Web Engineering

Lecture 2
PROTOCOLS
(TCP/IP, HTTP)

Zulfiqar Ahmad
Lecturer
Department of Information Technology
Hazara University Mansehra
zulfiqarahmad@hu.edu.pk

Sockets

- Sockets, or ports, are a very low level software construct that allows computers to talk to one another
- When you send information from one computer to another, you send it to a port on the receiving computer
  - If the computer is "listening" on that port, it receives the information
  - In order for the computer to "make sense" of the information, it must know what protocol is being used
- Common port numbers are 80 (for web pages), 23 (for telnet) and 25 and 110 (for mail)
- Port numbers above 1024 are available for other kinds of communication between our programs

Protocols

- In order, for computers, to communicate with one another, they must agree on a set of rules for **who says what**, **when they say it**, and **what format they say it in**
- The set of rules used by computers for communication is called a protocol.
- Some common protocols are HTTP (for web pages), FTP (for file transfer), and SMTP (Simple Mail Transfer Protocol)

What is a protocol?

- A protocol is the set of rules governing a conversation between people
- We have seen that the client and server carry on a machine-to-machine conversation
- A network protocol is the set of rules governing a conversation between a client and a server
- There are many protocols, e.g. TCP/IP, HTTP etc.

TCP/IP

- The Internet (and most other computer networks) are connected through TCP/IP networks
- TCP/IP is actually a combination of two protocols:
  - IP, Internet Protocol, is used to move packets (chunks) of data from one place to another
    - Places are specified by IP addresses: four single-byte (0..255) numbers separated by periods
    - Example: 192.168.1.1
  - TCP, Transmission Control Protocol, ensures that all necessary packets are present, and puts them together in the correct order
- TCP/IP forms a "wrapper" around data of *any* kind
- The data uses its own protocol, for example, FTP

Hostnames and DNS servers

- The "real" name of a computer on the internet is its four-byte IP address
- People, however, don't like to remember numbers, so we use hostnames instead
- For example, the hostname www.cis.upenn.edu is 158.130.12.9
- A DNS (Domain Name Server) is a computer that translates hostnames into IP addresses
- A domain name server (DNS) is a machine that keeps a table of names and corresponding IP addresses
  - Think of it as like a phone book--names to useful numbers
  - Of course, you have to know the IP address of the DNS in order to use it!

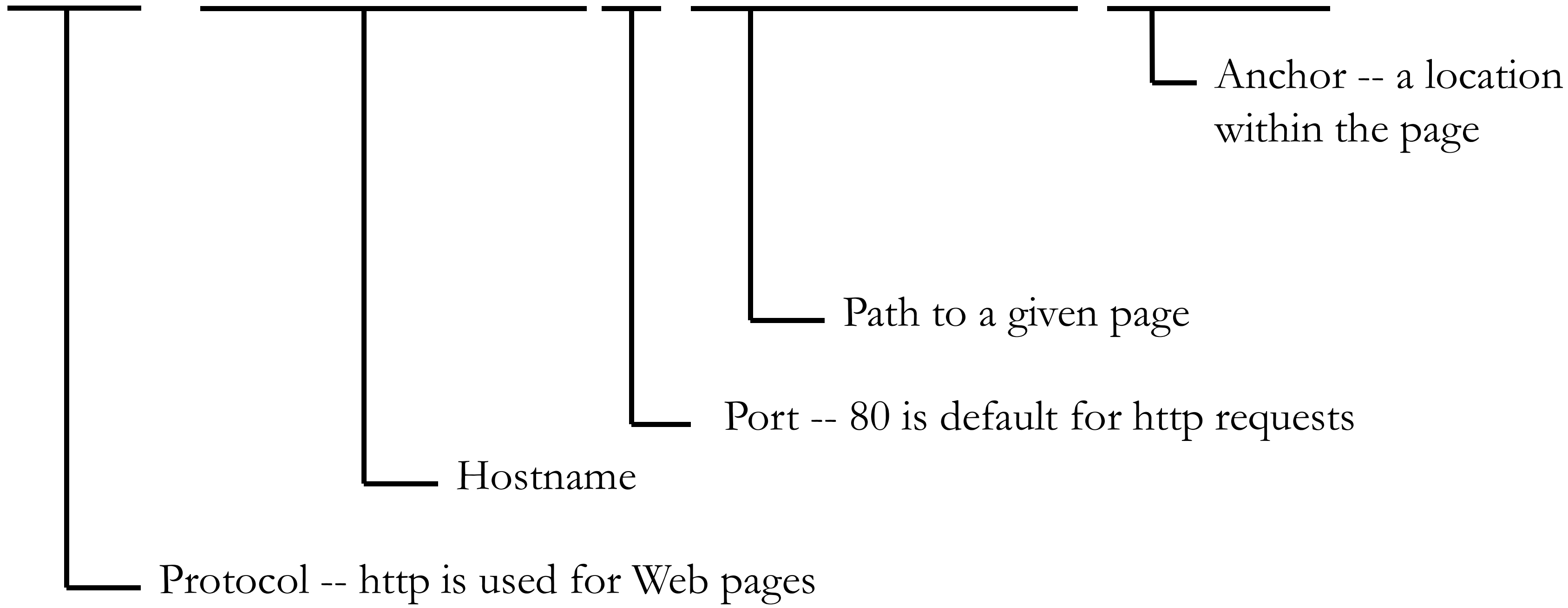
DHCP

- If you have a web site, it must be hosted on a computer that is "permanently" on the Web
  - This computer must have a permanent IP address
  - There aren't enough IP addresses for the number of computers there are these days
- If you have no permanent web site, you can be given a *temporary* (dynamically allocated) IP address each time you connect to the Web
- Similarly, if you have a home or office network, only one computer needs a permanent IP address
  - The rest of the computers can be assigned *internal,* permanent IP addresses (not known to the rest of the world)
  - They can also be assigned internal IP addresses dynamically
- DHCP (Dynamic Host Configuration Protocol) is a way of assigning temporary IP addresses as needed

URLs

- A URL, Uniform Resource Locater, defines a location on the Web
- A URL has up to five parts:

http://www.xyz.com:80/ad/index.html#specials



Anchor -- a location within the page

Path to a given page

Port -- 80 is default for http requests

Hostname

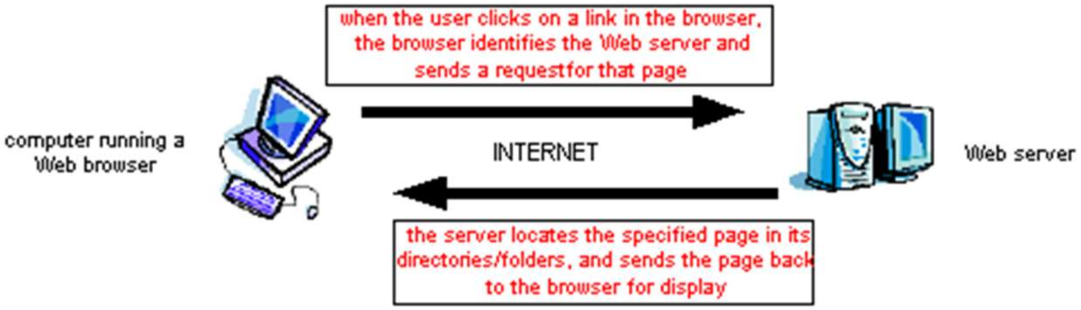
Protocol -- http is used for Web pages

World Wide Web

- the Web is the world's largest client/server system

communication occurs via message passing

- within browser, select URL of desired page
- browser requests page from server
- server responds with message containing
  - type of page (HTML, gif, pdf, zip, …)
  - page contents
- browser uses type info to correctly display page
- if page contains other items (images, applets, …), browser must request each separately



when the user clicks on a link in the browser, the browser identifies the Web server and sends a requestfor that page

the server locates the specified page in its directories/folders, and sends the page back to the browser for display

HTTP vs HTML

- HTML: hypertext markup language
  - Definitions of tags that are added to Web documents to control their appearance
- HTTP: hypertext transfer protocol
  - The rules governing the conversation between a Web client and a Web server
  -
- Both were invented at the same time by the same person

- The HTTP protocol used for Web applications was invented by Tim Berners Lee

HTTP

- Hypertext Transfer Protocol (HTTP):

  application-level protocol for distributed, collaborative, hypermedia information systems

  - generic, stateless, object-oriented
  - can be used for many tasks, such as name servers & distributed object management systems
  - underlying language of the Web

HTTP/1.0 allowed only connectionless message passing

- each request/response required a new connection
- to download a page with images required multiple connections

  can overload the server, require lots of overhead

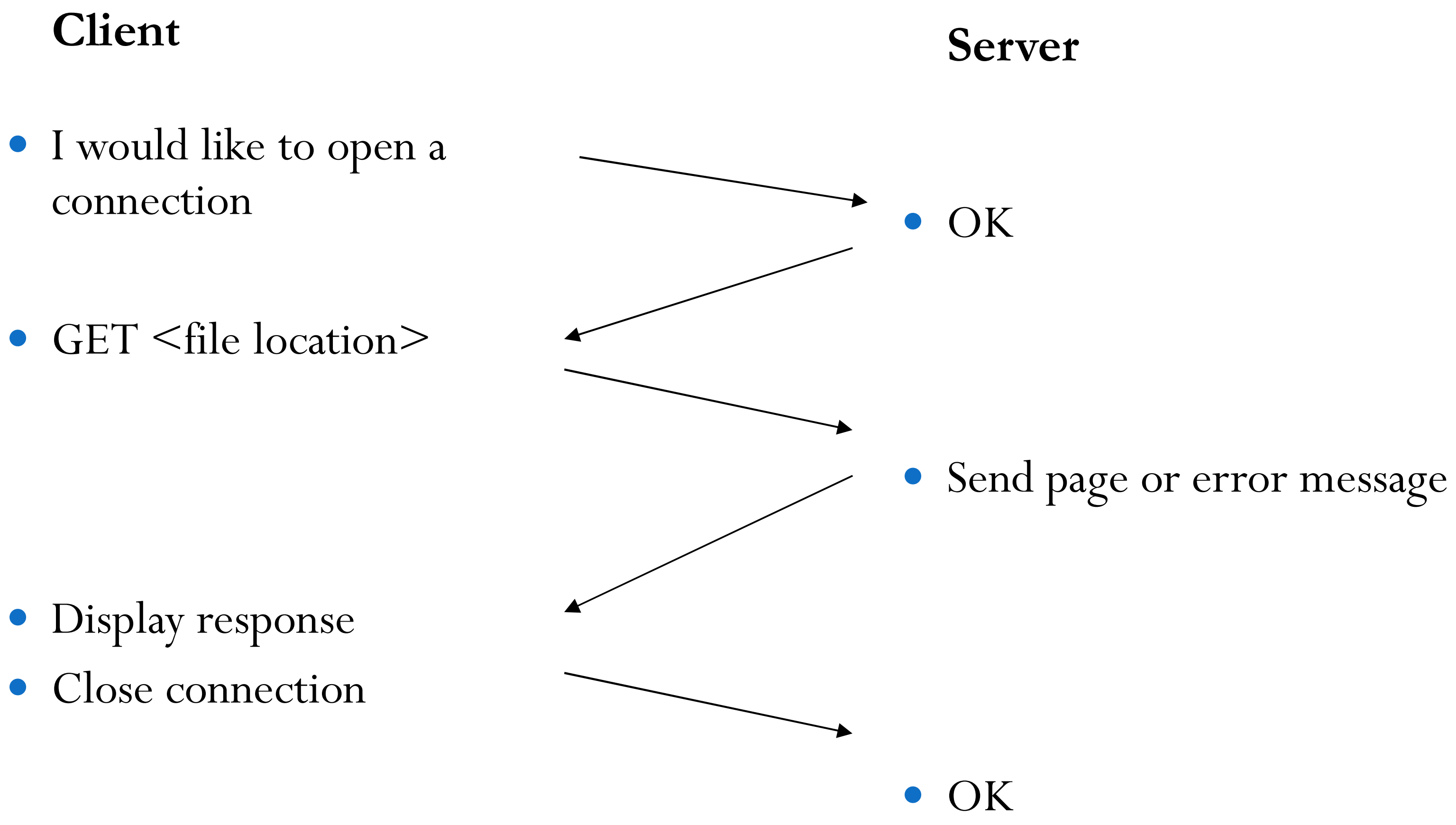
HTTP/1.1 provides persistent connection by default

- once client & server connect, remains open until told to close it *(or timeout)*

  reduces number of connections, saves overhead
- client can send multiple requests without waiting for responses

  e.g., can request all images in a page at once

An HTTP conversation



Client

- I would like to open a connection
- GET <file location>
- Display response
- Close connection

Server

- OK
- Send page or error message
- OK

HTTP is the set of rules governing the format and content of the conversation between a Web client and server

An HTTP example

The message requesting a Web page must begin with the work "GET" and be followed by a space and the location of a file on the server, like this:

```
GET /fac/lpress/shortbio.htm
```

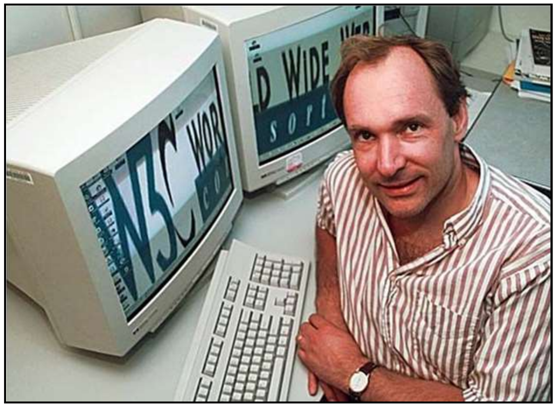
The protocol spells out the exact message format, so any Web client can retrieve pages from any Web server.

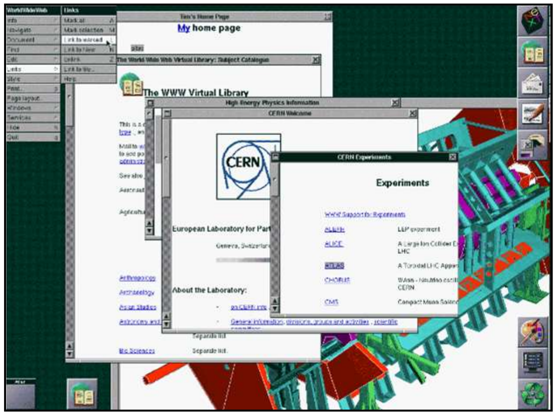
Network protocols

- The details are only important to developers.
- The rules are defined by the inventor of the protocol – may be a group or a single person.
- The rules must be precise and complete so programmers can write programs that work with other programs.
- The rules are often published as an RFC along with running client and server programs.
- The HTTP protocol used for Web applications was invented by Tim Berners Lee.

RFC = request for comments

Tim Berners-Lee





My home page

The WWW Virtual Library

European Laboratory for Part

Geneva, Switzerlan

About the Laboratory:

Anthropology

Archaeology

Asian Studies

Bio Sciences

Separate list.

Experiments

WWW Support for Experiments

ALEPH LEP experiment

ALICE A Large Ion Collider E LHC

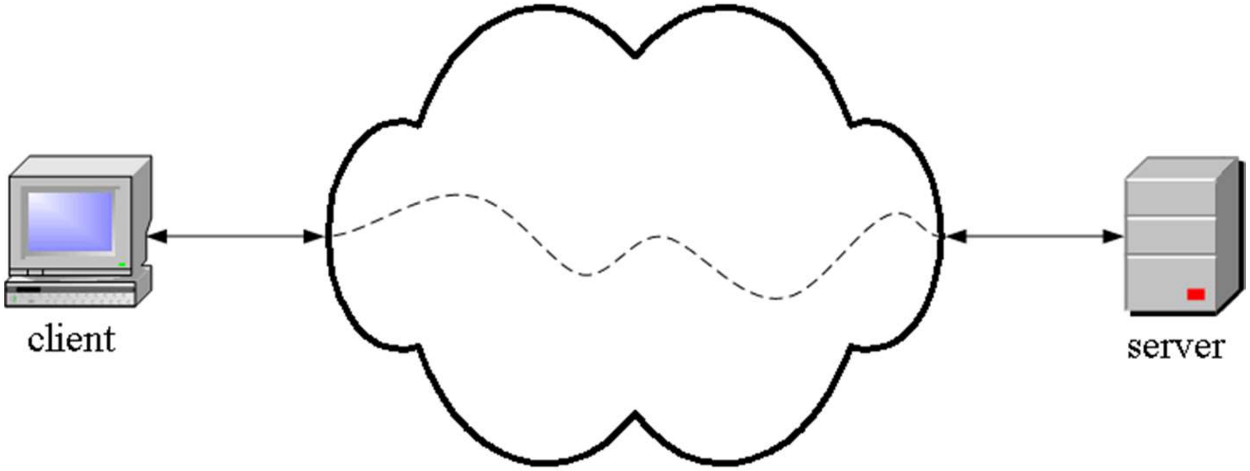
ATLAS A Toroidal LHC Appar

CHORUS WAnn - Neutrino oscil CERN

CMS Compact Muon Soleno

Tim Berners-Lee was knighted by Queen Elizabeth for his invention of the World Wide Web. He is shown here, along with the first picture posted on the Web and a screen shot from an early version of his Web browser.

HTTP is an application layer protocol



- The Web client and the Web server are application programs
- Application layer programs do useful work like retrieving Web pages, sending and receiving email or transferring files
- Lower layers take care of the communication details
- The client and server send messages and data without knowing anything about the communication network

Many application layer protocols are used on the Internet, HTTP is only one

Protocol	**Application**
HTTP: Hypertext Transfer	Retrieve and view Web pages
FTP: File Transfer	Copy files from client to server or from server to client
SMTP: Simple Mail Transfer Protocol	Send email
POP: Post Office Protocol	Read email

The TCP/IP protocol layers

The application program is king – it gets work done using the lower level layers for communication between the client and server.

Application	Get useful work done – retrieve Web pages, copy files, send and receive email, etc.
Transport	Make client-server connections and optionally control transmission speed, check for errors, etc.
Internet	Route packets between networks
Data link	Route data packets within the local area network
Physical	Specify what medium connects two nodes, how binary ones and zeros are differentiated, etc,

Caching

- browsers temporarily store pages or content of pages for future use
  - maintain temporary storage (cache) for recent pages
  - when a page is requested, check to see if already in cache
  - if not in the cache, issue GET request
    - when response message arrives,
      - display page and store in cache (along with header info)
  - if already stored in the cache, send GET request with If-Modified-Since header set to the data of the cached page
    - when response message arrives,
      - if status code 200, then display and store in cache
      - if status code 304, then display cached version instead

Cookies

HTTP message passing may be transaction-based,

- many e-commerce apps require persistent memory of customer interactions

e.g., amazon.com

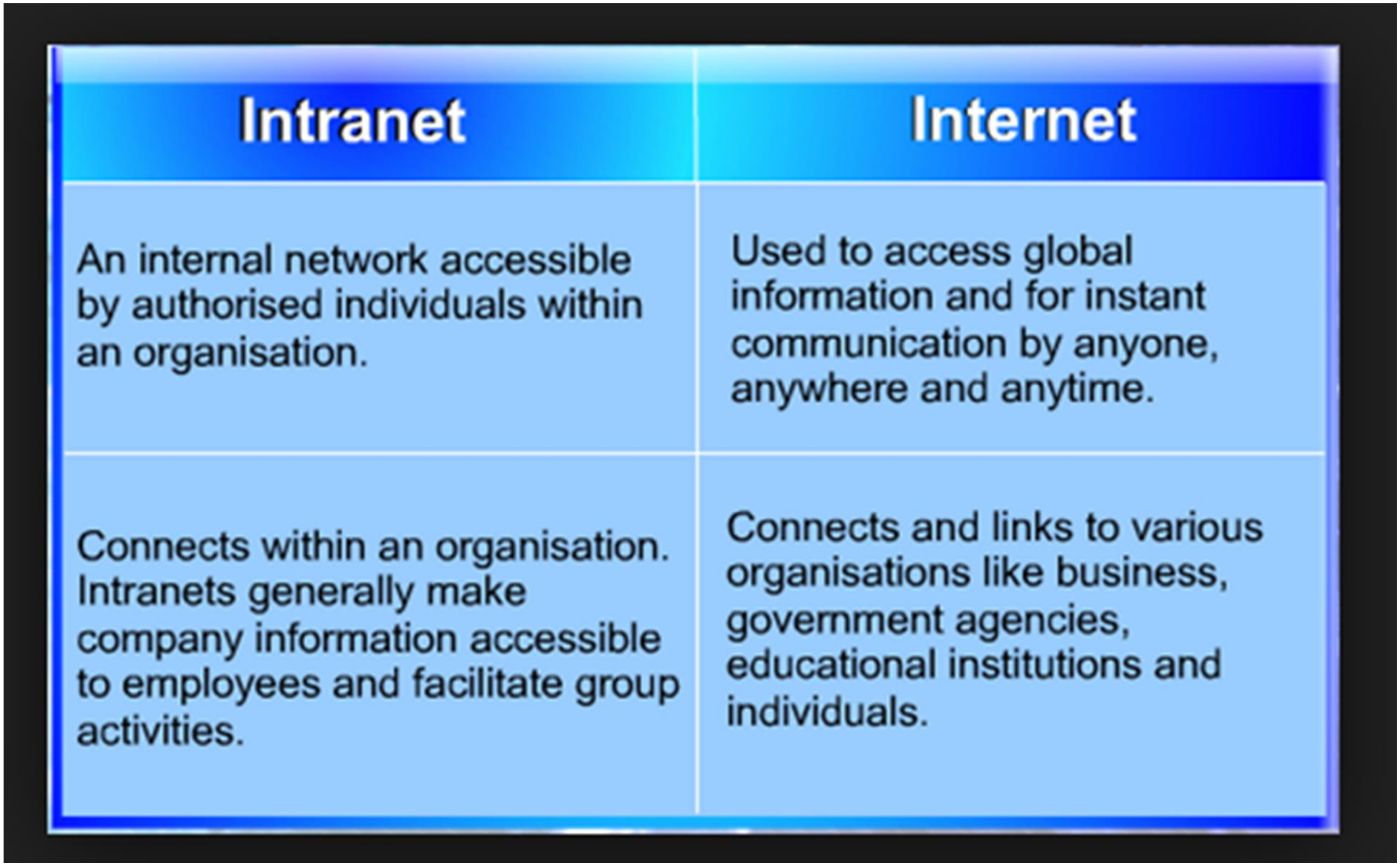
remembers your name, credit card, past purchases, interests

- Netscape's solution: cookies
  - **a *cookie* is a collection of information about the user**
  - **server can download a cookie to the client's machine using the "Set-cookie" header in a response**

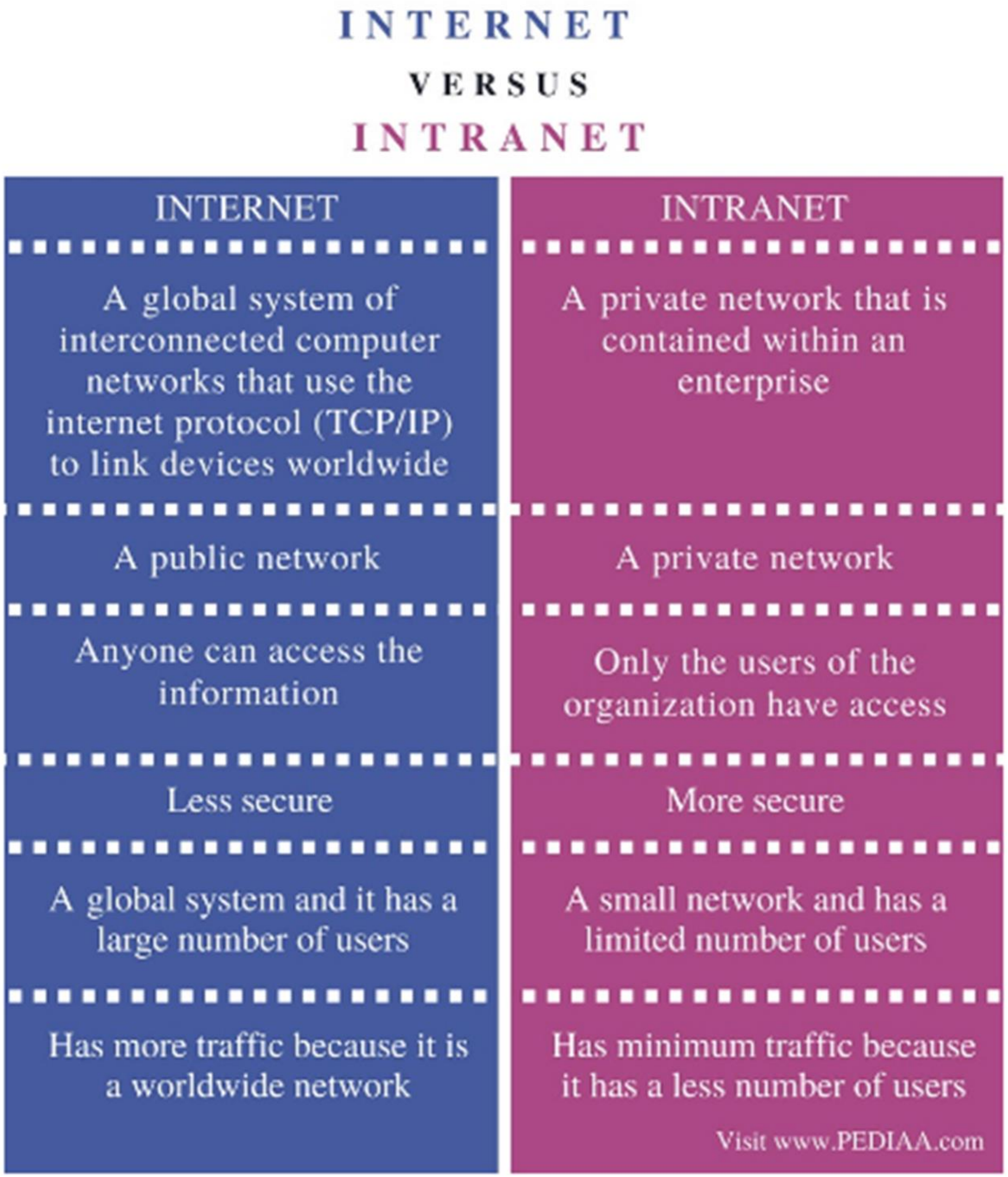
```
Set-cookie: CUSTOMER=Dave_Reed; PATH=/; EXPIRES=Thursday, 29-Jan-04 12:00:00
```

  - when user returns to URL on the specified path, the browser returns the cookie data as part of its request

```
Cookie: CUSTOMER=Dave_Reed
```



Intranet	Internet
An internal network accessible by authorised individuals within an organisation.	Used to access global information and for instant communication by anyone, anywhere and anytime.
Connects within an organisation. Intranets generally make company information accessible to employees and facilitate group activities.	Connects and links to various organisations like business, government agencies, educational institutions and individuals.



INTERNET VERSUS INTRANET

INTERNET	INTRANET
A global system of interconnected computer networks that use the internet protocol (TCP/IP) to link devices worldwide	A private network that is contained within an enterprise
A public network	A private network
Anyone can access the information	Only the users of the organization have access
Less secure	More secure
A global system and it has a large number of users	A small network and has a limited number of users
Has more traffic because it is a worldwide network	Has minimum traffic because it has a less number of users

Visit www.PEDIAA.com