Dynamic Web Development

Lecture 1 – What is a web server

What is a Web Server?

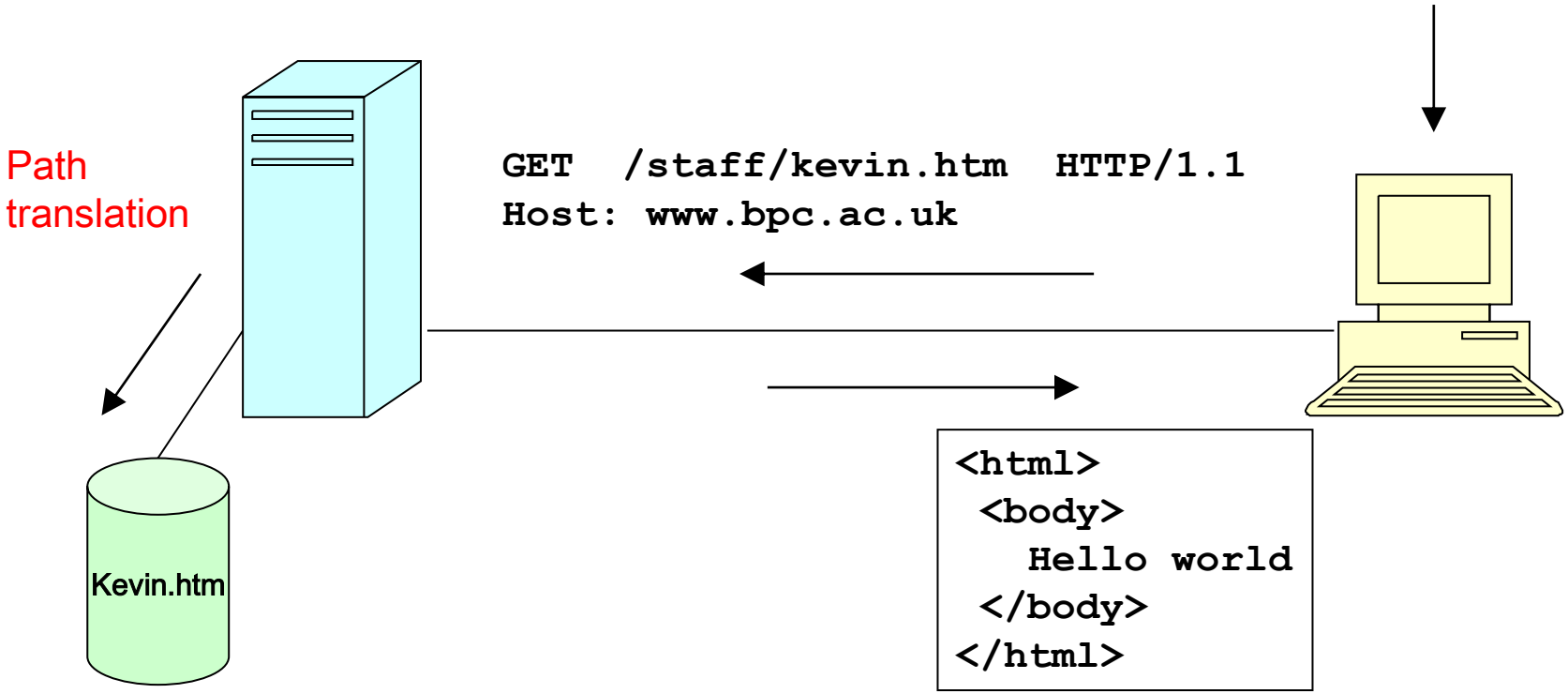
Hardware

A computer that accepts HTTP requests from client software (web browser) and sends back the web page that was asked for.

Software

A computer program that does the same job.

http://www.bpc.ac.uk/staff/kevin.htm



`C:\inetpub\wwwroot\staff\kevin.htm`

History 1

1980 Tim Berners-Lee developed programs to store information with random links.

1989 He proposed idea that any piece of information on a network could be identified by a “Universal Document Identifier” (UDI)

1990 While working at CERN, he wrote a program called "WorldWideWeb", a point and click hypertext editor which ran on the "NeXT" microcomputer.
Also wrote the first Web server

History 2

1991 Ideas spread throughout CERN, via their internal network, then to the hypertext and NeXT communities via the internet.

The specifications of UDIs (now URIs), HyperText Markup Language (HTML) and HyperText Transfer Protocol (HTTP) published on the first server in order to promote wide adoption and discussion.

Web clients were developed for other platforms: Erwise, Viola, Cello and Mosaic.

History 3

Other people set up their own web servers – mainly in academic and research communities.

Between the summers of 1991 and 1994, the load on the first Web server ("info.cern.ch") rose steadily by a factor of 10 every year.

1993 Industry and the rest of the world starts taking notice.

1994 Berners-Lee formed the World Wide Web Consortium (W3C).

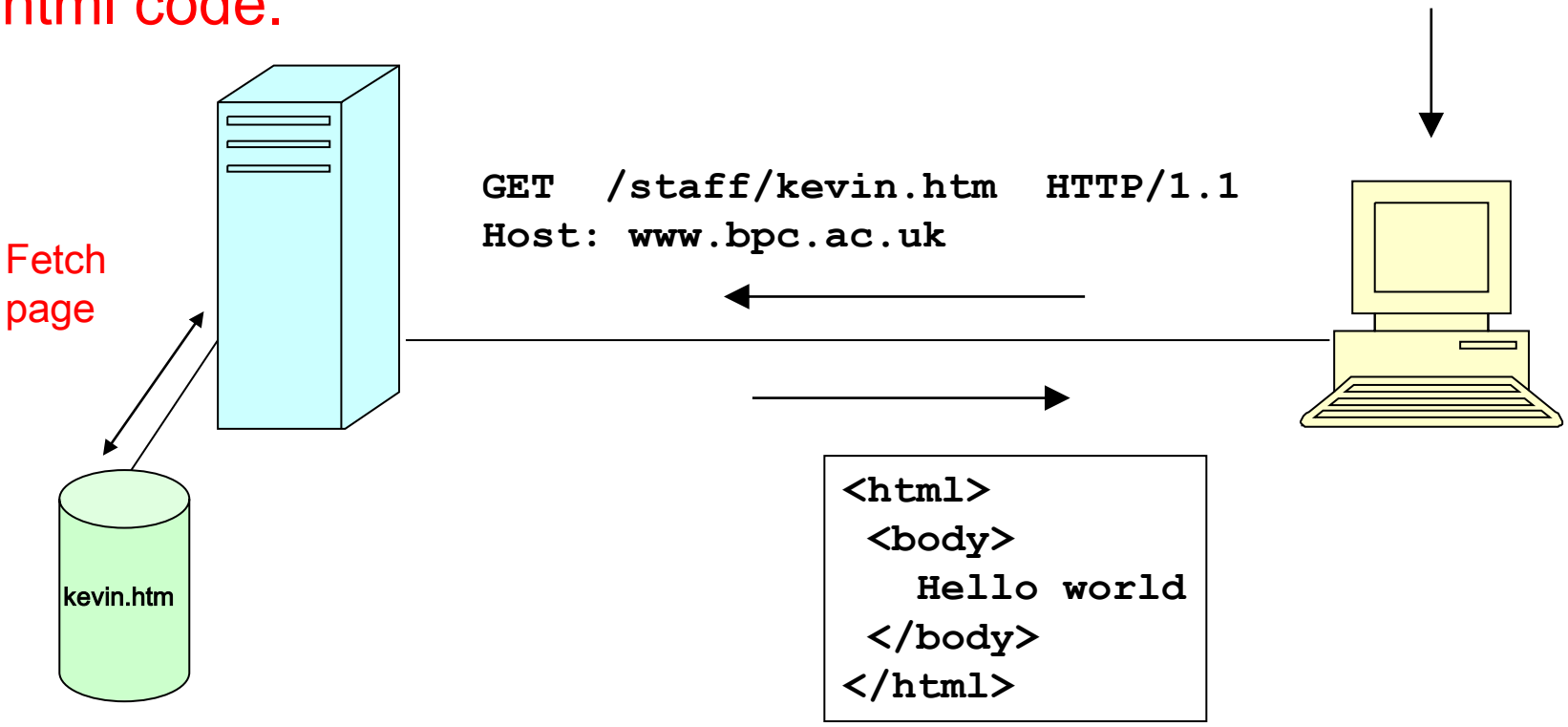
Generic Features of a Web Server

- Listen for http requests on i/o port 80 and respond to them.
- Serve up static and dynamic web pages.
- Keep log files of activity.
- Support plug-in software modules.
- Compression of web pages.
- Virtual hosting of many sites on one IP address.
- HTTPS support to allow encrypted connections to server.
- Bandwidth throttling to limit speed of responses.

Static Web Pages

File on server contains fixed html code.

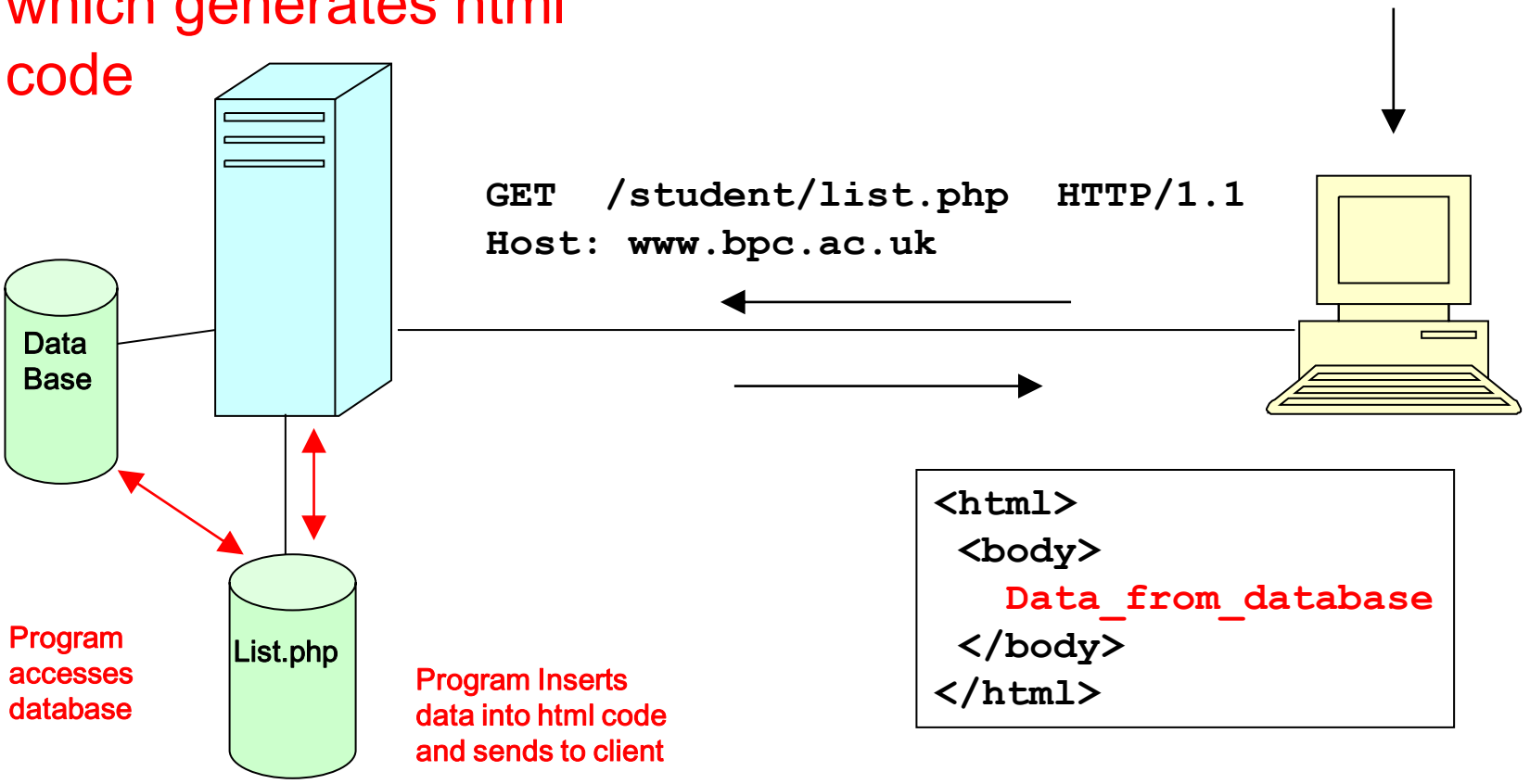
http://www.bpc.ac.uk/staff/kevin.htm



Dynamic Web Pages

File on server contains a program which generates html code

http://www.bpc.ac.uk/student/list.php



Scripting Languages

Server Side

Active Server Pages (ASP Microsoft)
(uses VBScript, JScript)
Coldfusion (Macromedia)
JavaServer Pages (JSP)
Server Side Includes (SSI)
PHP (Open source)
Perl (Open source)

Client Side

Javascript (Netscape)
Jscript (Microsoft)
Adobe Flash (Web Animations)
Adobe Shockwave (Multimedia)

Server side Scripting 1

Used to be done via:

- Common Gateway Interface (CGI) protocol
- Programs written in a standard language (e.g. Perl, C++)
- Stored in folder called cgi-bin (CGI binaries)
- Executed by the Operating System.

`http://www.kev.com/cgi-bin/namecheck.c`

Problems:

- Poor coding can lead to security holes.
- Operating system itself could be compromised.
- New process created every time program is executed. Could overwhelm server.

Server side Scripting 2

Now done by:

- Programs in languages designed for purpose (e.g. PHP, ASP)
- Can be stored in secure location on server
- Executed by the web server.

`http://www.kev.com/register.php`

Note:

- Better (but not perfect) security.
- Only web server might be compromised.
- Language module must be installed on the web server.
- Language designed to make minimal use of system resources.

Client side Scripting

- Can be used to modify web pages to suit local settings on users computer.
- Language must be installed in browser.
- Difficult to ensure standard behaviour.
- Could be used to introduce malware into system.

Popular Server Software

Apache HTTP server

- Runs on 70% of servers on internet.
- Free, open source software
- Available on Unix, Windows, Mac OS X and others.
- Small basic core but with many add on modules for specific functions.

Microsoft Internet Information Server (IIS)

- Runs on 20% of servers on internet.
- Free (with Windows) but not open source.
- Available on Windows only.
- Easy to set up and configure

Index files

http://www.kevweb.co.uk

This URL is ‘pointing to’ the home directory of the web server software on the host machine www.kevweb.co.uk.

Most servers will default to displaying a file called index.htm.

There may be an ordered list of filenames that it will check for.

In the absence of a default file, some servers may display the contents of the server’s root directory.

This may create a security risk, and should be disabled.

Index of /wp-includes

Name	Last modified	Size	Description
Parent Directory	22-Feb-2007 20:57	-	
author-template.php	22-Feb-2007 04:47	6k	
bookmark-template.php	22-Feb-2007 04:47	11k	
bookmark.php	22-Feb-2007 04:47	4k	
cache.php	22-Feb-2007 04:47	11k	
capabilities.php	22-Feb-2007 04:47	12k	
category-template.php	22-Feb-2007 04:47	8k	
category.php	22-Feb-2007 04:47	9k	
class-IXR.php	22-Feb-2007 04:47	27k	
class-pop3.php	22-Feb-2007 04:47	21k	
class-snoopy.php	22-Feb-2007 04:47	37k	
classes.php	22-Feb-2007 04:47	22k	
comment-functions.php	05-Jan-2007 23:31	30k	
comment-template.php	22-Feb-2007 04:47	10k	
comment.php	22-Feb-2007 04:47	28k	
compat.php	22-Feb-2007 04:47	3k	
cron.php	22-Feb-2007 04:47	5k	
default-filters.php	22-Feb-2007 04:47	6k	
deprecated.php	22-Feb-2007 04:47	18k	

Unix Directory Structures

Windows Path

```
C:\Documents and Settings\Kevin\workfile.doc
```

Unix Path

```
/home/kevin/workfile.doc
```

Which is why URL's have forward slashes.

HTTP and HTML were developed by Tim Berners-Lee on an operating system called NEXTSTEP, a modified version of BSD-UNIX.

(As were DOOM and QUAKE, apparently).
http://rome.ro/2006/12/apple-next-merger-birthday.html

HTTP Response Headers

Once the server locates the file, it will send it back to the client, preceded by some HTTP response headers.

These provide the browsers with information about the arriving file, including its media type (Content type or MIME type).

This enables the browser to load the correct software to handle the file.

The web server will send the HTML for a web page in a HTTP packet(s).

Each resource (image, sound clip, video clip, javascript script etc) for that page will be sent in a separate HTTP packet.

A large file is usually split into several packets.

MIME File Types

MIME (Multipurpose Internet Mail Extension) was originally developed for sending attachments in email.

Borrowed by HTTP developers to help identify the different types of file that could be requested from a web server.

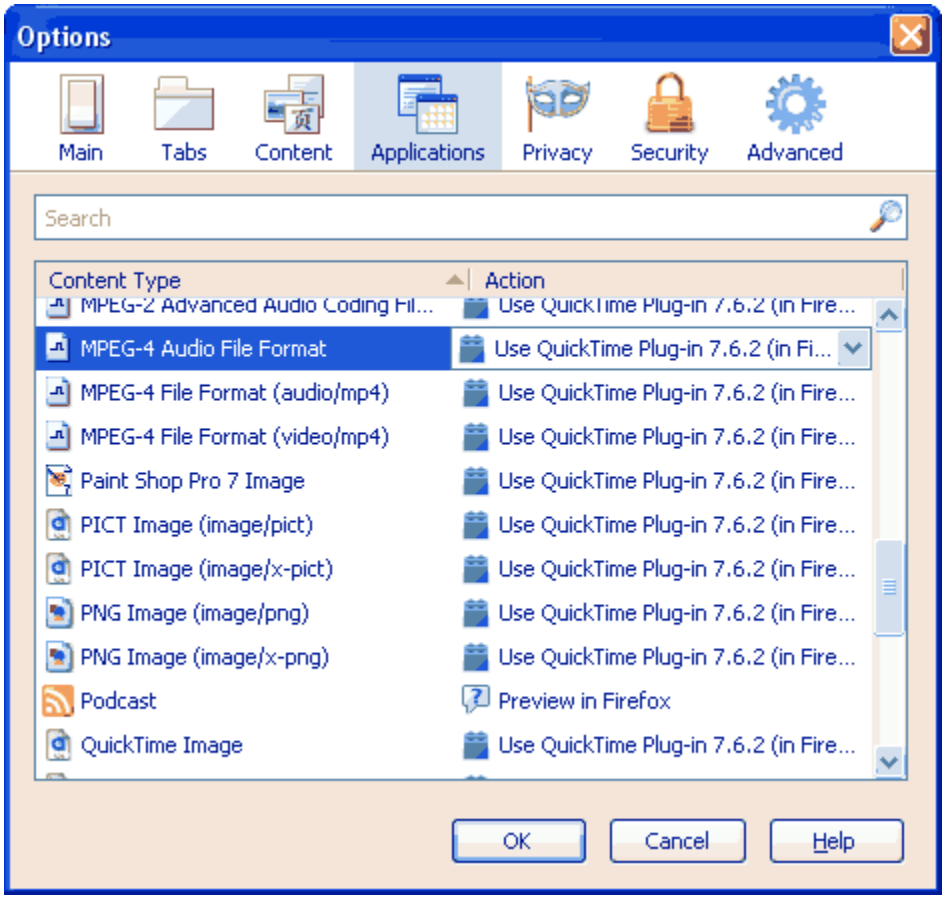
You need to ensure that the server and browser is configured to recognise each MIME type you are interested in using.

They all require: type, subtype and extension.

MIME types

application/excel	.xl
application/msword	.doc, .dot, .word
application/pdf	.pdf
audio/x-wav	.wav, .aiff
audio/x-mpeg	.mp3
image/gif	.gif
image/jpeg	.jpg, .jpeg, .jfif
text/html	.html, .htm
text/plain	.txt
video/quicktime	.mov
video/x-ms-wmv	.wmv

Firefox MIME type Configuration



Handling MIME Types in Internet Explorer

Internet Explorer sets the file extension of a downloaded file based on a few pieces of information available from the HTTP server and from the file itself. Internet Explorer confirms or resets the file extension of a downloaded file to ensure that if the file is executed, Microsoft Windows launches the proper application to handle it.

Internet Explorer first consults the Content-Type and Content-Disposition headers for the MIME type from the HTTP server, if they exist.

Next, Internet Explorer consults the registry entry for that MIME type and looks at the Extension key for the correct file extension.

Internet Explorer then ensures that the file has that extension, before placing it in the local browser cache.

If Internet Explorer knows the Content-Type specified and there is no Content-Disposition data, Internet Explorer performs a "MIME sniff," scanning the first 200 bytes of the file to determine if the file structure matches any known MIME types.

If the MIME sniff detects a MIME type known to Internet Explorer, and the file has not been loaded by a mimefilter already, Internet Explorer sets that file extension before placing the file in the local browser cache.

Lastly, if there is no Content-Type or Content-Disposition data, and the MIME sniff does not recognize a known MIME type, the file extension is set to the same extension as the URL used to download the file.

Executing a Downloaded File

In Windows XP SP2 and later, after downloading a file, Internet Explorer consults the registry to find out which ProgID and CLSID is associated with the MIME type of the file.

The browser then finds the ProgID and CLSID associated with the file's extension. If these ProgIDs do not match, and the CLSIDs do not match, Internet Explorer prompts the user before executing the file, as a safety precaution.

Internet Explorer launches only the program associated with the MIME type of the downloaded file and not the program associated with the file extension (if they are different programs).

In Windows XP SP2 and later, if the MIME handler rejects the file, and the MIME handler and shell extension handler don't match, Internet Explorer shows an error dialog.

If the MIME handler rejects the file with an error code of INET_E_CANNOT_LOAD_DATA, then Internet Explorer does not load the shell extension handler even if the ProgIDs or CLSIDs match. This is an extra security measure designed to prevent the shell extension handler from executing corrupt files.

If the file is marked as "content-disposition=attachment" in the HTTP header, Internet Explorer does not look at the MIME handler, shows a file download prompt, and, if instructed to run the attachment, it executes the file.

http://msdn.microsoft.com/en-us/library/ms775148(VS.85).aspx

Summary

What is a web server?

What translation does a web server carry out to every http request?

Who wrote the first web server?

What year was the W3C founded?

List the generic features of a web server.

What is a static web page?

What is a dynamic web page?

Name the server side languages.

Name the client side languages.

Name one problem with server side scripting.

Name one problem with client side scripting.