

THE AUSTRALIAN NATIONAL UNIVERSITY

WEB SERVERS

COMP1710 – New Media

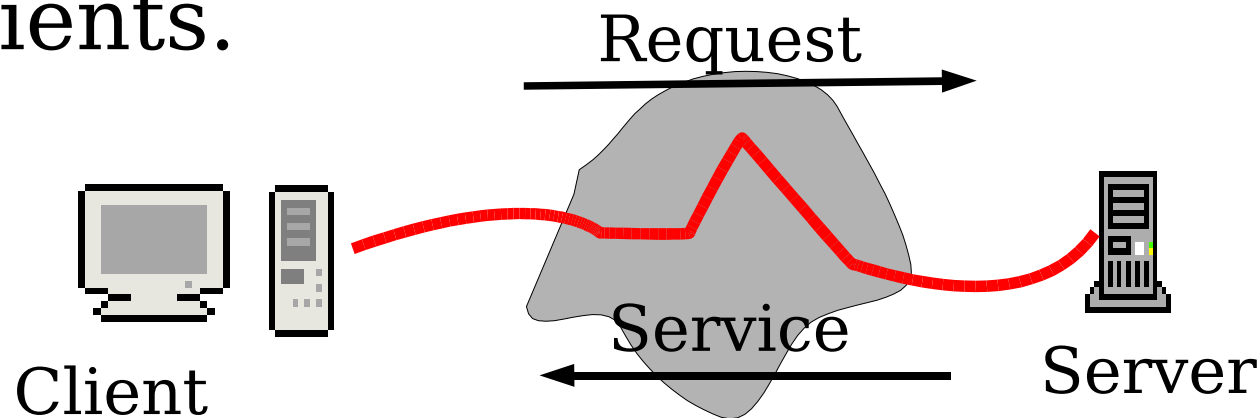
Eric McCreath

2004

- The server model
- TCP/IP revisited(PN, NAT)
- Ports and standard services
- HTTP Server – Apache
- GET and POST
- Files and directory structures
- Authorization
- Cookies (maintaining a session)
- caching

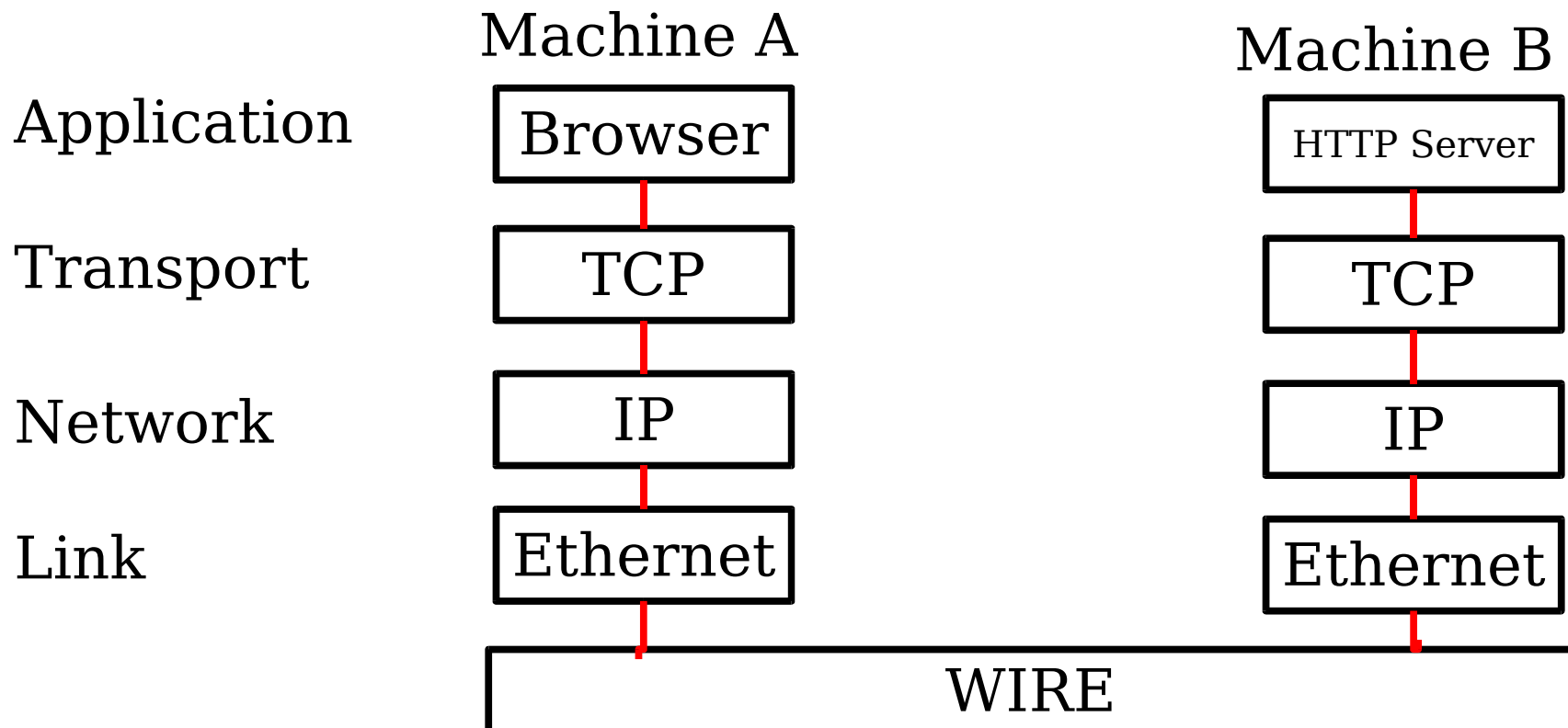
The Client-Server Model

- HTTP uses a client-server model for your web browser to interact with the web server.
- This means the client initiates and maintains the interaction with the web server.
- The web server simply listens for request that come from possibly many different clients.



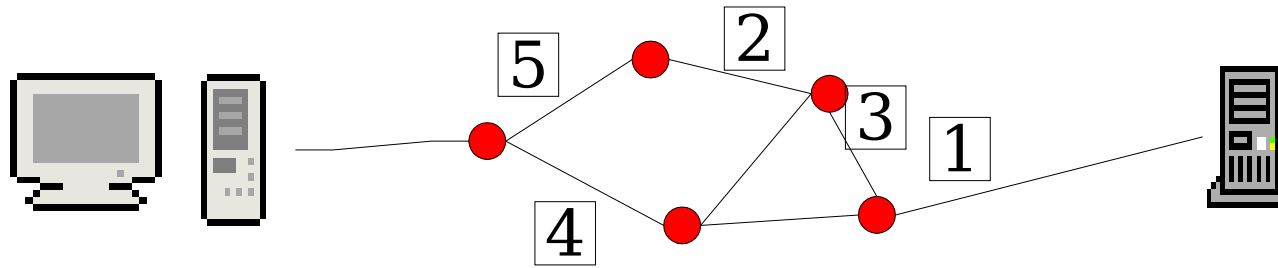
TCP/IP Revisited

- A protocol stack provides a standard and modular way for computers to communicate. The internet uses the TCP/IP suite of protocols.

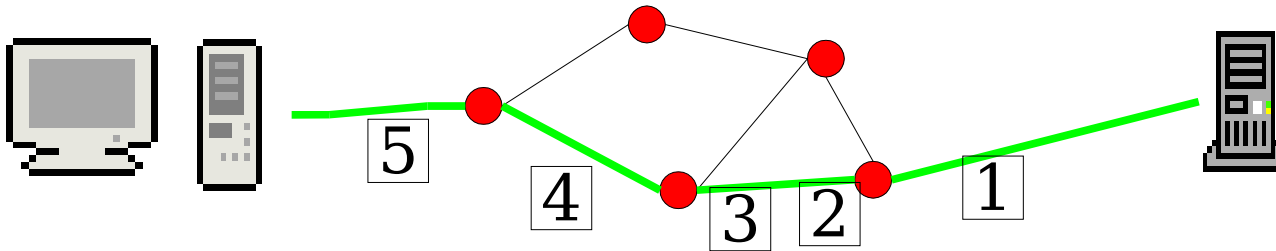


- IP – is a network layer which has the task of unreliably get a packet of data from a source machine to the destination machine.
- Machines are addressed using a 32 bit IP address. These are often displayed as four decimal number separated by a period.
(eg 150.203.160.43)
- If you want something to work on the internet it needs an IP address. (Either fixed or temporary via DHCP(Dynamic Host Configuration Protocol))

- IP uses a datagram service model as aposed to a virtual-circuit service model.



Packet Switch Approach



Circuit Approach

- If you wish to communicate on the internet you need an IP address. If an organization does not need there hosts to act as servers on the internet then they can use a private subnet.

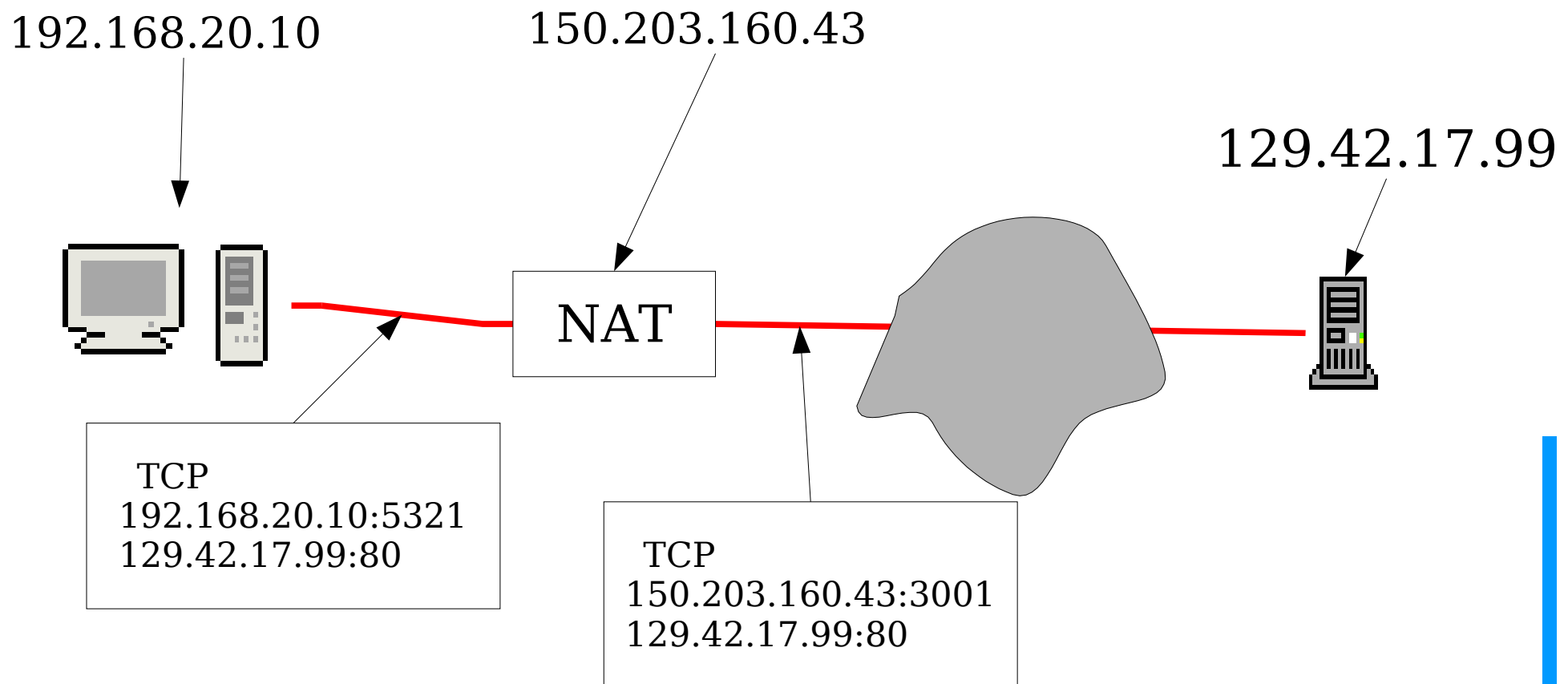
From : <http://www.faqs.org/rfcs/rfc1918.html>

```
:
3. Private Address Space
The Internet Assigned Numbers Authority (IANA) has reserved the
following three blocks of the IP address space for private internets:
  10.0.0.0      - 10.255.255.255 (10/8 prefix)
  172.16.0.0   - 172.31.255.255 (172.16/12 prefix)
  192.168.0.0  - 192.168.255.255 (192.168/16 prefix)
:
An enterprise that decides to use IP addresses out of the address
space defined in this document can do so without any coordination
with IANA or an Internet registry. The address space can thus be used
by many enterprises. Addresses within this private address space will
only be unique within the enterprise, or the set of enterprises which
choose to cooperate over this space so they may communicate with each
other in their own private internet.
```

- TCP provides a reliable connection between 'ports' on two different hosts.
- TCP uses IP to send the data between the hosts.
- A TCP connection is uniquely defined by the following:
 - source IP address(32 bits)
 - destination IP address(32 bits)
 - source port number(16 bits)
 - destination port number(16 bits)

NAT(Network Address Translation)

- NAT aka IP Masquerading
- This enable hosts within a private network to connect to hosts on the internet.



- HTTP(HyperText Transfer Protocol) Servers use TCP connections for communication. Basically they wait to receive a connection from a client on port 80.
- Generally a HTTP Server we deal with a number of requests at the same time. This will involve the server running a number of processes that handle the different requests.
- If a server is used greatly it will have to employ techniques that scale up(or scale out!).

- There are a number of HTTP commands that exist. The main two that are used are GET and POST. In many respects these two commands are very similar, they both request a page. However POST also enables the browser to give information that was set in a form. (See - RFC2616)

Setup a TCP connection. → [~] telnet boole 80
Trying 127.0.0.1...
Connected to localhost.localdomain.
Escape character is '^]'.
HTTP command → GET /
<html>
HTML Document back ↗ <head>
 <title> My Laptop Web Server </title>
 </head>
 :
 :

- The web content is stored within files within a directory structure on the server.
- Apache by default use the location :
`/var/www`
- So a URL such as
<http://boole/images/logo.gif>
- Would be looking to send back the contents of the file:
`/var/www/images/logo.gif`
- The Apache server needs permission to read this file.

- HTTP/1.0 – uses non persistent connections. This means a new TCP connection must be established for each object within a web page.
- Parallel connections can be made to improve performance.
- HTTP/1.1 – may also use persistent connections. This permits many object to be transferred over the one connection.
- Persistent connections may also allow pipelining(new requests are sent before the information is back from previous requests.)

- A HTTP Request will have the following format:

```
<method> <URL> <Version>  
<field name>: <value>  
<field name>: <value>  
:  
<field name>: <value>  
  
<entity body>
```

```
GET /streams HTTP/1.1  
User-Agent: Mozilla/5.0  
Accept: text/xml, application/xml, ...  
Accept-Encoding: gzip, deflate  
Accept-Charset: ISO-8859-1, utf-8;q=0.7..  
Keep-Alive:300  
Connection: keep-alive
```

Example request



- A HTTP Response will have the following format:

<version> <status code> <phrase>
<field name>: <value>

<field name>: <value>

:

:

<field name>: <value>

<entity body>

Common status codes:

200 OK

301 Moved Permanently

302 Found

400 Bad Request

404 Not Found

505 HTTP Version Not Supported

HTTP/1.1 302 Found

Date: Mon, 11 Oct 2004 00:46:23 GMT

Server: Apache/1.3.26 ...

Location: <https://cs.anu.edu.au/streams>

Keep-Alive: timeout=15, max=100

Connection: Keep-Alive

Transfer-Encoding: chunked

Content-Type: text/html; charset=iso-8859-1

<web page that says this web page has moved>

Example Response

- HTTP is stateless. This makes it simpler to develop high performance web servers.
- However, often you need to be able to identify users. This would generally be either to restrict access or enable to the web page's content to be a function of the user accessing these pages.
- The two simple approaches the HTTP uses to identify users are:
 - Authorization and
 - Cookies.

- Web pages can be restricted to a group of users by requiring a user name and password. HTTP includes this simple authorization mechanism.
- A page is requested as usual. The server returns '401 Authorization Required' status code. The browser prompts the user for a name and password which is then included in the 'WWW-Authenticate:' headers.
- This is rather insecure as the name and password is sent in plain text.

- Cookies enable a site to keep track of users.
- When someone first uses a web site they are given a cookie.
eg Set-cookie: student=12345
- When they revisit this site the browser sends their cookie.
e.g. 'Cookie: student=12345'
- Now the web site knows it was the same person.

- Often a browser will request an object that it already has a copy of. In these cases the browser can use a cache to save on bandwidth.

```
GET /logo.gif HTTP/1.0  
User-agent: Mozilla/4.0
```

```
HTTP/1.0 200 OK  
Date: Fri, 1 Jan 2004 10:35:10 GMT  
Server: Apache/1.3.26  
Last-Modified: Mon, 12 Feb 2003 9:30:37 GMT  
Content-Type: image/gif  
<image data>
```

```
GET /logo.gif HTTP/1.0  
User-agent: Mozilla/4.0  
If-modified-since: Mon, 12 Feb 2003 9:30:37 GMT
```

```
HTTP/1.0 304 Not Modified  
:  
:
```

- Kurose and Ross, Computer Networking : A Top-Down Approach Featuring the Internet, Pearson, 2nd Ed, 2003.
- Google search