairlane.txt`
`;` is the separator
`4414ee7a` is for `fairlane`

Note that a structured entity tag is itself also an entity tag. The structured nature of the tag allows caching proxies capable of transparent content negotiation to perform some optimizations defined in section 10. When not performing such optimizations, a structured tag SHOULD be treated as a single opaque value, according to the general rules in HTTP/1.1. Examples of structured entity tags are:

"xyzzy;1234" W/"xyzzy;1234" "gonkxxxx;1234" "a;b;c;;1234"

In the last example, the normal entity tag is "a;b;c;" and the variant list validator is "1234".

UA prefers images

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Accept: image/*; q=1.0
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 04:06:45 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: fairlane.jpeg
Vary: negotiate,accept
TCN: choice
Last-Modified: Sun, 12 Mar 2006 17:37:16 GMT
ETag: "3b64bd-9639-44145c4c;4414ee7a"
Accept-Ranges: bytes
Content-Length: 38457
Connection: close
Content-Type: image/jpeg

Connection closed by foreign host.
```

"quality value" or "qvalue"; RFC 7231, section 5.3.1

UA prefers png over gif, and gif over jpeg

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Accept: image/png; q=1.0, image/gif; q=0.5, image/jpeg; q=0.1
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 02:30:44 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: fairlane.png
Vary: negotiate,accept
TCN: choice
Last-Modified: Sun, 12 Mar 2006 17:37:31 GMT
ETag: "3b64bf-17f9b-44145c5b;4414d473"
Accept-Ranges: bytes
Content-Length: 98203
Connection: close
Content-Type: image/png

Connection closed by foreign host.
```

UA prefers tiff, but server has no tiff

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Accept: image/tiff; q=1.0, image/gif; q=0.001
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 02:37:10 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: fairlane.gif
Vary: negotiate,accept
TCN: choice
Last-Modified: Sun, 12 Mar 2006 20:46:10 GMT
ETag: "3b64c0-c28b-44148892;4414d473"
Accept-Ranges: bytes
Content-Length: 49803
Connection: close
Content-Type: image/gif

Connection closed by foreign host.
```

Server chooses an encoding

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/vt-uva HTTP/1.1
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 03:26:35 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: vt-uva.html.gz
Vary: negotiate,accept-encoding
TCN: choice
Last-Modified: Sun, 12 Mar 2006 20:52:54 GMT
ETag: "3b64c1-2c54-44148a26;4414d473"
Accept-Ranges: bytes
Content-Length: 11348
Connection: close
Content-Type: text/html
Content-Encoding: x-gzip

Connection closed by foreign host.
```

UA does not want gzip

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/vt-uva HTTP/1.1
Accept-Encoding: compress; q=0.1, gzip; q=0.0
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 03:29:54 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: vt-uva.html.Z
Vary: negotiate,accept-encoding
TCN: choice
Last-Modified: Sun, 12 Mar 2006 20:52:41 GMT
ETag: "3b64c3-5c0b-44148a19;4414d473"
Accept-Ranges: bytes
Content-Length: 23563
Connection: close
Content-Type: text/html
Content-Encoding: compress

Connection closed by foreign host.
```

UA wants only German

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/ HTTP/1.1
Accept-Language: de; q=1.0
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 03:32:36 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: index.html.de
Vary: negotiate,accept-language,accept-charset
TCN: choice
Last-Modified: Sun, 12 Mar 2006 16:39:23 GMT
ETag: "3b64b7-1c94-44144ebb;4414d473"
Accept-Ranges: bytes
Content-Length: 7316
Connection: close
Content-Type: text/html
Content-Language: de

Connection closed by foreign host.
```

Representations for this resource can vary in the dimension of accept-charset, even though this resource does not (i.e., no charset parameter on Content-type)

UA prefers iso-2022-jp character set

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/index.html HTTP/1.1
Accept-Language: ja; q=1.0
Accept-Charset: iso-2022-jp; q=1.0
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 200 OK
Date: Mon, 13 Mar 2006 03:35:17 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Content-Location: index.html.ja.jis
Vary: negotiate,accept-language,accept-charset
TCN: choice
Last-Modified: Sun, 12 Mar 2006 16:39:23 GMT
ETag: "3b64ba-1dd3-44144ebb;4414d473"
Accept-Ranges: bytes
Content-Length: 7635
Connection: close
Content-Type: text/html; charset=iso-2022-jp
Content-Language: ja

Connection closed by foreign host.
```

UA wants Japanese, but only in EUC-JP

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/index.html HTTP/1.1
Accept-Language: ja; q=1.0
Accept-Charset: euc-jp; q=1.0
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 406 Not Acceptable
Date: Mon, 13 Mar 2006 03:39:29 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Alternates: {"index.html.de" 1 {type text/html} {language de} {length 7316}},
{"index.html.en" 1 {type text/html} {language en} {length 7233}},
{"index.html.es" 1 {type text/html} {language es} {length 7643}},
{"index.html.ja.jis" 1 {type text/html} {charset iso-2022-jp} {language ja}
{length 7635}},
{"index.html.ru.koi8-r" 1 {type text/html} {charset koi8-r} {language ru}
{length 7277}}
Vary: negotiate,accept-language,accept-charset
TCN: list
Connection: close
Content-Type: text/html; charset=iso-8859-1

Connection closed by foreign host.
```

UA wants only tiff

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
HEAD /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Accept: image/tiff; q=1.0, */*;q=0.0
Connection: close

HTTP/1.1 406 Not Acceptable
Date: Mon, 13 Mar 2006 04:03:01 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Alternates: {"fairlane.gif" 1 {type image/gif} {length 49803}},
{"fairlane.jpeg" 1 {type image/jpeg} {length 38457}},
{"fairlane.png" 1 {type image/png} {length 98203}},
{"fairlane.txt" 1 {type text/plain} {length 193}}
Vary: negotiate,accept
TCN: list
Connection: close
Content-Type: text/html; charset=iso-8859-1

Connection closed by foreign host.
```

Same, but with GET

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
GET /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Host: www.cs.odu.edu
Accept: image/tiff; q=1.0, *;q=0.0
Connection: close

HTTP/1.1 406 Not Acceptable
Date: Tue, 14 Mar 2006 03:30:23 GMT
Server: Apache/1.3.26 (Unix) ApacheJServ/1.1.2 PHP/4.3.4
Alternates: {"fairlane.gif" 1 {type image/gif} {length 49803}},
{"fairlane.jpeg" 1 {type image/jpeg} {length 38457}}, {"fairlane.png" 1 {type image/png} {length 98203}},
{"fairlane.txt" 1 {type text/plain} {length 193}}
Vary: negotiate,accept
TCN: list
Connection: close
Transfer-Encoding: chunked
Content-Type: text/html; charset=iso-8859-1


27d
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<HTML><HEAD>
<TITLE>406 Not Acceptable</TITLE>
</HEAD><BODY>
<H1>Not Acceptable</H1>
An appropriate representation of the requested resource /~mln/teaching/cs595-s06/a3-test/fairlane
could not be found on this server.<P>
Available variants:
<ul>
<li><a href="fairlane.gif">fairlane.gif</a> , type image/gif
<li><a href="fairlane.jpeg">fairlane.jpeg</a> , type image/jpeg
<li><a href="fairlane.png">fairlane.png</a> , type image/png
<li><a href="fairlane.txt">fairlane.txt</a> , type text/plain
</ul>
<HR>
<ADDRESS>Apache/1.3.26 Server at www.cs.odu.edu Port 80</ADDRESS>
</BODY></HTML>

0
```

300 Multiple Choices

(It's *kind of* like a redirection… A user-drive redirection)

6.4.1. 300 Multiple Choices

The 300 (Multiple Choices) status code indicates that the target resource has more than one representation, each with its own more specific identifier, and information about the alternatives is being provided so that the user (or user agent) can select a preferred representation by redirecting its request to one or more of those identifiers. In other words, the server desires that the user agent engage in reactive negotiation to select the most appropriate representation(s) for its needs (Section 3.4).

…

For request methods other than HEAD, the server SHOULD generate a payload in the 300 response containing a list of representation metadata and URI reference(s) from which the user or user agent can choose the one most preferred. The user agent MAY make a selection from that list automatically if it understands the provided media type. A specific format for automatic selection is not defined by this specification because HTTP tries to remain orthogonal to the definition of its payloads. In practice, the representation is provided in some easily parsed format believed to be acceptable to the user agent, as determined by shared design or content negotiation, or in some commonly accepted hypertext format.

Transparent Content Negotiation

- Defined in RFC 2295
- Client requests CN
  - "transparent" does not mean "hide from client", it means "making transparent all available representations for a given URI"
- Three kinds of responses:
  - list: "here is a list for the client to choose from"
  - choice: "the server made this choice for the client"
  - adhoc: "weird things are happening, so the server made a choice" (a (hopefully) rare situation)

Negotiate Directive

```
$ telnet www.cs.odu.edu 80
Trying 128.82.4.2...
Connected to xenon.cs.odu.edu.
Escape character is '^]'.
GET /~mln/teaching/cs595-s06/a3-test/fairlane HTTP/1.1
Negotiate: 1.0
Host: www.cs.odu.edu
Connection: close

HTTP/1.1 300 Multiple Choices
Date: Sun, 07 Jan 2007 18:20:04 GMT
Server: Apache/2.2.0
Alternates: {"fairlane.gif" 1 {type image/gif} {length 49803}}, {"fairlane.jpeg" 1 {type image/jpeg} {length 38457}},
{"fairlane.png" 1 {type image/png} {length 98203}}, {"fairlane.txt" 1 {type text/plain} {length 193}}
Vary: negotiate,accept
TCN: list
Content-Length: 524
Connection: close
Content-Type: text/html; charset=iso-8859-1

<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>300 Multiple Choices</title>
</head><body>
<h1>Multiple Choices</h1>
Available variants:
<ul>
<li><a href="fairlane.gif">fairlane.gif</a> , type image/gif</li>
<li><a href="fairlane.jpeg">fairlane.jpeg</a> , type image/jpeg</li>
<li><a href="fairlane.png">fairlane.png</a> , type image/png</li>
<li><a href="fairlane.txt">fairlane.txt</a> , type text/plain</li>
</ul>
<hr>
<address>Apache/2.2.0 Server at www.cs.odu.edu Port 80</address>
</body></html>
Connection closed by foreign host.
```

CN in the wild

```
$ curl -Is https://www.cnn.com/
HTTP/1.1 200 OK
Content-Type: text/html; charset=utf-8
x-servedByHost: ::ffff:172.17.64.4
access-control-allow-origin: *
cache-control: max-age=60
content-security-policy: [deletia]
x-content-type-options: nosniff
x-xss-protection: 1; mode=block
Content-Length: 1717605
Accept-Ranges: bytes
Date: Mon, 22 Oct 2018 23:46:55 GMT
Via: 1.1 varnish
Age: 214
Connection: keep-alive
Set-Cookie: [deletia]
X-Served-By: cache-iad2644-IAD
X-Cache: HIT
X-Cache-Hits: 8
X-Timer: S1540252016.528240,VS0,VE0
Vary: Accept-Encoding
$ curl -is https://www.cnn.com/ | wc
      43   46991 1719081
$ curl -is -H "Accept-encoding: gzip" https://www.cnn.com/ | wc
     665    5521  183355
$ curl -is -H "Accept-encoding: compress" https://www.cnn.com/ | wc
      43   47025 1719555
$ curl -is -H "Accept-encoding: gzip" https://www.cnn.com/ | grep -ia "Content-Encoding:"
Content-Encoding: gzip
$ curl -is -H "Accept-encoding: compress" https://www.cnn.com/ | grep -ia "Content-Encoding:"
$              # Blank line means no match
```

CN in more than just Accept-Encoding

```
$ curl -ILs -A "iphone" en.wikipedia.org | grep -iE "^(http|location|Vary)"
HTTP/1.1 301 TLS Redirect
Location: https://en.wikipedia.org/
HTTP/2 301
location: https://en.wikipedia.org/wiki/Main_Page
vary: Accept-Encoding,X-Forwarded-Proto,Cookie,Authorization
HTTP/2 302
location: https://en.m.wikipedia.org/wiki/Main_Page
HTTP/2 200
vary: Accept-Encoding,Cookie,Authorization
```

CN with Host

```
$ curl -IL www.odu.edu
HTTP/1.0 302 Found
Location: https://www.odu.edu/
Server: BigIP
Connection: Keep-Alive
Content-Length: 0

HTTP/1.1 200 OK
Date: Wed, 24 Oct 2018 16:38:44 GMT
Server: Apache/2.2.15 (Red Hat)
Vary: Host,Accept-Encoding
Accept-Ranges: bytes
Connection: close
Content-Type: text/html; charset=UTF-8
Set-Cookie: BIGipServerWEB_HTTPS_PROD.app~WEB_HTTPS_PROD_pool_int=
 rd741o00000000000000000000ffffc0a86094o80; path=/
```

CN with User-Agent

```
$ curl -I https://www.grosbill.com/
HTTP/1.1 200 OK
Date: Tue, 23 Oct 2018 00:15:16 GMT
Server: Apache/2.4.25 (Debian)
X-Powered-By: PHP/5.6.23
Set-Cookie: [deletia]
Cache-Control: max-age=1
Expires: Tue, 23 Oct 2018 00:15:17 GMT
Vary: User-Agent
Content-Type: text/html; charset=UTF-8
Via: 1.1 google
Transfer-Encoding: chunked
Alt-Svc: clear
$ curl -is https://www.grosbill.com/ | wc
      13    1218   34823
$ curl -is -A "iphone" https://www.grosbill.com/ | wc
      13    1218   34823
$ curl -is -A "Mozilla/5.0 (iPhone; CPU iPhone OS 8_4_1 like Mac OS X) AppleWebKit/600.1.4
(KHTML, like Gecko) CriOS/45.0.2454.68 Mobile/12H321 Safari/600.1.4" https://www.grosbill.com/
| wc
      13     758   27868
```
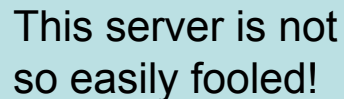
This server is not so easily fooled!

This method used to be common but has largely been replaced by responsive design
See: https://en.wikipedia.org/wiki/Responsive_web_design
https://www.spiria.com/en/blog/mobile-development/dynamic-serving-vs-responsive-web-design
https://www.clickseed.com/responsive-design-vs-separate-mobile-site-vs-dynamic-serving/

An anonymous representation

```
$ curl -is -A "Mozilla/5.0 (iPhone; CPU iPhone OS 8_4_1 like Mac OS X) AppleWebKit/600.1.4
(KHTML, like Gecko) CriOS/45.0.2454.68 Mobile/12H321 Safari/600.1.4" https://www.grosbill.com/
| less

HTTP/1.1 200 OK
Date: Tue, 23 Oct 2018 23:21:55 GMT
Server: Apache/2.4.25 (Debian)
X-Powered-By: PHP/5.6.23
Set-Cookie: versiontype=6; expires=Thu, 22-Nov-2018 23:21:55 GMT; Max-Age=2592000; path=/;
  domain=grosbill.com
Cache-Control: max-age=1
Expires: Tue, 23 Oct 2018 23:21:56 GMT
```
Vary: User-Agent
```
Transfer-Encoding: chunked
Content-Type: text/html; charset=UTF-8
Via: 1.1 google
Alt-Svc: clear

<!DOCTYPE html><html lang=fr><head><meta charset=utf-8 /> [deletia]
```

Cf. slide 13, where the representation had its own URI specified in `Content-Location`. This representation does not have its own separate, unique URI and can only be accessed by repeating this (or a similar) user-agent string.

302 CN, No Vary?

```
$ curl -I https://www.sephora.com/
HTTP/1.1 200 OK
Server: Apache
UFE-Page: Y
Content-Language: en-US
X-Content-Type-Options: nosniff
X-XSS-Protection: 1; mode=block
Content-Type: text/html;charset=UTF-8
Expires: Tue, 23 Oct 2018 03:23:48 GMT
Cache-Control: max-age=0, no-cache, no-store
Pragma: no-cache
Date: Tue, 23 Oct 2018 03:23:48 GMT
Connection: keep-alive
Set-Cookie: [deletia]
Strict-Transport-Security: max-age=31536000
```

```
$ curl -ILs -A "iphone"  https://www.sephora.com/
HTTP/1.1 302 Moved Temporarily
Server: AkamaiGHost
Content-Length: 0
Location: https://m.sephora.com/
Expires: Tue, 23 Oct 2018 03:23:58 GMT
Cache-Control: max-age=0, no-cache, no-store
Pragma: no-cache
Date: Tue, 23 Oct 2018 03:23:58 GMT
Connection: keep-alive
Set-Cookie: [deletia]
Strict-Transport-Security: max-age=31536000

HTTP/1.1 200 OK
Server: Apache
UFE-Page: Y
Content-Language: en-US
X-Content-Type-Options: nosniff
X-XSS-Protection: 1; mode=block
Content-Type: text/html;charset=UTF-8
Date: Tue, 23 Oct 2018 03:23:58 GMT
Connection: keep-alive
Set-Cookie: [deletia]
Strict-Transport-Security: max-age=31536000
```

sephora.com prohibits caching but a `Vary: User-Agent` response header could be used, cf. 7.1.4 of RFC 7231:

"Likewise, an origin server might use Cache-Control directives (Section 5.2 of [RFC7234]) to supplant Vary if it considers the variance less significant than the performance cost of Vary's impact on caching."

Slightly off-topic, but note: "Cache-control: max-age" trumps "Expires", so "Expires" is redundant.
"Pragma: no-cache" and "Connection: Keep-alive" are deprecated.