HTTP Protocol and Server-Side Basics

Web Programming

Uta Priss
ZELL, Ostfalia University

2013

Outline

Manual interaction with the HTTP protocol

Manual interaction with a webserver is possible by typing

```
telnet www.somename.ac.uk 80
```

then typing

```
GET / HTTP/1.0
```

and then pressing enter twice.

Format of the first line of an HTTP request:

```
METHOD space Request-URI space HTTP-Version \n\n
```

Methods are GET, HEAD, OPTIONS, POST, DELETE, etc.

The Request-URI is usually the path and name of the document (eg. "/" or "/somedir/index.html").

The HTTP-Version is currently either 1.0 or 1.1.

A typical HTTP request with GET

The HTTP request is normally generated by the user's browser. A typical request with "GET" is

```
GET /cgi-bin/somefile.cgi HTTP/1.0
CONNECTION: Keep-Alive
USER-AGENT: Mozilla/4.0 (compatible; MSIE 5.22)
PRAGMA: no-cache
HOST: www.somename.ac.uk
ACCEPT: image/gif, image/x-xbitmap, image/jpeg, */*
QUERY_STRING: name=Mary&comments=Hello
```

A typical HTTP request with POST

```
POST /cgi-bin/somefile.cgi HTTP/1.0
REFERER: http://www.somename.ac.uk/cgi-bin/someform.html

CONNECTION: Keep-Alive
USER-AGENT: Mozilla/4.0 (compatible; MSIE 5.22)
HOST: www.somename.ac.uk
ACCEPT: image/gif, image/x-xbitmap, image/jpeg, */*
CONTENT-TYPE: application/x-www-form-urlencoded
CONTENT-LENGTH: 42
name=Mary&comments=Hello
```

A typical answer from a webserver

```
HTTP/1.1 200 OK
Date: Thu, 18 Nov 2004 15:41:04 GMT
Server: Apache/2.0.40 (Red Hat Linux)
Accept-Ranges: bytes
X-Powered-By: Open SoCks 1.2.0000
Last modified: Thu, 18 Nov 2004 15:41:04 GMT
Connection: close
Content-Type: text/html; charset=ISO-8859-1
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//
"http://www.w3.org/TR/html4/loose.dtd">
<html>

...
```

Different possible first lines

```
HTTP/1.1 200 OK
HTTP/1.1 404 Not Found
HTTP/1.1 302 Found
HTTP/1.0 500 internal error
```

(302 indicates a redirection):

Low-level interaction with the HTTP protocol

Normally users need not interact with the HTTP protocol on a low level because browsers do all the work. But there are reasons why programmers may need more low level access to the web:

- Testing of websites from the command-line. For example, if a whole website is to be checked for dead links this can be conveniently done from the command-line.
- Testing of the security of one's own website (especially for scripts) by trying to break in.
- Web crawlers, spiders, robots: programs that navigate the web and collect information, for example, for creating search engines.

What tools to use

Most modern programming languages contain libraries that allow for direct interaction with the HTTP protocol (such as libwww or socket libraries).

There are plugins for browsers (such as Firefox) which support viewing and editing of HTML headers.

Environment Variables

- Environment variables are a means for server-side web languages to exchange information between the server and the client.
- They control the information which is disclosed by either server or client or which refers to a specific HTTP request.
- Not all environment variables are always available.

Client-Side Information

Environment variables which contain client-side information:

HTTP_USER_AGENT	type and version of the client's browser client
HTTP_ACCEPT	accepted MIME types
REMOTE_ADDR	IP address of the client

Sever-Side Information

Environment variables which contain server-side information:

SERVER_SOFTWARE	software used for webserver (e.g. Apache)
SERVER_NAME	name of server
SERVER_PROTOCOL	protocol used by server (e.g. HTTP/1.1)

Request-Specific Information

Environment variables which contain information about a specific HTTP request:

SCRIPT_NAME	the URL of the script
REQUEST_METHOD	usually either GET or POST
QUERY_STRING	form parameters as part of URL (only for GET)
CONTENT_LENGTH	(only for POST)
HTTP_COOKIE	content of a cookie
REMOTE_USER	username - if authentication is used
HTTP_REFERER	URL of previous page

Can environment variables be trusted?

Can environment variables be trusted?

- The value of environment variables cannot be trusted.
- For example: clients can lie about which type of browser they use.
- HTTP_REFERER should never be used for security because it can be modified by a client.
- For highly secure applications, the content of cookies should be encrypted because otherwise it can be modified by a client.

Why can environment variables not be trusted?

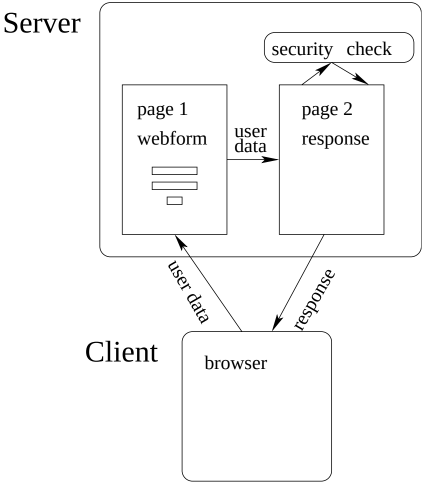
Why can environment variables not be trusted?

Because they are usually transmitted as plain text via the HTTP protocol. A client can use a scripting language or browser plugin to read and modify environment variables.

A challenge for server-side web languages

Session management poses a challenge for server-side web languages because, in contrast to stand-alone applications, server-side applications transfer data between servers and clients. A server-side application does not have full control over the data at the client's computer. A client can delete, modify or add to any data. Therefore client data cannot be trusted.

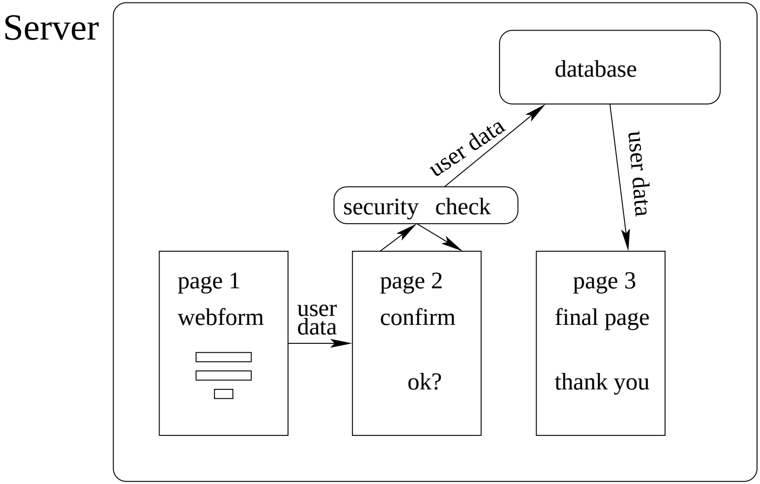
A webform and response without sessions



Using session IDs

- Sessions are created to minimise the risk of data corruption.
- All user data is stored on the server and associated with a session ID. Data retrieval is based on the session ID.
- A security check must be performed on any data, which a user enters, before the data is stored on the server.
- It is not necessary to perform security checks on data that is retrieved from the server. This is in contrast to CGI applications without sessions: in that case a security check must be performed on user data every time the data is used!

Webform, confirmation and response

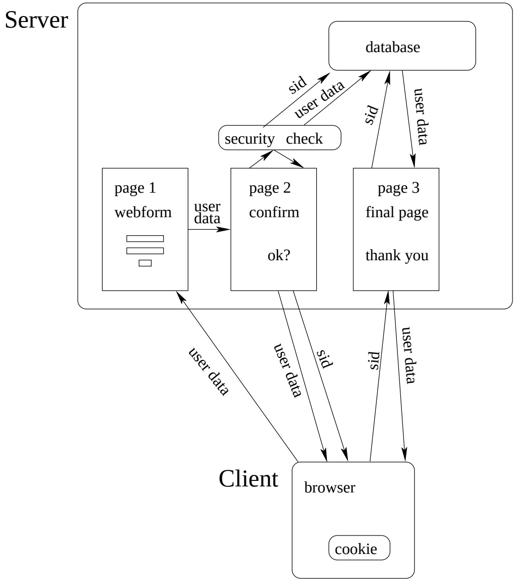


Session IDs

Session IDs are passed from page to page either

- by using hidden HTML form text or
- by using the query string or
- by using cookies (on the client computer).

The session IDs should be generated by the server and should be long and complicated enough to deter users from manual tampering with the IDs.



Server
database
sid
user data
sid
user data
security check
page 1
webform
user
data
page 2
confirm
ok?
page 3
final page
thank you
user data
user data
sid
sid
user data
Client
browser
cookie

Storing user data

The user data, which is associated with a session ID can be stored on the server

- in a file
- in a database.

This approach reduces the number of security checks that must be performed on user data.

Storing user data

The user data, which is associated with a session ID can be stored on the server

- in a file
- in a database.

This approach reduces the number of security checks that must be performed on user data.

Further measures can be taken for applications which require high security:

- the session can be password protected;
- the cookie data can be encrypted;
- the whole session can be encrypted using SSL.

Cookies

Cookies are transmitted between client and server via the HTTP_COOKIE environment variable.

This means that any cookie data needs to be **written** before the HTML header is sent.

Cookies

Cookies are transmitted between client and server via the HTTP_COOKIE environment variable.

This means that any cookie data needs to be **written** before the HTML header is sent.

Cookie data can be **read** at any time.

Workflow for cookies

- First: read any existing cookie.
- Second: compile cookie data and send it.
- Third: Print the HTML header.

Workflow for cookies

- First: read any existing cookie.
- Second: compile cookie data and send it.
- Third: Print the HTML header.

Note:
it is counterintuitive to first read the cookie and then write it!

Example: creating a counter with a cookie

- Read the cookie data.

Example: creating a counter with a cookie

- Read the cookie data.
  If no prior cookie is received, set counter $= 0$.

Example: creating a counter with a cookie

- Read the cookie data.
  If no prior cookie is received, set counter $= 0$.
  Read counter value and increase by 1.

Example: creating a counter with a cookie

- Read the cookie data.
  If no prior cookie is received, set counter = 0.
  Read counter value and increase by 1.
- Write the cookie.
- Print the HTML header.