Edinburgh Napier UNIVERSITY

CSN11121
System Administration and Forensics
Week 5: Essential Apache and Log Analysis

Module Leader: Dr Gordon Russell
Lecturers: G. Russell, R.Ludwiniak

Aliases: CSN11122 (Distance Learning Version)

Edinburgh Napier UNIVERSITY

This lecture

- Configuring Apache
- Log analysis
- Discussions

Edinburgh Napier
UNIVERSITY

Configuring Apache

Edinburgh Napier UNIVERSITY

Apache

- Very well known and respected http server.
- Used commercially.
- Freely available from http://www.apache.org
- Plenty of plugins.
- Relatively easy and flexible to configure.
- Fast and Reliable.

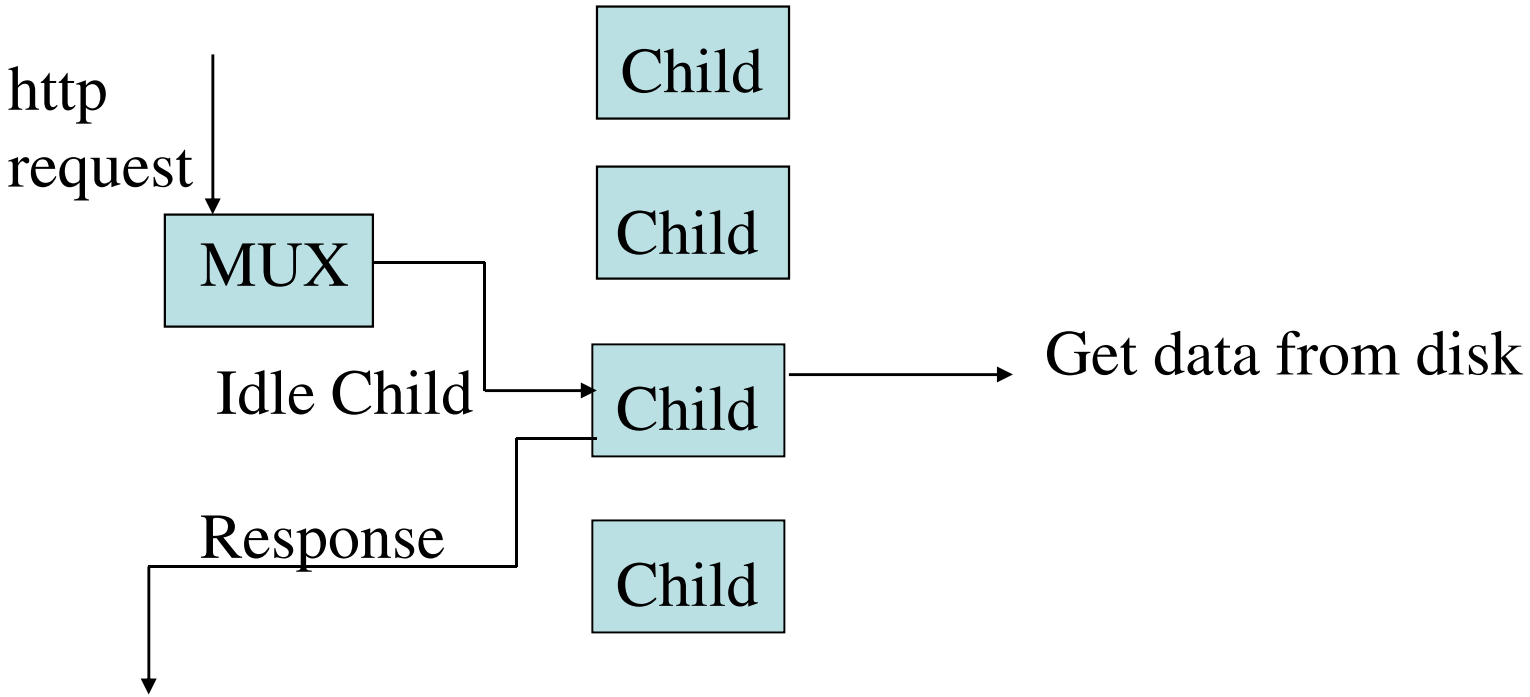
Edinburgh Napier UNIVERSITY

Server Architectures

- In most designs of server, you either use
  - Threaded model
  - Forking model
  - Asynchronous Architecture
- A threaded model needs special OS support to provide lightweight threads. Not used in Apache for security and reliability reasons.
- Forking means that each new request which arrives is handled by a whole process. This is the Apache way.
- Asynchronous. Some web servers exist with this model, where one process handles everything with complex IO code. Good for fast processing of simple web pages.

Edinburgh Napier UNIVERSITY

Apache Forking Model



Edinburgh Napier UNIVERSITY

Initial Settings

StartServers	8
MinSpareServers	5
MaxSpareServers	20
MaxClients	150
MaxRequestsPerChild	1000

- These options are important, but often the least likely to be changed from the defaults!

Edinburgh Napier UNIVERSITY

Important Files

- /etc/init.d/httpd – the server control script
- /etc/httpd/conf/http.confg – the main conf file.

- Remember when changing the configurations it is only reread on a server reload or restart.
- Errors and other details are logged by default in /var/log/httpd/ as access_log, error_log, as suexec.log.

Edinburgh Napier UNIVERSITY

Reload or Restart

- Reload is the best option to use.
- With a reload, apache checks your configuration file, and switches to it only if it contains no errors.
- If it has errors, it keeps using the old configuration.
- This allows you to reconfigure a server with no downtime.
- Restart shuts down then starts the server…
- Look in the error log for help (e.g. /var/log/httpd/error_log), or syslog (e.g. /var/log/messages).
- Remember to use the service command for this:
  - Service httpd start|stop|reload|restart|status
- You can easily make errors in the config file. You can check for errors using
  - Service httpd configtest

Edinburgh Napier UNIVERSITY

Mimic a Browser

- To understand how a sever is running is it sometimes useful to make requests at the keyboard of a server and see the results as text.
- Telnet can do this, so long as you have learned some basic HTTP commands.
- The two important ones are:
  - HEAD – Give information on a page.
  - GET – Give me the whole page.

Edinburgh Napier UNIVERSITY

- In HTTP 1.1 we can use virtual hosts.
- This allows multiple hosts to share a single server.
- Each host has a different name.
- The name of the host you want to answer a query is given as part of a page request.
- This is only supported in HTTP 1.1 and beyond.

Edinburgh Napier UNIVERSITY

$ telnet linuxzoo.net 80
HEAD / HTTP/1.1
Host: linuxzoo.net

HTTP/1.1 200 OK
Date: Mon, 01 Nov 2008 15:06:44 GMT
Server: Apache/2.0.46 (Red Hat)
Last-Modified: Fri, 29 Oct 2008 14:47:22 GMT
ETag: "4981dd-920-22ea7280"
Accept-Ranges: bytes
Content-Length: 2336
Content-Type: text/html; charset=UTF-8

Edinburgh Napier UNIVERSITY

$ telnet linuxzoo.net 80
HEAD / HTTP/1.1
Host: db.grussell.org

HTTP/1.1 200 OK
Date: Mon, 01 Nov 2008 15:08:52 GMT
Server: Apache/2.0.46 (Red Hat)
Last-Modified: Thu, 21 Oct 2008 09:12:33 GMT
ETag: "3c8066-a37-86c9a240"
Accept-Ranges: bytes
Content-Length: 2615
Content-Type: text/html; charset=UTF-8

Edinburgh Napier UNIVERSITY

VirtualHosts

- The sharing of a single IP to provide multiple hostnames is well supported in Apache.
- The part of the conf file which handles this is called <VirtualHost>
- Each part holds a list of hostnames it can handle
- The first host found in the file is always considered the default, so if no VirtualHost section matches the first block is done instead.

Edinburgh Napier UNIVERSITY

```
<VirtualHost>
  ServerAdmin me@grussell.org
  DocumentRoot /home/gordon/public_html
  ServerName grussell.org
  ServerAlias www.grussell.org grussell.org.uk
  ErrorLog logs/gr-error_log
  CustomLog logs/gr-access_log combined
</VirtualHost>
```

Edinburgh Napier UNIVERSITY

public_html

- Where apache runs on a server used by many different servers, it would be useful for each user to be able to build their own web pages which the server could serve.
- But the virtualhost configuration takes only a single document root, and each user has their own directories in /home.
- You could make the root /home
  - All of the files in /home would be accessible, not just web pages.
  - It's a bit disgusting…
- Instead, apache supports web pages appearing in a users home directory, under the subdirectory public_html.

public_html access

- Urls of the form
  - http://linuxzoo.net/~gordon/file.html
- Refer to
  - /home/gordon/public_html/file.html
- This feature must first be switched on in httpd.conf.
- To activate it, find the line
  - UserDir disable
- Then either delete the line, or put “#” (the comment character) in front of it.
- Then find the following line and delete the ‘#’ character.
  - #UserDir public_html
- Remember to reload the server.

Edinburgh Napier UNIVERSITY

Linuxzoo tutorials

- Each time you book a linuxzoo machine, you will likely get a different IP and hostname.
- Each time you come in, check your hostname with “hostname”.

$ hostname

host-5-5.linuxzoo.net

- In this example, virtual hosts vm-5-5.linuxzoo.net, as well as host-5-5 and web-5-5 will be proxied to your machine.
- Warning: If the server on which your virtual machine fails, you will be moved to a different machine and a different IP. You need to check your hostname when you boot!

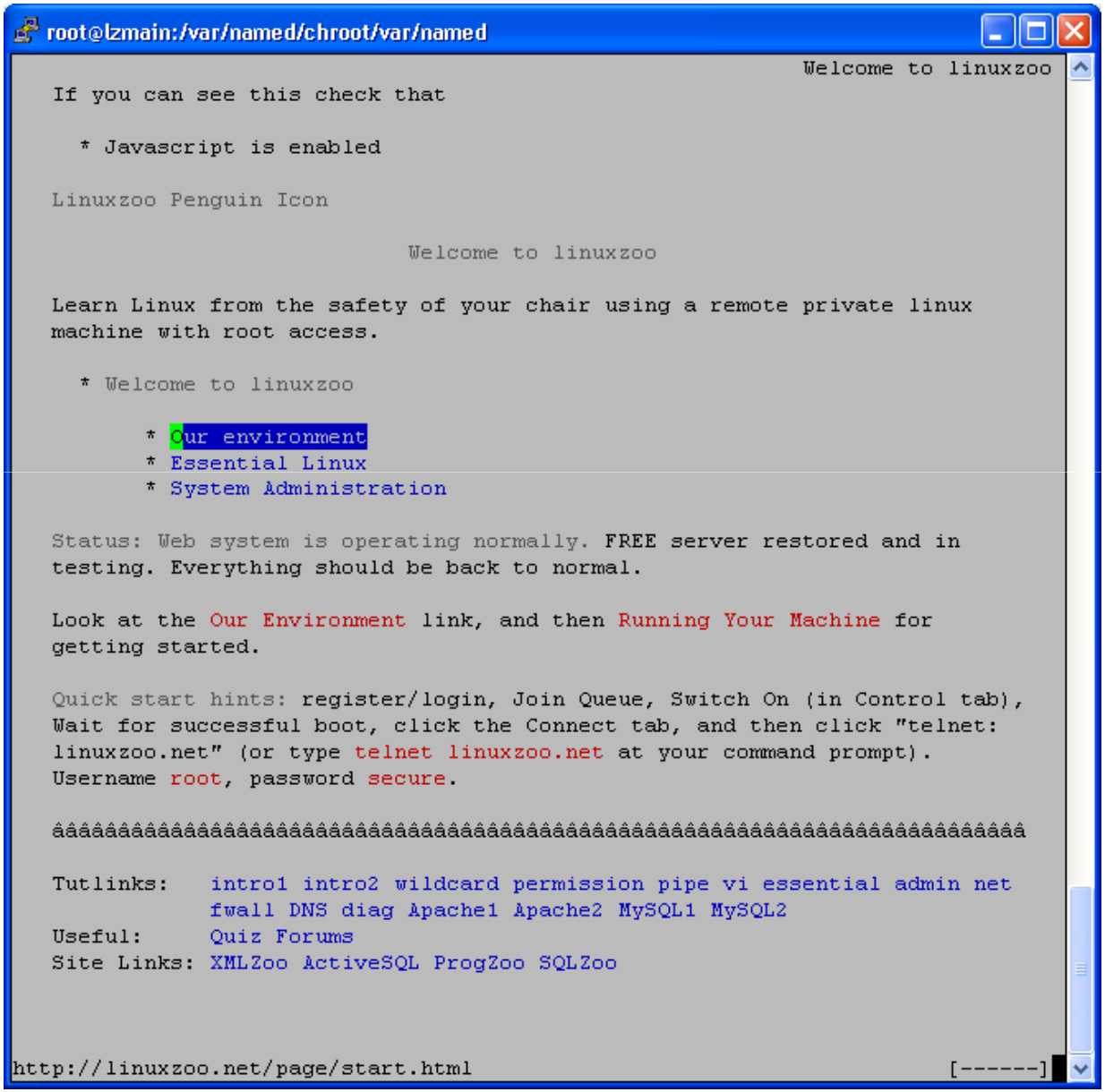
Edinburgh Napier UNIVERSITY

Web access from the prompt

- The prompt is fast and convenient for admin purposes, but when you are debugging http sometimes “telnet” is not sufficient.
- There are a few other tools you can use at the prompt.
  - elinks
  - lwp-request
  - wget
- However, there is no simple replacement for actually using a real browser to check your pages.

$ elinks http://linuxzoo.net

Edinburgh Napier UNIVERSITY



Edinburgh Napier UNIVERSITY

Copy http to your directory

- lwp-request http://linuxzoo.net > file.html
  - The data is obtained and then printed to the screen.
  - In this case that is redirected to file.html
- wget http://linuxzoo.net

```
$ wget http://linuxzoo.net
--19:20:11--  http://linuxzoo.net/
Resolving linuxzoo.net... 146.176.166.1
Connecting to linuxzoo.net|146.176.166.1|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4785 (4.7K) [text/html]
Saving to: `index.html'
100%[======================================>] 4,785       --.-K/s   in 0s
19:20:11 (304 MB/s) - `index.html' saved [4785/4785]
```

Edinburgh Napier UNIVERSITY

SELinux and Apache

- SELinux secures apache, and SELinux security of files in public_html is by default quite strong.
- Check if SELinux allows files to be published from public_html by
  - getsebool httpd_read_user_content
  - If this is 0 then publishing files is forbidden.
- Set SELinux to allow public_html publishing using:
  - setsebool -P httpd_read_user_content 1
  - This may take 20 or more seconds. Be patient.
  - The setting will be forgotten if you get a new image in the linuxzoo interface.
- SELinux requires the file security (shown by ls –Z) to be:
  - unconfined_u:object_r:httpd_user_content_t:s0
  - However this should happen automatically provided you create files in public_html
  - You can set the type of say filename.html (but remember you should not have to) using:
    - chcon –t httpd_user_content_t filename.html

Edinburgh Napier
UNIVERSITY

Log Analysis

Edinburgh Napier UNIVERSITY

Logs

- Apache produces two types of log files
  - Error Logs
  - Access Logs
- Error logs are useful for debugging
- Access logs are excellent for monitoring how your site is being used.
  - Fun for people who have hobby sites
  - Life or death if your business relies on the web site.

Edinburgh Napier UNIVERSITY

Where are the logs

- Normally they go to /var/log/httpd/access_log and error_log
- In a virtual host we set them to what we liked:

```
<VirtualHost>
…
    ErrorLog logs/gr-error_log
    CustomLog logs/gr-access_log combined
</VirtualHost>
```

Edinburgh Napier UNIVERSITY

Logging in /var/log/http access file

- The normally used log format is called “combined”.
- It contains significant amounts of information about each page request.
- Specifically, the log format is:

%h %l %u %t %r %>s %b Referrer UserAgent

Edinburgh Napier UNIVERSITY

%h %l %u %t %r %>s %b Referrer UserAgent

- h – IP of the client
- l – useless ident info
- u – username in basic authentication
- t – time of request
- r – the request itself
- s – The response code (e.g. 200 is a successful request)
- b – size of the response page
- Referrer – who the client things told it to come here
- User Agent – identification info of the browser

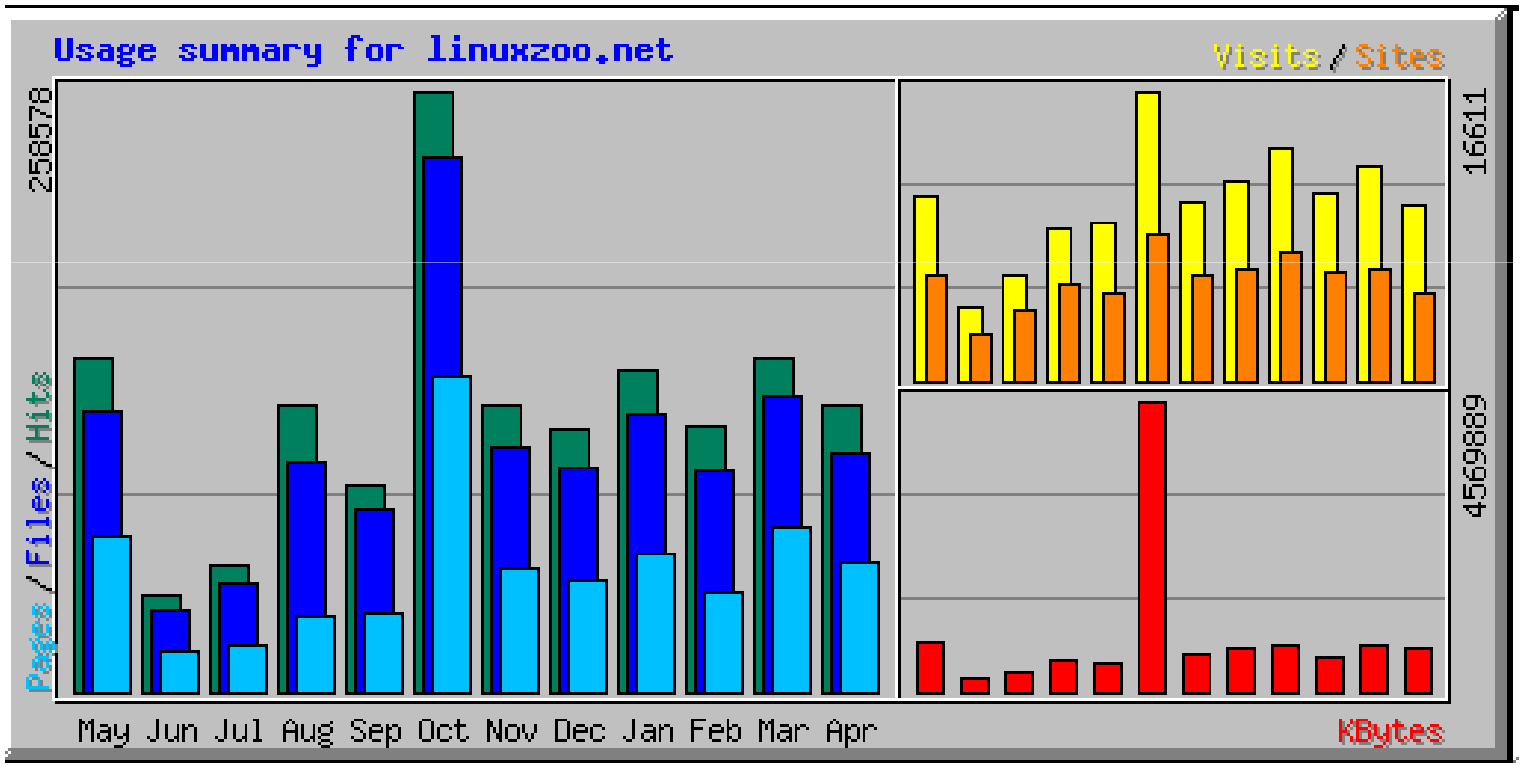
Edinburgh Napier UNIVERSITY

Analysing the log

- The log is useful in itself for checking the proper function of the server.
- However, traffic analysis is also valuable.
- There are a number of tools available to do this.
- One of the best free ones is webaliser.

Edinburgh Napier
UNIVERSITY

Webaliser Summary



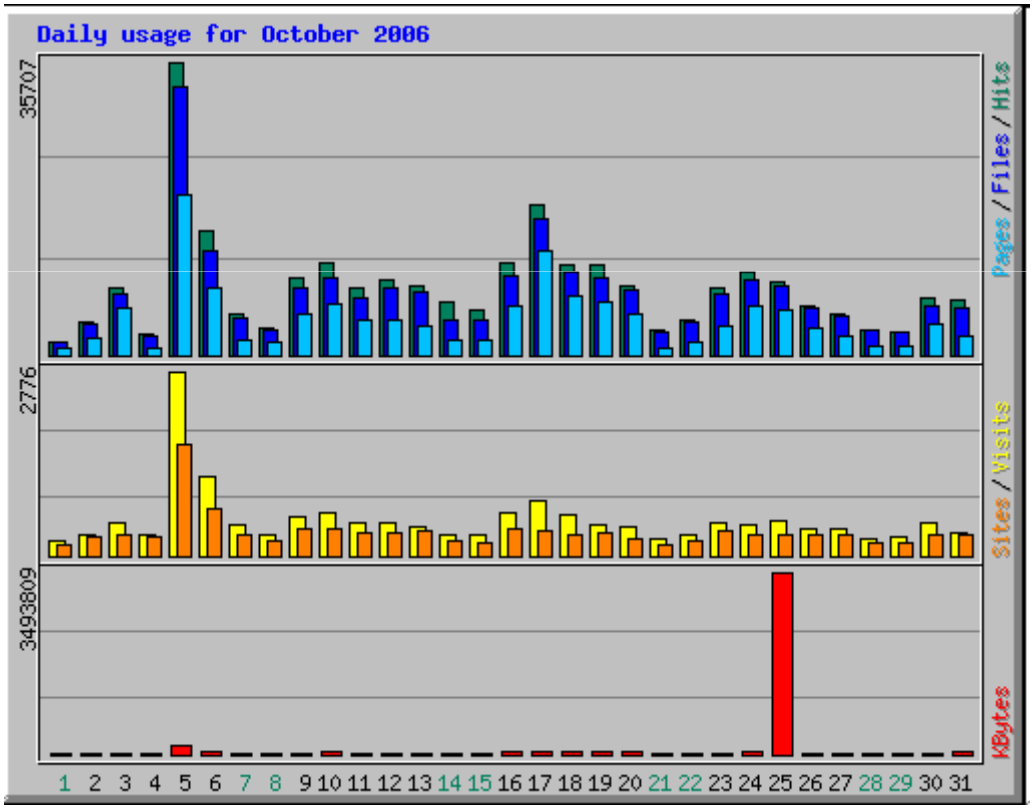
Edinburgh Napier UNIVERSITY

Analysis

- The summer is quiet for linuxzoo.
- Students are enthusiastic in October…
- After that it settles down to “kept busy”.

Edinburgh Napier UNIVERSITY

Per day activity – October

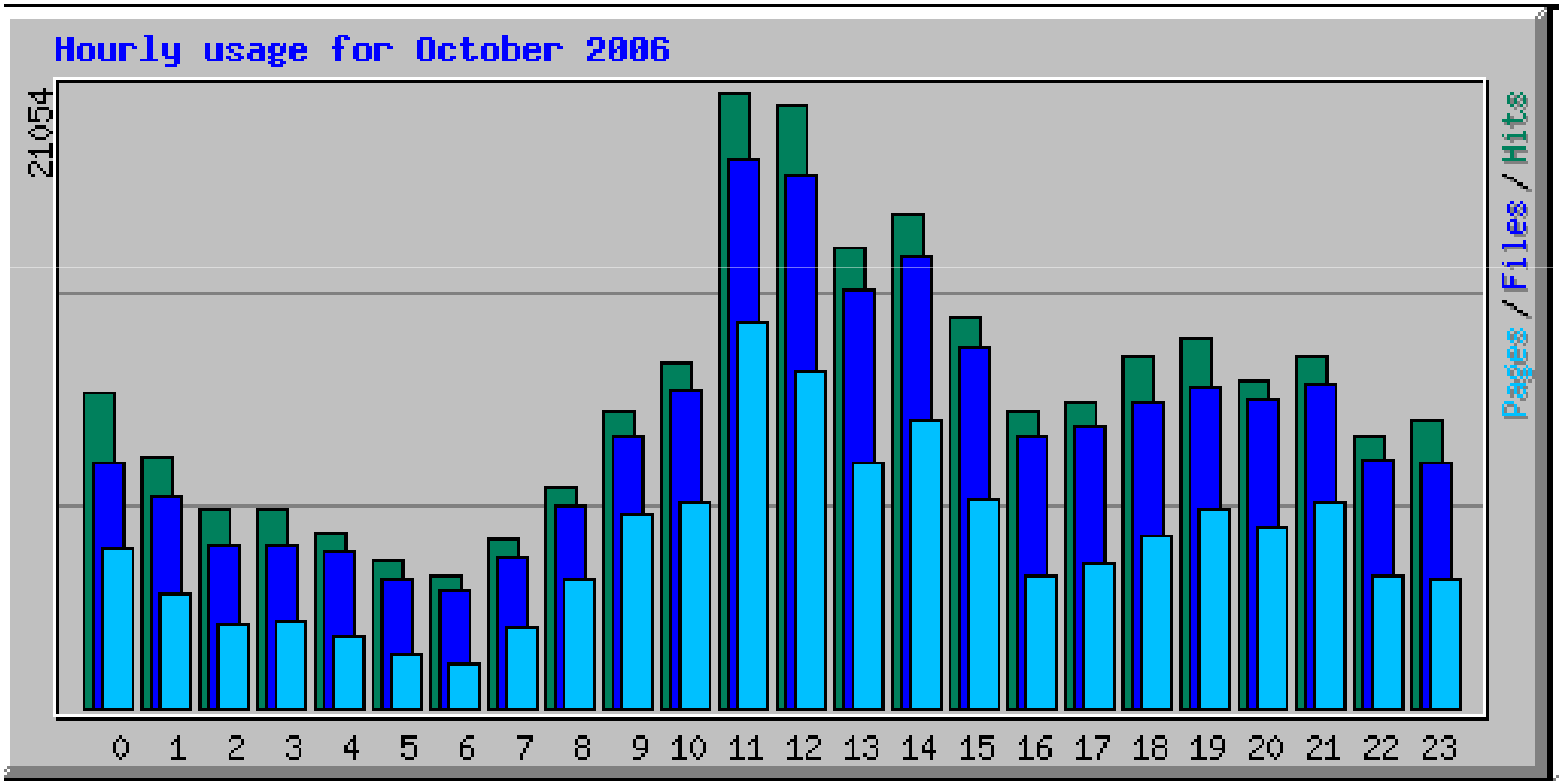


Edinburgh Napier UNIVERSITY

- I wonder which day was the first tutorial?
- Look at the 7 day oscillations. This is common in many web sites.
- Who stole all my web site data on the 25th?

Edinburgh Napier UNIVERSITY

Hour analysis – October



Edinburgh Napier UNIVERSITY

- Peak learning time (so they say) is 11am.
- Students here seem to like 9am-4pm.
- American students produce another bump later at night.

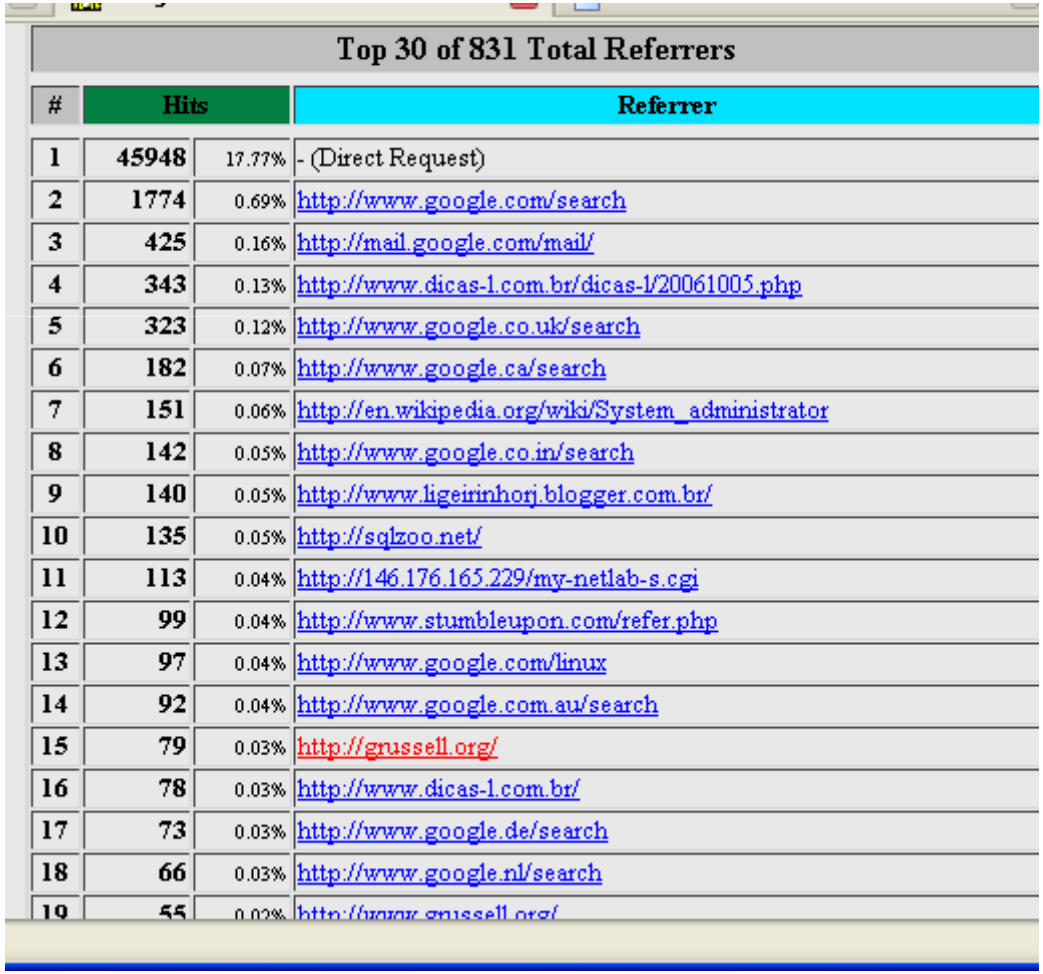
Edinburgh Napier UNIVERSITY

Users

#	Hits		Files		KBytes		Visits		Hostname
1	25319	9.79%	24605	10.69%	94474	2.07%	1051	6.33%	gtw-12.nhs.uk
2	16906	6.54%	16906	7.35%	14933	0.33%	1	0.01%	ds1081-020-069.nyc1.dsl.speakeasy.net
3	6400	2.48%	6381	2.77%	32857	0.72%	122	0.73%	200.182.252.4
4	6005	2.32%	5887	2.56%	24881	0.54%	245	1.47%	200.217.233.139
5	4903	1.90%	2230	0.97%	8144	0.18%	39	0.23%	mvx-200-196-55-148.mundivox.com
6	4506	1.74%	4506	1.96%	6580	0.14%	4	0.02%	200-221-128-3.corp.uolinc.com
7	3679	1.42%	9	0.00%	20	0.00%	2	0.01%	59.145.136.1
8	2728	1.06%	2726	1.18%	10462	0.23%	47	0.28%	59.163.124.54.static.vsnl.net.in
9	2690	1.04%	2590	1.13%	4733	0.10%	103	0.62%	58.68.28.66
10	2647	1.02%	2645	1.15%	11453	0.25%	193	1.16%	glenlivet.spc.eeng.liv.ac.uk
11	2479	0.96%	2455	1.07%	11568	0.25%	69	0.42%	mx2.queirozgalvao.com
12	2381	0.92%	2376	1.03%	50698	1.11%	1	0.01%	200.4.171.30
13	2156	0.83%	270	0.12%	2494	0.05%	6	0.04%	186.112-84-212.staticip.namesco.net
14	1657	0.64%	1657	0.72%	5821	0.13%	8	0.05%	80-41-249-174.dynamic.dsl.as9105.com
15	1606	0.62%	1601	0.70%	7577	0.17%	42	0.25%	146.176.242.35
16	1479	0.57%	347	0.15%	2545	0.06%	6	0.04%	r200-40-197-174.static.adinet.com.uy
17	1478	0.57%	1478	0.64%	0	0.00%	1476	8.89%	host.avidnetwork.com
18	1323	0.51%	1056	0.46%	12408	0.27%	3	0.02%	80-192-78-217.cable.ubr13.edin.blueyonder.co.uk
19	1157	0.45%	1155	0.50%	5840	0.13%	42	0.25%	146.176.242.54

Edinburgh Napier UNIVERSITY

Referrer Info



Top 30 of 831 Total Referrers

#	Hits		Referrer
1	45948	17.77%	- (Direct Request)
2	1774	0.69%	http://www.google.com/search
3	425	0.16%	http://mail.google.com/mail/
4	343	0.13%	http://www.dicas-l.com.br/dicas-l/20061005.php
5	323	0.12%	http://www.google.co.uk/search
6	182	0.07%	http://www.google.ca/search
7	151	0.06%	http://en.wikipedia.org/wiki/System_administrator
8	142	0.05%	http://www.google.co.in/search
9	140	0.05%	http://www.ligeirinhorj.blogger.com.br/
10	135	0.05%	http://sqlzoo.net/
11	113	0.04%	http://146.176.165.229/my-netlab-s.cgi
12	99	0.04%	http://www.stumbleupon.com/refer.php
13	97	0.04%	http://www.google.com/linux
14	92	0.04%	http://www.google.com.au/search
15	79	0.03%	http://grussell.org/
16	78	0.03%	http://www.dicas-l.com.br/
17	73	0.03%	http://www.google.de/search
18	66	0.03%	http://www.google.nl/search
19	55	0.02%	http://www.grussell.org/

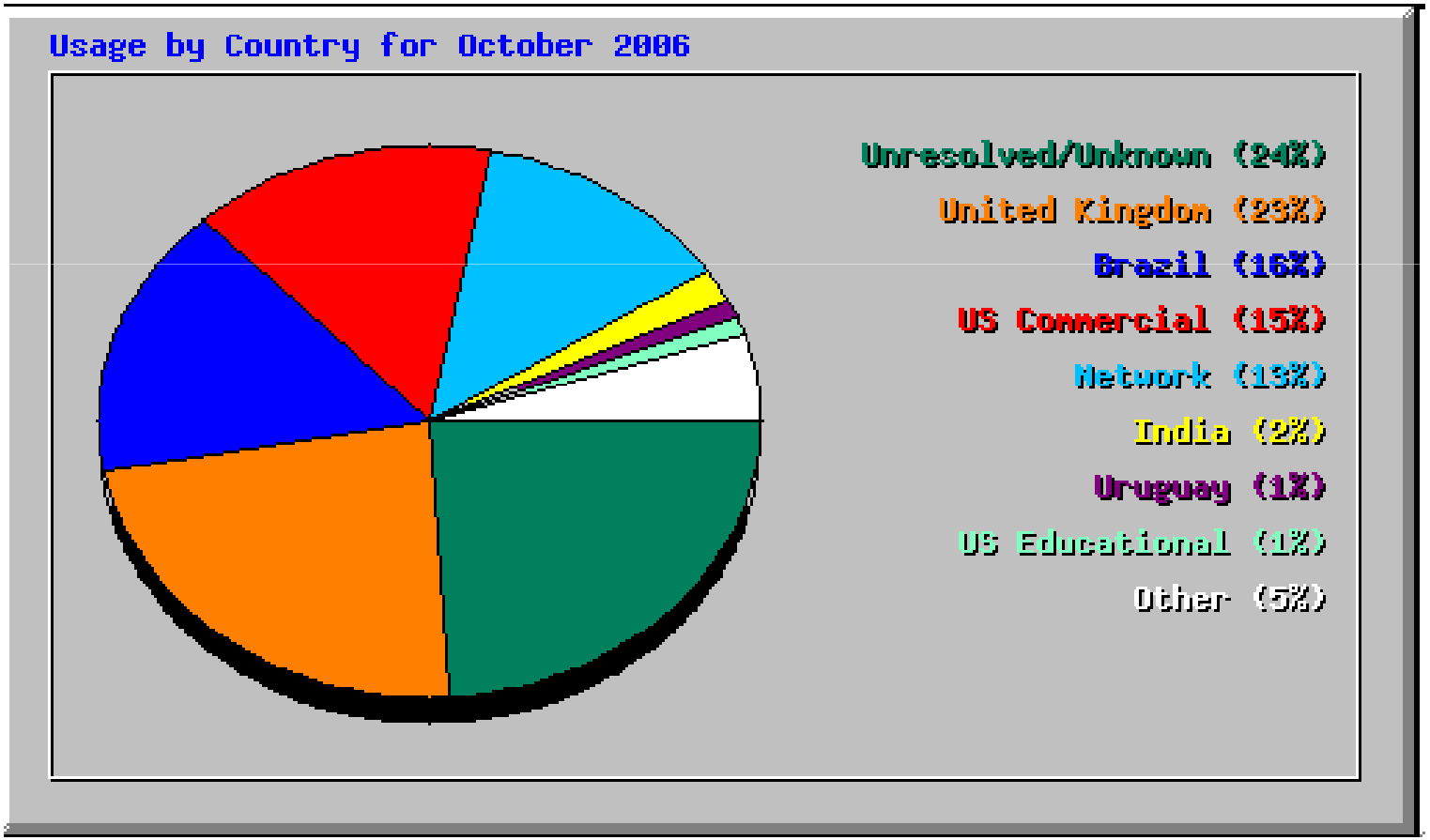
Edinburgh Napier UNIVERSITY

What search terms?

Top 20 of 1160 Total Search Strings			
#	Hits		Search String
1	176	8.17%	umask 022
2	68	3.16%	linuxzoo
3	46	2.13%	unix file structure
4	45	2.09%	apache2 authentication
5	31	1.44%	rndc.key
6	30	1.39%	apache basic authentication
7	27	1.25%	apache public_html
8	20	0.93%	generate rndc key
9	18	0.84%	linux zoo
10	18	0.84%	named service
11	16	0.74%	iptables -f
12	16	0.74%	mysql replicate
13	16	0.74%	symbolic link example
14	15	0.70%	apache2 basic authentication
15	15	0.70%	setting umask
16	14	0.65%	apache basic auth
17	13	0.60%	named forward
18	12	0.56%	rndc key

Edinburgh Napier UNIVERSITY

Where from?

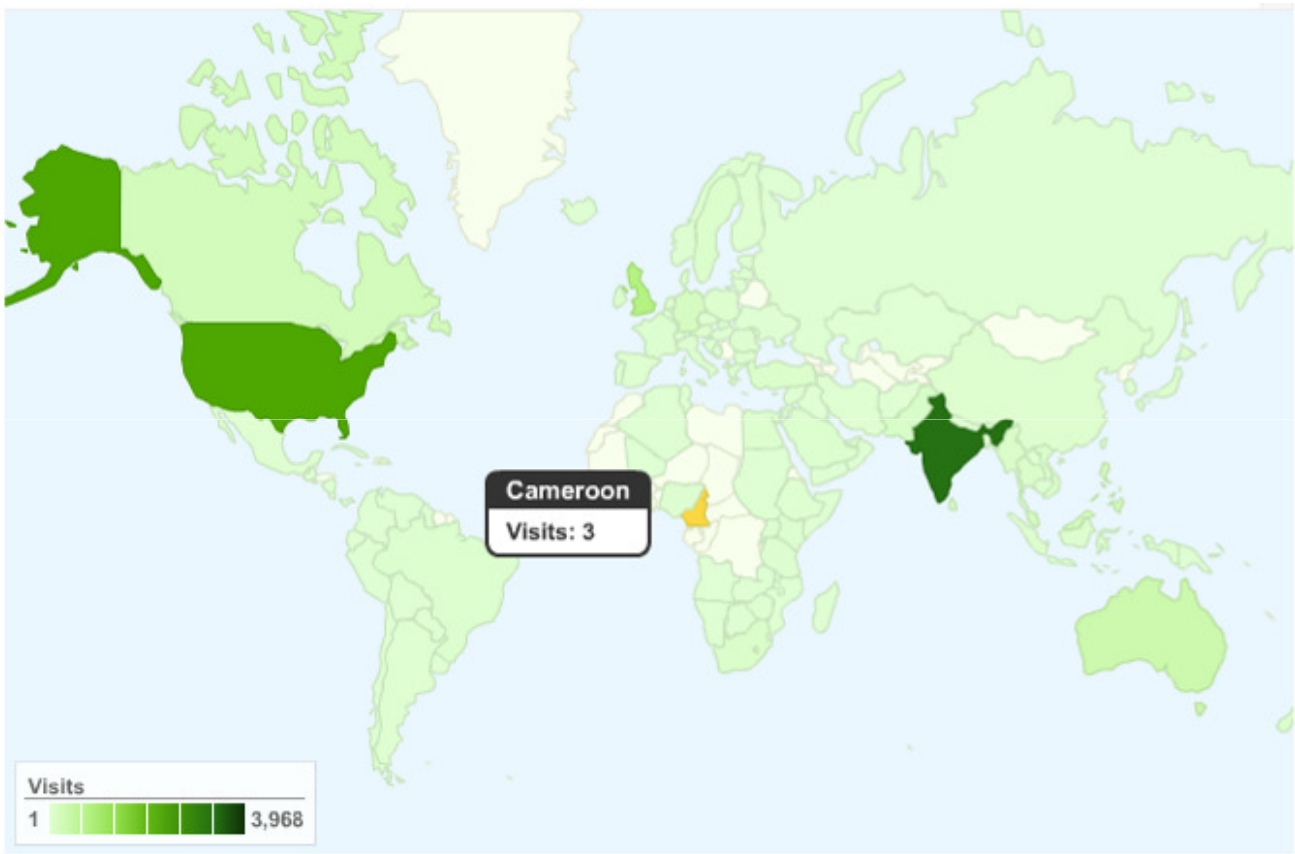


Edinburgh Napier UNIVERSITY

Google Analytics

- Another approach to web logging is to use JavaScript embedded in each web page.
- This does away with the need to access the web log.
  - Good if you don't have access!
- It does mean that
  - You only get logs where there is javascript switched on.
  - Each page is slowed by having extra stuff on it.
  - It's a little more complex.

Edinburgh Napier UNIVERSITY



11,998 visits came from 141 countries/territories

Edinburgh Napier UNIVERSITY

Logging Summary

- What is best?
- I have used both and have mixed feelings…
- Things to consider
  - Convenience
  - Reliability
  - Availability
  - Performance
  - Cost
  - Privacy
  - Complexity

Edinburgh Napier
UNIVERSITY

Discussions

Edinburgh Napier UNIVERSITY

Discussion

- Apache runs as a user, usually “apache” or “httpd”. For apache to serve a file from a user’s public_html directory, what permissions would be required?

Edinburgh Napier UNIVERSITY

Discussion

- Here are some mock exam questions you should now be able to answer:

Edinburgh Napier UNIVERSITY

Question 1

- To test a web server which is hosting the virtual host “grussell.org”, using only telnet, what would you type at the telnet prompt?

Edinburgh Napier UNIVERSITY

Question 2

What fields would you expect to have to define in a VirtualHost definition in apache?

Edinburgh Napier UNIVERSITY

Question 3

- Below is a line from a webserver logfile:

157.55.18.25 - - [31/Aug/2011:12:48:04 +0100] "GET /robots.txt HTTP/1.1" 200 48 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)"

- What kind of request was this? Was this a successful request (i.e. was a document found)?