SENG1050

Web Technologies

Week 01

Lecture Content

Course Outline

- ❑ Course Description and Content Overview
- ❑ Why are you Here?
- ❑ Lectures and Lab Times
- ❑ Course Resources – Blackboard
- ❑ Textbook
- ❑ Course Timetable
- ❑ Assessment
- ❑ Academic Integrity
- ❑ Other Administration
- ❑ Where to find answers

Introduction to course – teaching group

- Coordinator / Lecturer
  - Joe Ryan
    - Email: joe.ryan@newcastle.edu.au
    - Office: ES (Engineering Science Building) 234
    - Phone: x16071 (02 4921 6071)
- Lab demonstrators – depending on class

Course Description

- **Goal**: To understand the potential and limitations of the Internet
- High-level coverage of many aspects of the Internet
- Detailed coverage of some practical skills, including:
  - various communication techniques,
  - building Web pages,
  - securing information via encryption

Content Overview

- Introduction to computer networking
- Internet protocols
- Web languages: HTML, CSS, XML, XSLT, Javascript
- Client-server computing
- Cryptography
- Compression
- HCI and Social Aspects of the Internet

Lectures Schedule

- Lectures: Friday 9:00am – 11:00am, RW149
- Labs
  - You should be enrolled in ONE two hour lab class
- Consultation
  - Contact times: Tuesday 9:00am ~ 11:00am

Laboratories

- Lab work starts in Week 2
- You must register for **one and only one** lab
  http://webapps.newcastle.edu.au/rego/index.cfm
- Labs are...
  - Strongly recommended, but are not compulsory
  - Provide the practical "hands-on" learning
  - Will help you understand the assignment work
  - Will help you prepare the exams

Course Resources

- Course resources are available online using the “Blackboard” system
  - http://blackboard.newcastle.edu.au/
  - Lecture slides, lab notes, assignment specifications
  - References
  - Discussion boards: monitored by lecturer and demonstrators (and other students)
- Course materials will only be available on-line: no printed handouts will be supplied!
- Slides *usually* available the night before the lecture

Text

- **Textbook**
  - *No Textbook*
- **Reference text**
  - **Web Development and Design Foundations with HTML5 (6e)**
    - *By Terry Felke-Morris*
    - *ISBN 9780132783392*
  - **Programming the World Wide Web (7e)**
    - *By Robert W. Sebesta*
    - *ISBN 9780132665810*
- **Alternative resources**

  Some additional resources will be linked with the lectures or posted in Blackboard.

Assessment

Midterm Exam	in Week 6 (during the lecture, 1st Sept)	20%
Assignment 1	due week 7	15%
Assignment 2	due week 11	25%
Final Examination		40%

You must score $\geq$ 50% with a final exam mark $\geq$ 40% to pass.

If your total mark $\geq$ 50%, but your final exam mark $\in$ [25,40), you may be given a second chance to sit the final exam.

Assessment

- All assignments are due 11:59pm Friday
  - 11:59pm Saturday = -10% (of total marks for an assessment)
  - 11:59pm Sunday = -20%
  - 11:59pm Monday = -30%
  - 11:59pm Tuesday = -40%
  - 11:59pm Wednesday = -50%
  - Any later = zero

Academic Integrity

- Copying someone else's work (with or without their consent) and presenting it as if it were your own is PLAGIARISM
  - This is **VERY, VERY BAD**
  - and not just because it can get you into trouble
  - http://www.newcastle.edu.au/about-uon/governance-and-leadership/policy-library/document?RecordNumber=D09/1899P

Other Administrative Matters

- Adverse Circumstances / Special Consideration

See http://www.newcastle.edu.au/current-students/learning/assessments-and-exams/adverse-circumstances

Only with **good** reason (significant and unexpected!)

- Having too many assignments due that week definitely does **not** count
- If you feel you may have cause to apply, then complete and submit the forms

Lecture Plan

Weekly program (lectures)

- Week 1 – The Internet, Protocols, TCP/IP, Email, HTTP
- Week 2 – HTML basics
- Week 3 – XML and DTD
- Week 4 – CSS
- Week 5 – More HTML with CSS
- Week 6 – Revision and Midterm
- Week 7 – XSLT
- Week 8 – JavaScript
- Week 9 – User Interface
- Week 10 – Encoding, Compressionl
- Week 11 – Security and Encryption
- Week 12 – Ethics and Course review

Lecture Plan

Weekly program (lectures)

- **Week 1 – The Internet, Protocols, TCP/IP, Email, HTTP**
- Week 2 – HTML basics
- Week 3 – XML and DTD
- Week 4 – CSS
- Week 5 – More HTML with CSS
- Week 6 – Revision and Midterm
- Week 7 – XSLT
- Week 8 – JavaScript
- Week 9 – User Interface
- Week 10 – Encoding, Compression and Information Retrieval
- Week 11 – Security and Encryption
- Week 12 – Ethics and Course review

Week 01 Lecture 01 Outline

The Internet, Protocols, TCP/IP, Email, HTTP

- Internet Communication
- Local Area Network (LAN)/ Wide Area Network (WAN)
- Internet / Intranet / Extranet / VPN
- World Wide Web (WWW)
- Switching – Circuit switching, Packet Switching
- TCP/IP model
- Transmission Control Protocol
- Internet Protocol (TCP/IP)
- IP Address: IPv4 and IPv6
- Doman Name System (DNS)
- Email
- Hyper Text Transfer Protocol (HTTP)

Internet Communications

- The Internet is …
  - A global network connecting millions of computers.
  - **A network of networks.**
  - Constantly growing.
  - Constantly changing.
  - The largest source of information in the world.
  - The biggest marketing machine since the telephone.

Networks

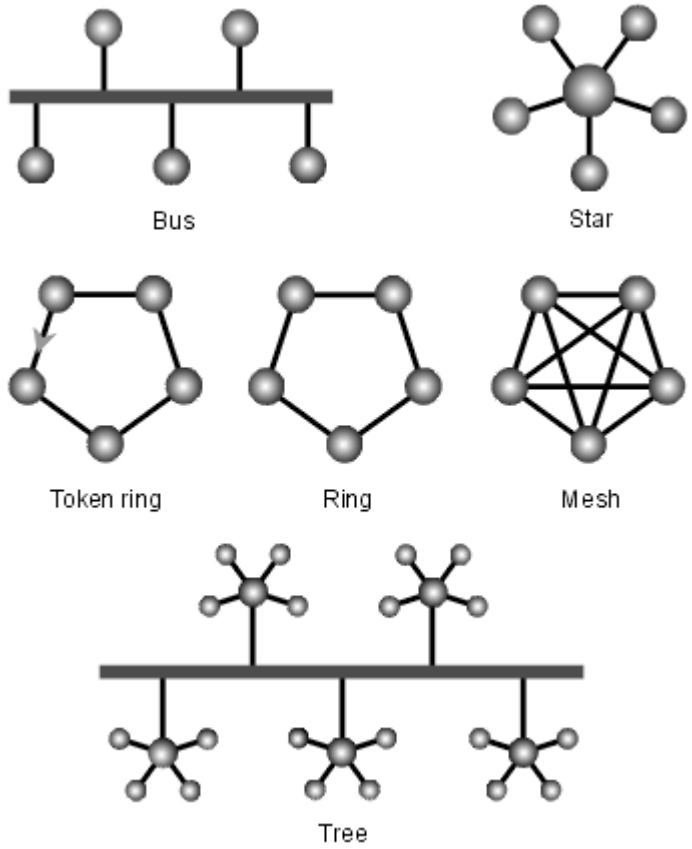
- A Network is …
  - A group of computers connected together to permit the (rapid) transfer of information between them.

- LAN = Local Area Network:
  + cheap and easy to install.
  + geographically close (easy to maintain and secure).
  + controlled at the Department/Group level.

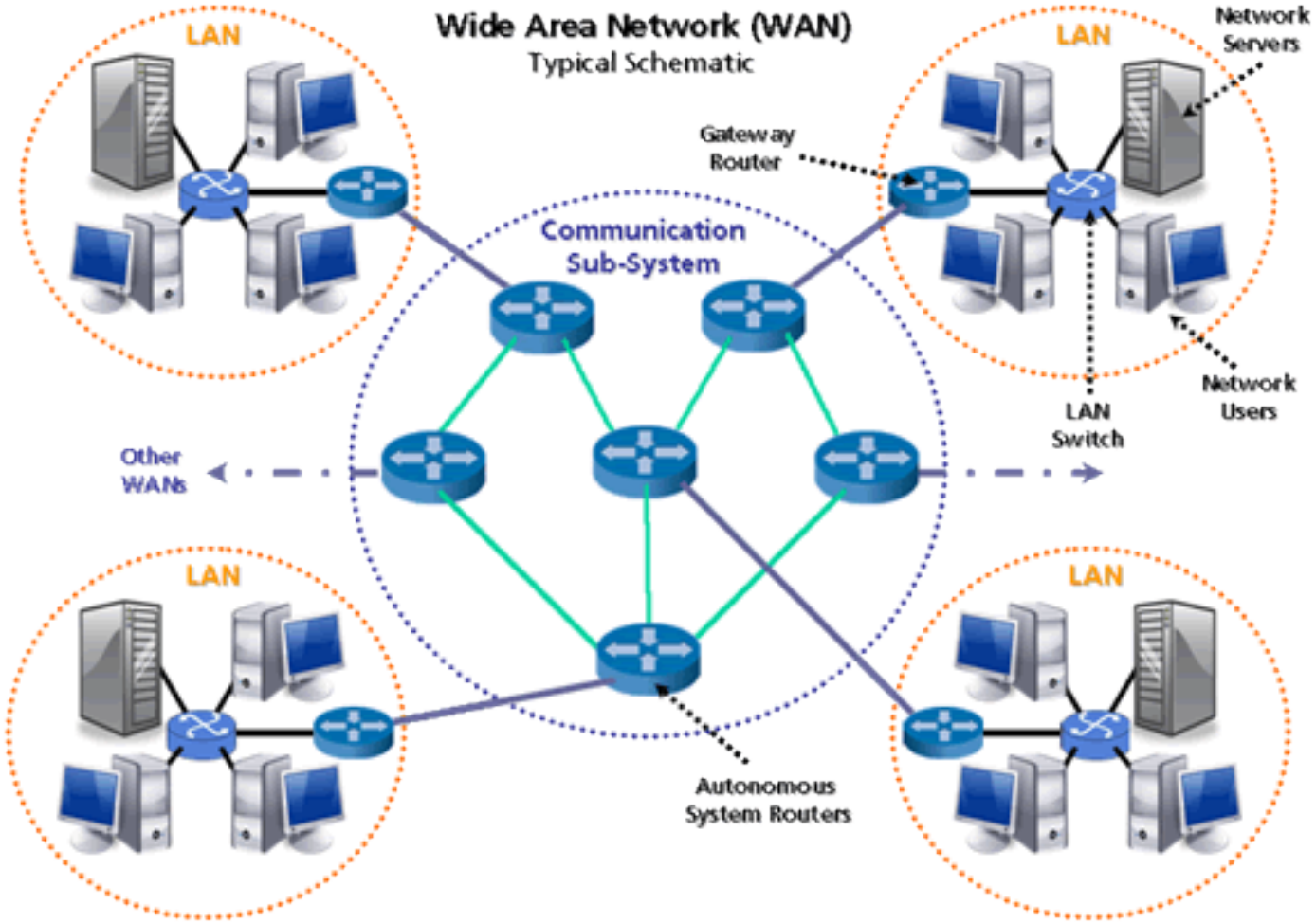
- WAN = Wide Area Network:
  + Effective for large transnational companies.
  + WAN transmissions don't burden the local LANs: they route through special purpose computers.
  - Costs of planning, infrastructure, hardware.
  - Different LAN/WAN technologies are often incompatible.

Networks (LAN) Topologies

- Bus – all computers share a single line; need a way to avoid/detect “collisions”
- Star – all computers communication via a central “switch/hub/router”
- Ring – each computer communicates only with its neighbours, forming a cycle or “ring”; information is relayed in one direction
- Mesh – every computer has a direct line to every other computer
- Tree – a bus with branches



Networks: Wide Area Networks (WAN)



Source: http://www.air-stream.org.au/network/what-wide-area-network

Where does the Internet fit in?

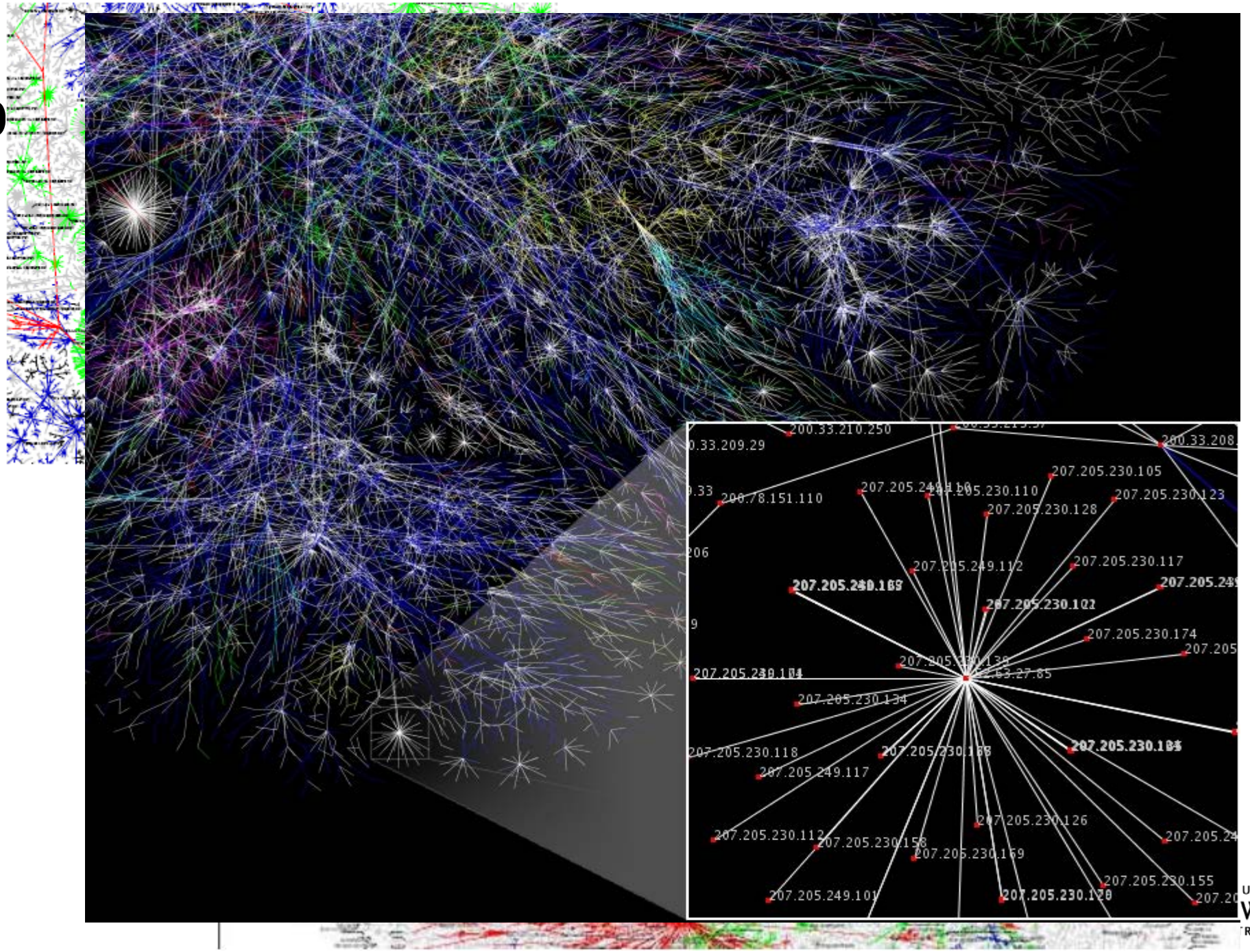
- The Internet is a network of otherwise incompatible networks.
- It joins together many LANs and WANs into a single network.

  "*The Internet allows computers to communicate and be **connected together** based on a set of **agreed upon standards** that a computer must use to be understood.*"
- Later, we will cover some technical details of just how the Internet does this.

Internet Backbone

- The lines that carry the majority of the information are known as the Internet backbone.
- “Information superhighway”:
  - An (old) analogy that suits the concepts of ‘bandwidth’ (the width of the highway),
  - ‘throughput’ (the number of cars through a point per hour),
  - but is not accurate when it comes to rules, regulations and who “owns” the highway.

Map


200.33.210.250
207.205.230.105
207.205.230.110
207.205.230.123
207.205.230.128
207.205.249.112
207.205.230.117
207.205.230.174
207.205.230.134
207.205.230.118
207.205.249.117
207.205.230.126
207.205.230.112
207.205.230.158
207.205.230.169
207.205.230.155
207.205.249.101

Intranets / Extranets / VPN

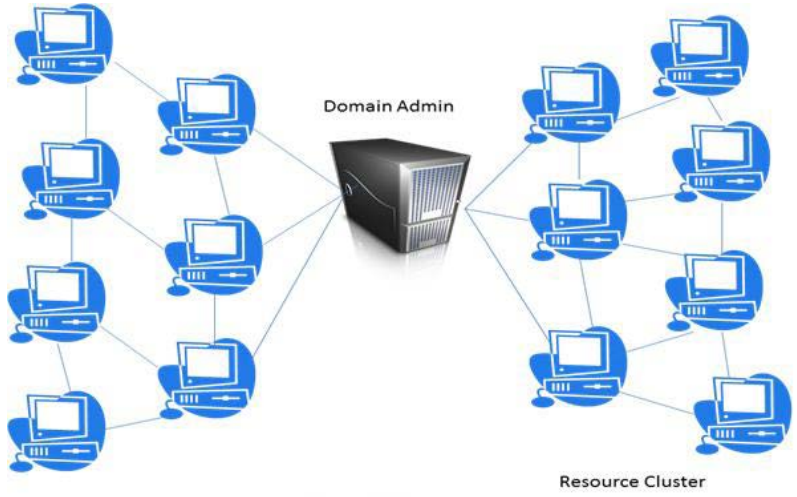
- Intranet
  - Like the Internet, but only accessible from within a certain area (such as a home/office network)
  - A private network, usually owned by a company, and used to provide information and applications for its employees
- Extranet
  - The parts of a company's network that can be accessed from outside
  - Often restricted to suppliers, vendors and/or customers of the company
  - Can include webpages and websites
- VPN (Virtual Private Network)
  - A way for a computer to securely access a company's intranet from outside via the Internet

The World Wide Web

- **WWW ≠ Internet**
- Based on hypertext – instant cross-referencing:
  - Links from a word, phrase, or image to another document.
- The Web contains billions of “pages” of information – but not all connected!
- Browsing through the Web is called “surfing”.
  - Chrome, Internet Explorer, Opera, Firefox, etc.

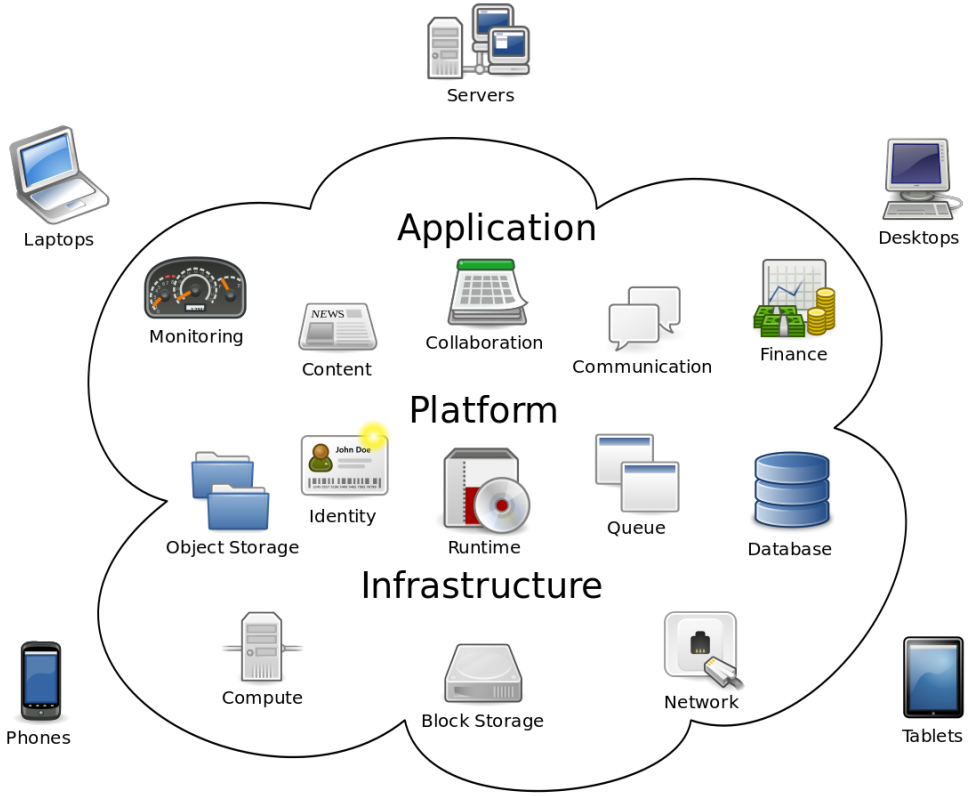
The Grid

- “Grid is a type of **parallel** and **distributed** system that enables the sharing, selection, and aggregation of geographically distributed “autonomous” resources dynamically at runtime depending on their availability, capability, performance, cost, and users’ quality-of-service requirements”: http://www.gridcomputing.com/gridfaq.html
  - Like the “electricity grid”, but the “electricity” is actually “computing power”
  - Show great promise, but still under development



The Cloud

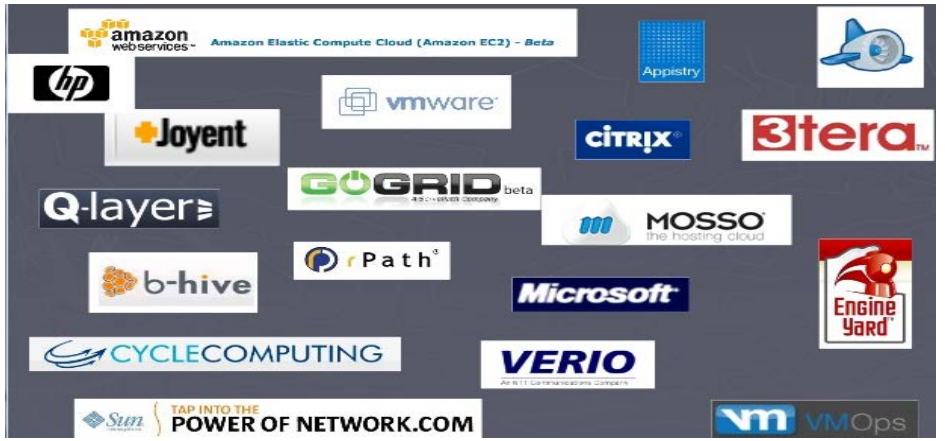
- **“Cloud computing** is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet)” -- Wikipedia



Source: http://www.wikipedia.org

The Cloud

- Advantage
  - Cloud Computing reduces E-Waste - Green IT
  - Cost-efficient
  - Almost limitless storage
  - Information accessibility
- Disadvantages
  - You should connect to the Internet
- Commercial Clouds



Source: http://www.wikipedia.org

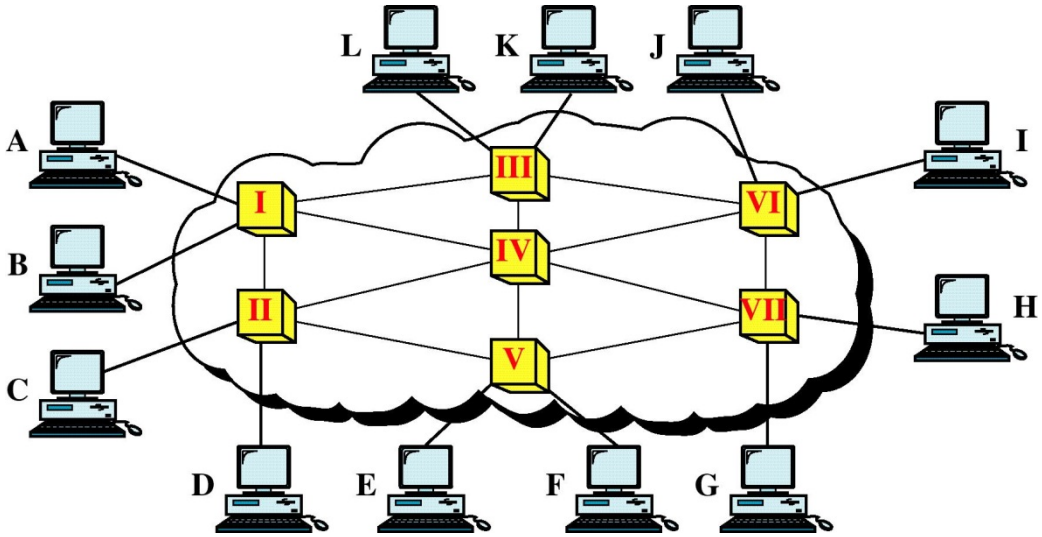
But how does the Internet Work?

- The Internet connects many otherwise incompatible networks into a single virtual network.

- It does this by defining a set of **protocols** that networks must use to become part of the Internet
  - Protocol = a set of rules for communicating

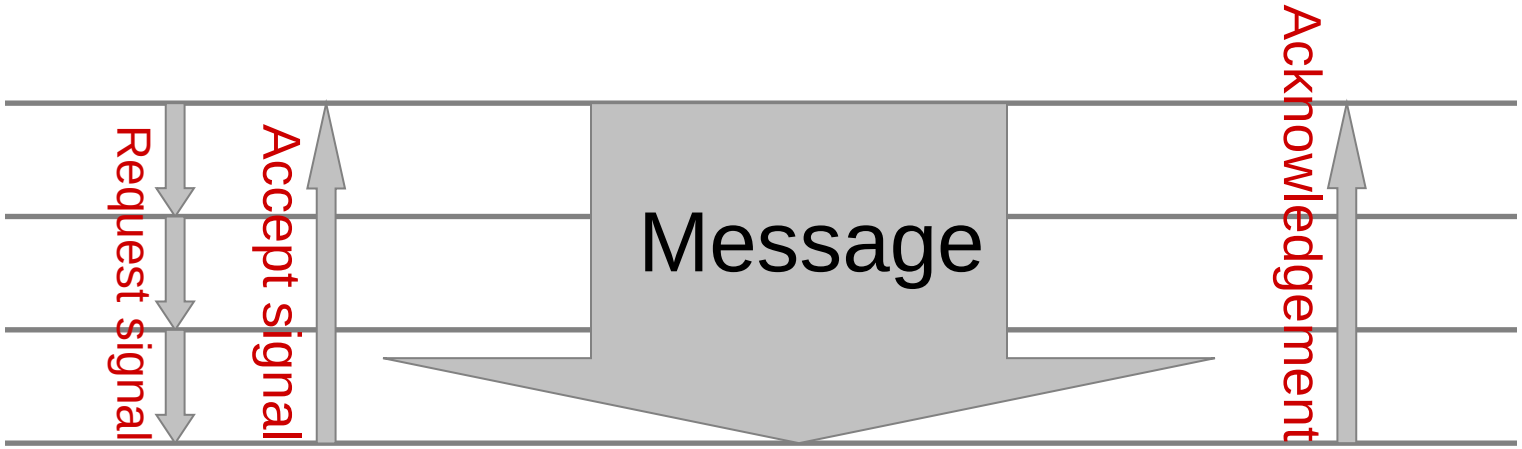
Sending Messages via Switching

- Point to point connect between each pair of devices is impractical
- Better solution is switching
  - Communicating via a series of interlinked nodes called switches/router



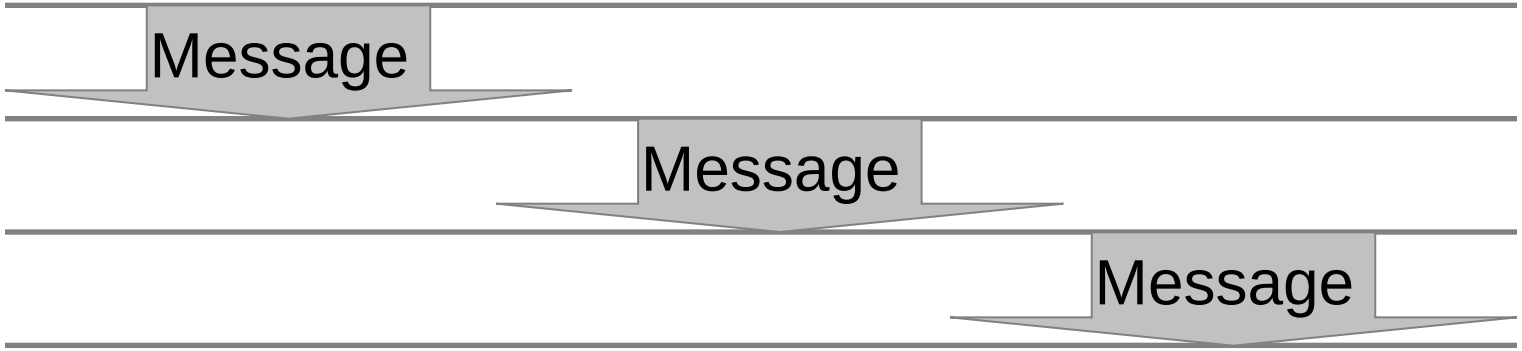
Sending Messages via Switching

- In general, different schemes can be used to co-ordinate the sending of a message via switching
  - Circuit switching – a “direct” connection is established from the sender to the receiver



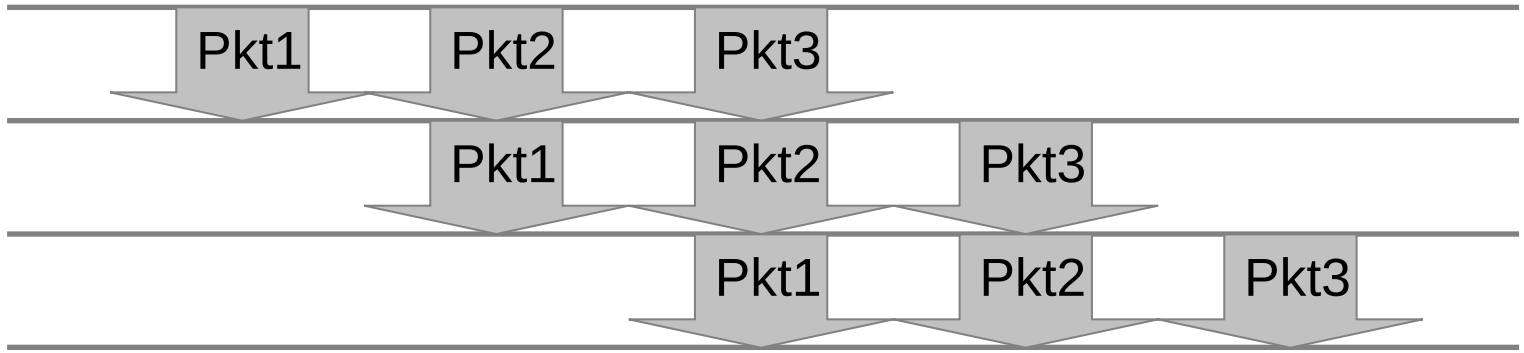
Sending Messages via Switching

- Different schemes (cont')
  - Message switching – the entire message is relayed; no guarantee of delivery



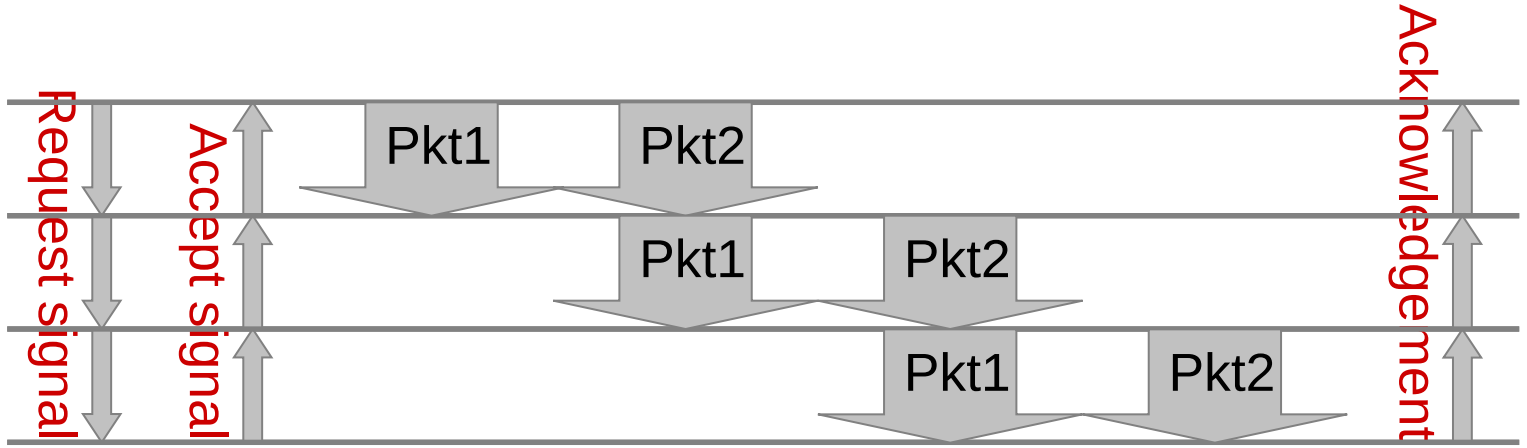
Sending Messages via Switching

- Different schemes (cont')
  - "Datagram" packet switching – the message is split into packets which are relayed; no guarantee of delivery

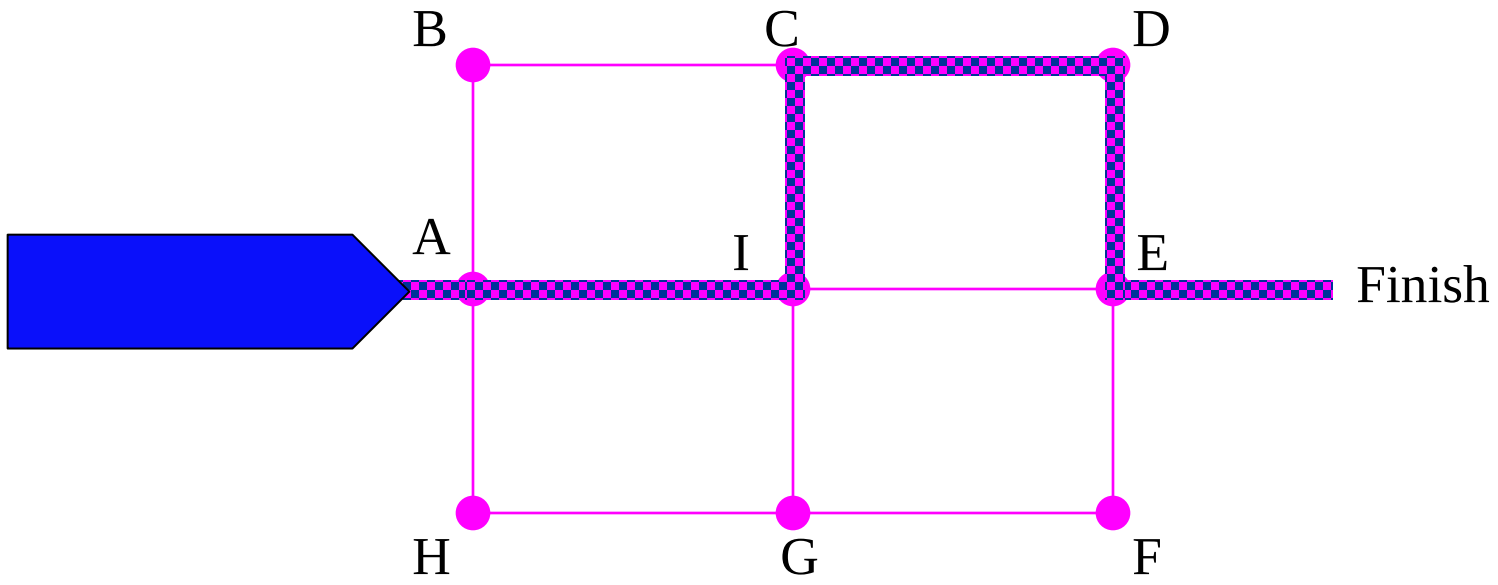


Sending Messages via Switching

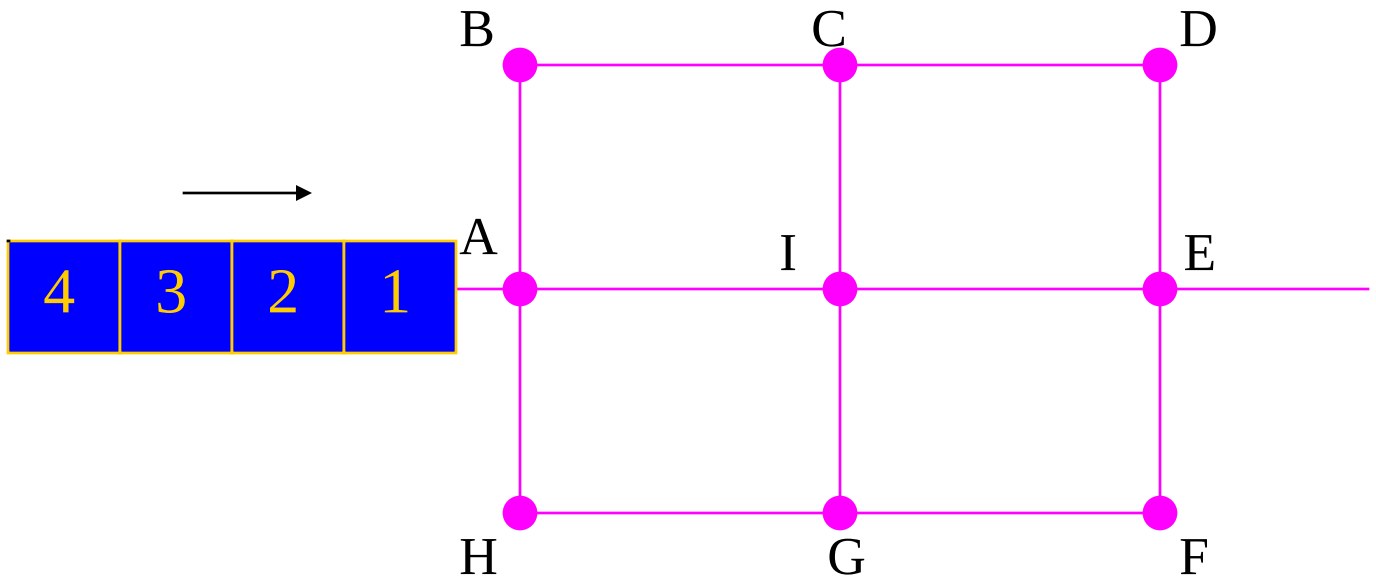
- Different schemes (cont')
  - "Virtual circuit" packet switching – simulating circuit switching using packets



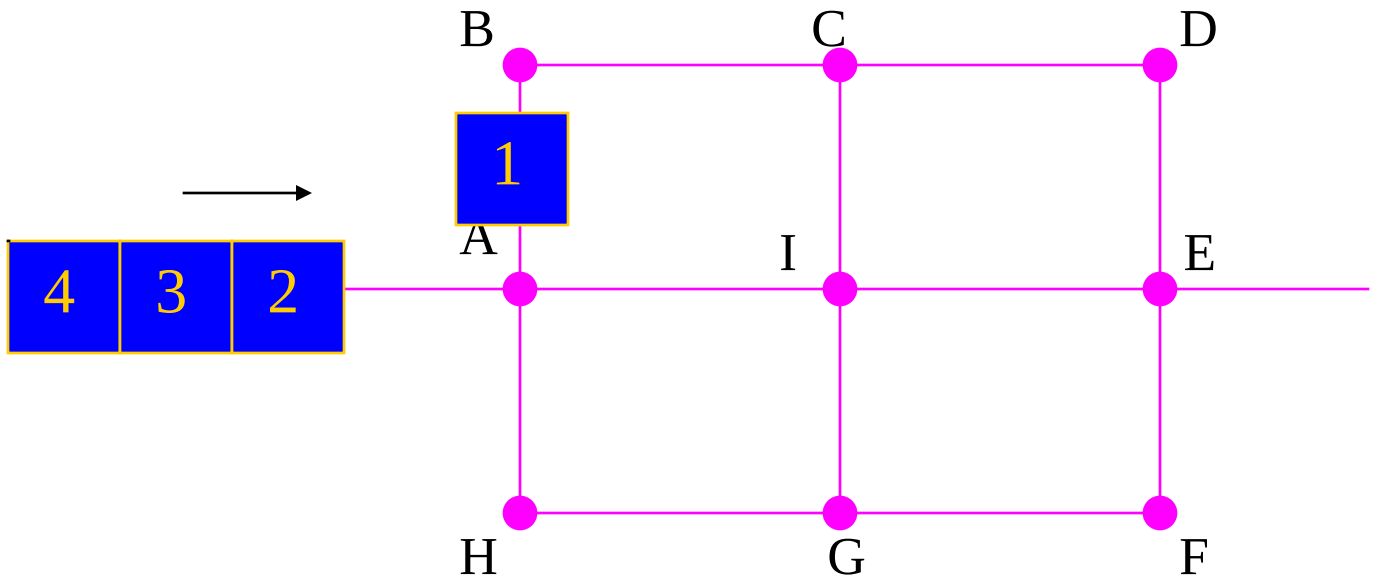
Circuit Switched



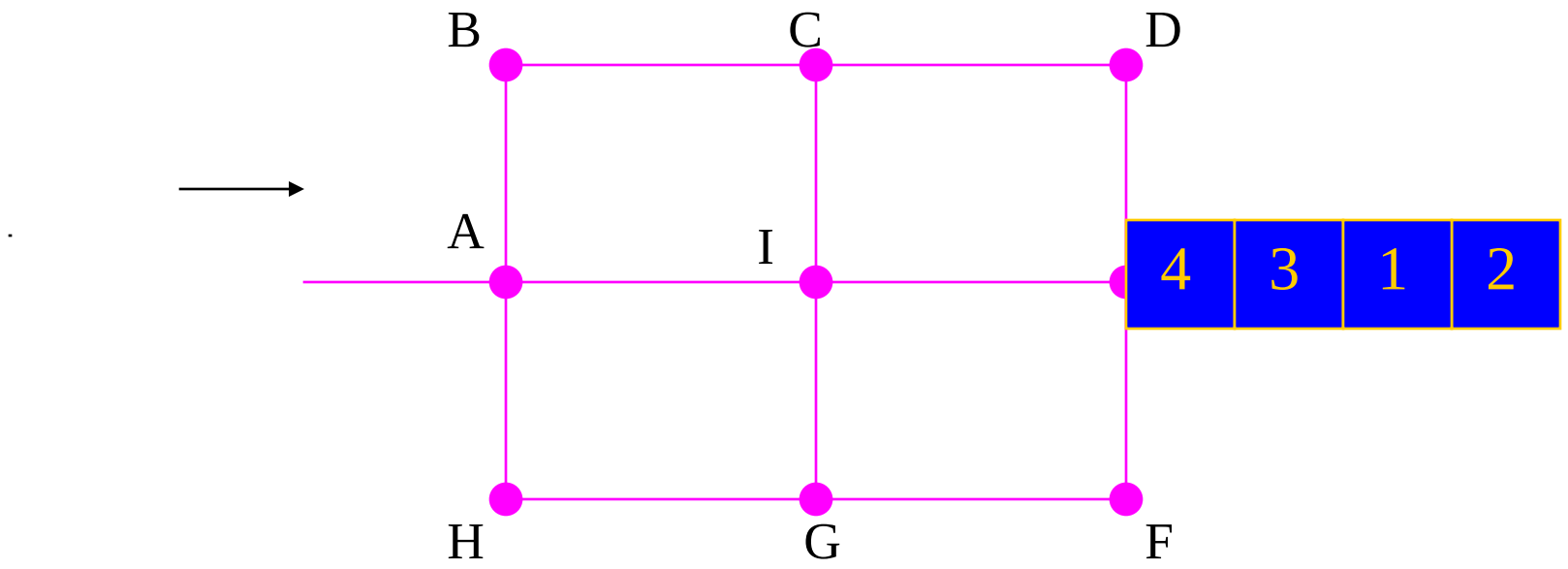
Datagram Packet Switched



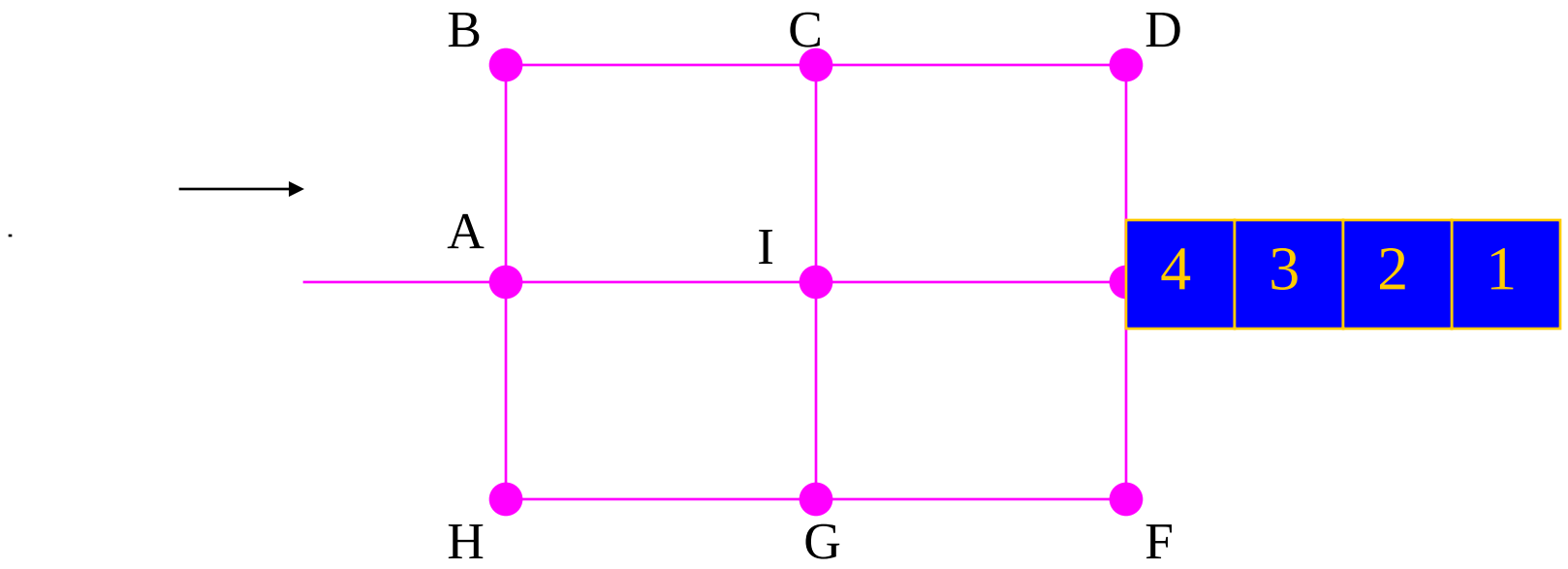
Datagram Packet Switched



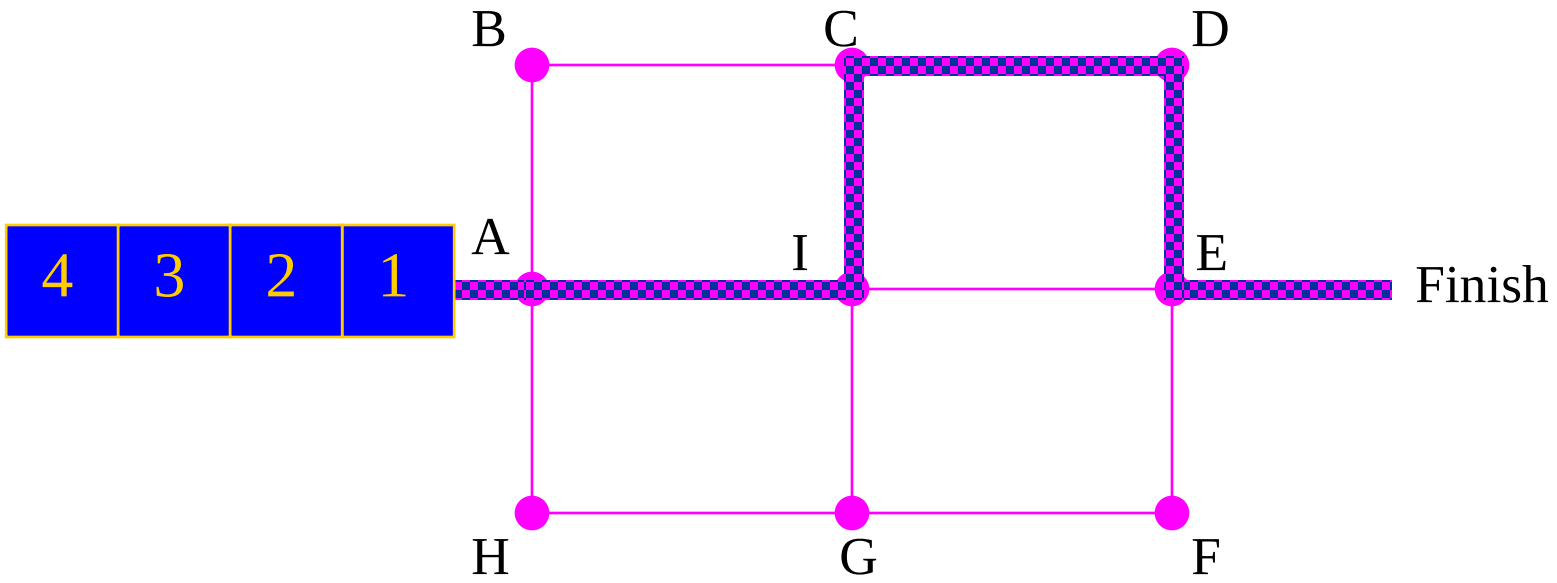
Datagram Packet Switched



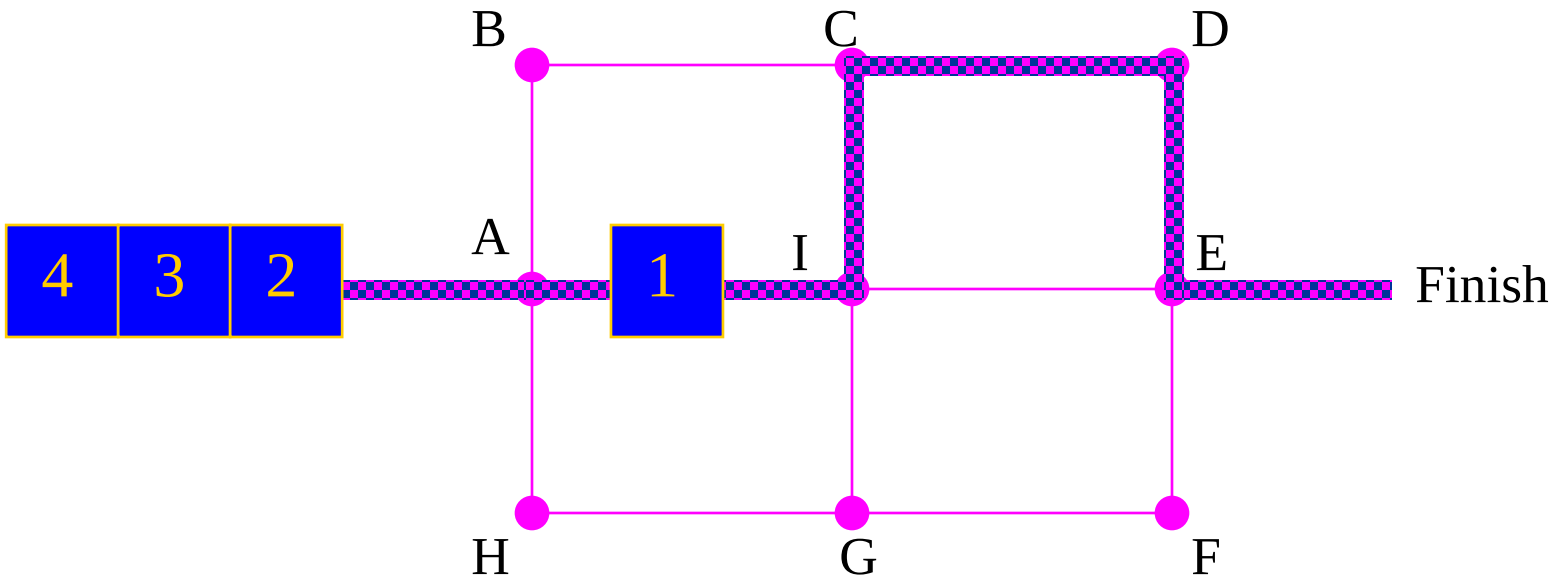
Datagram Packet Switched



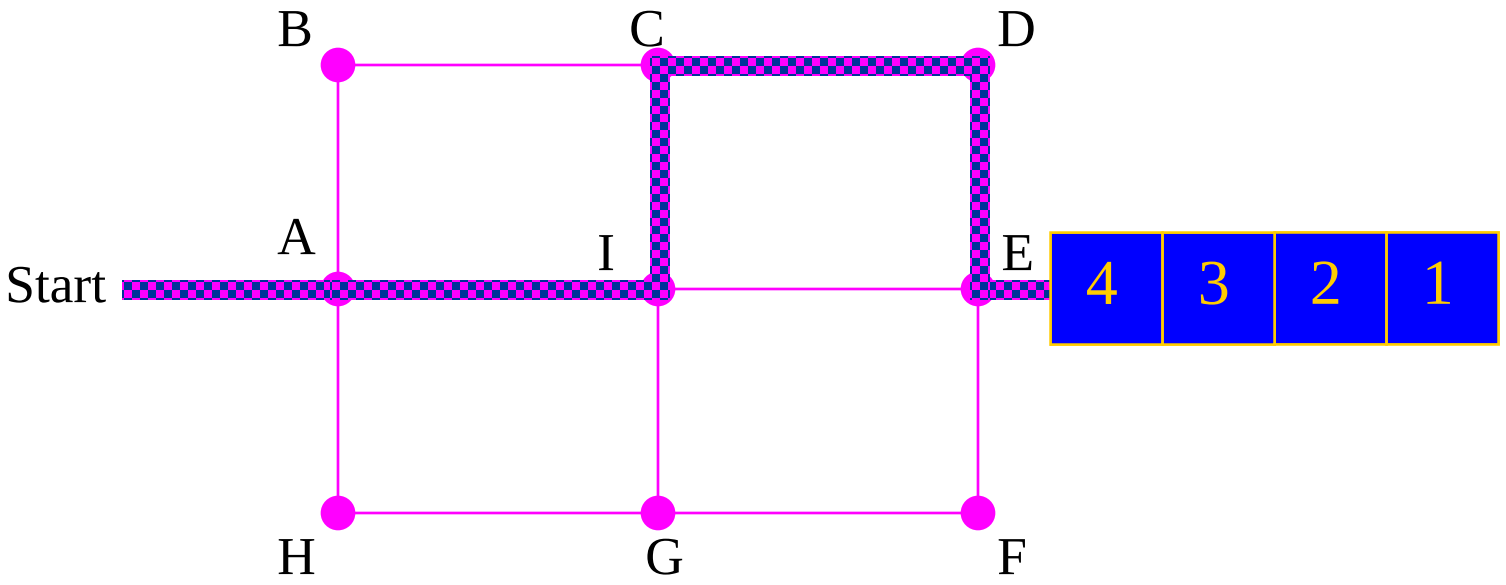
Virtual Circuit Packet Switched



Virtual Circuit Packet Switched



Virtual Circuit Packet Switched

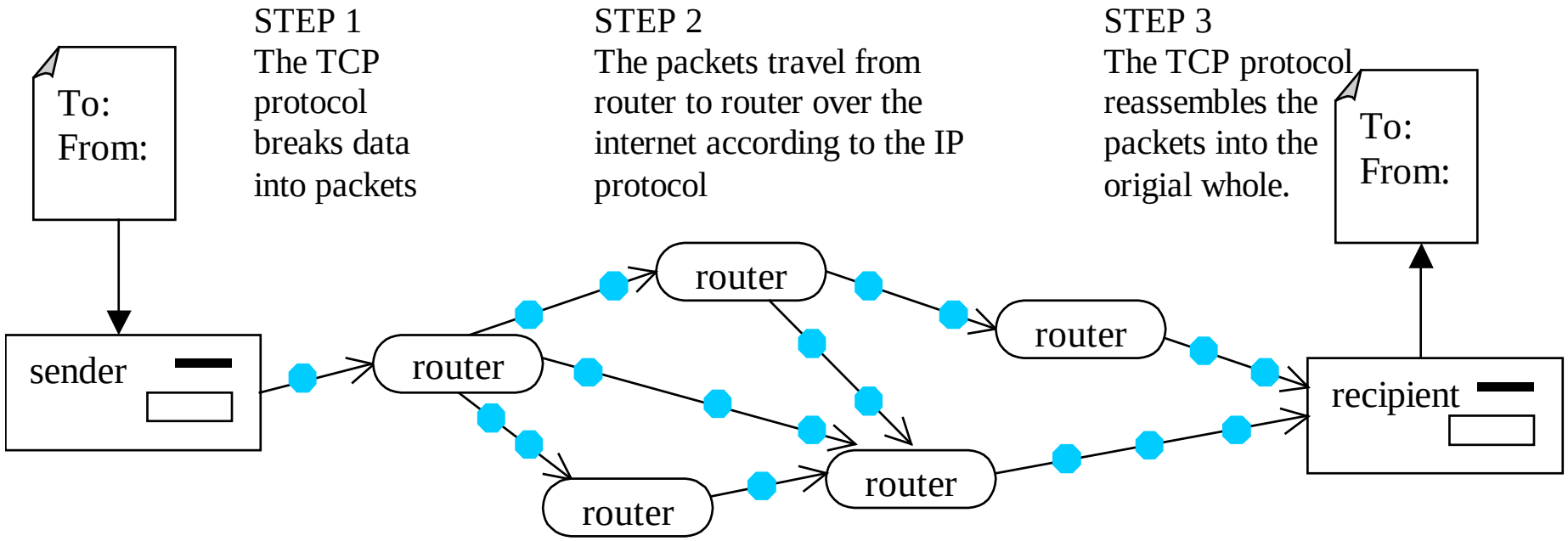


Routers

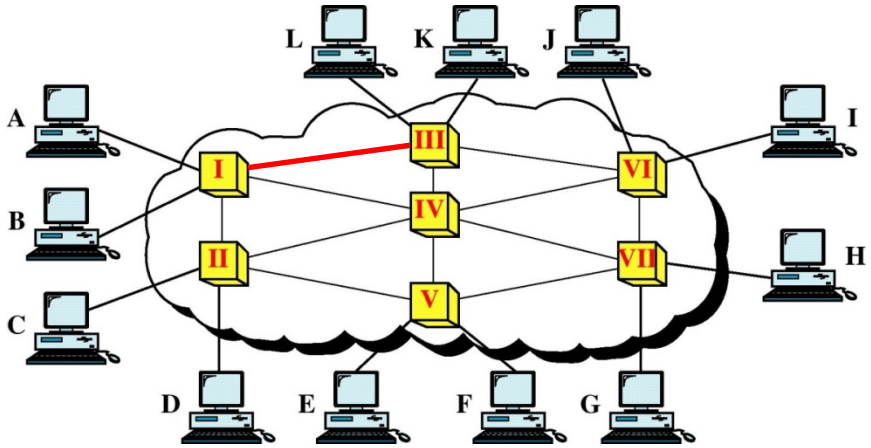
- Router = Gateway = a device connected to several networks = "*building blocks of the Internet*" .
- Routers use headers and a forwarding table (routing table) to determine where packets go.
- They also communicate with each other:
  - Routers determine all possible paths to the destination and select the best route, based on the traffic load and the number of hops.
  - They use the Internet Control Message Protocol, (ICMP) to communicate with each other – ICMP supports error, control, and informational messages.

Internet = Packets + Routing

- Routers do not care what data is in each packet.
- They allow different types of LANs to be connected to the Internet.



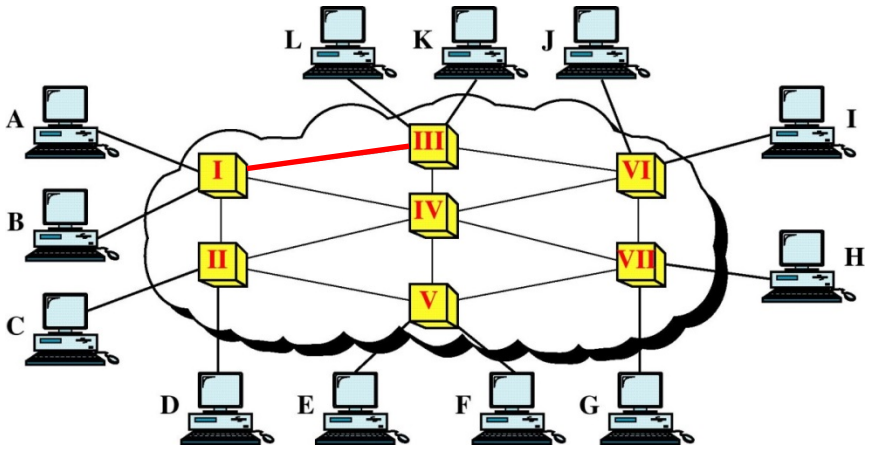
Overrun



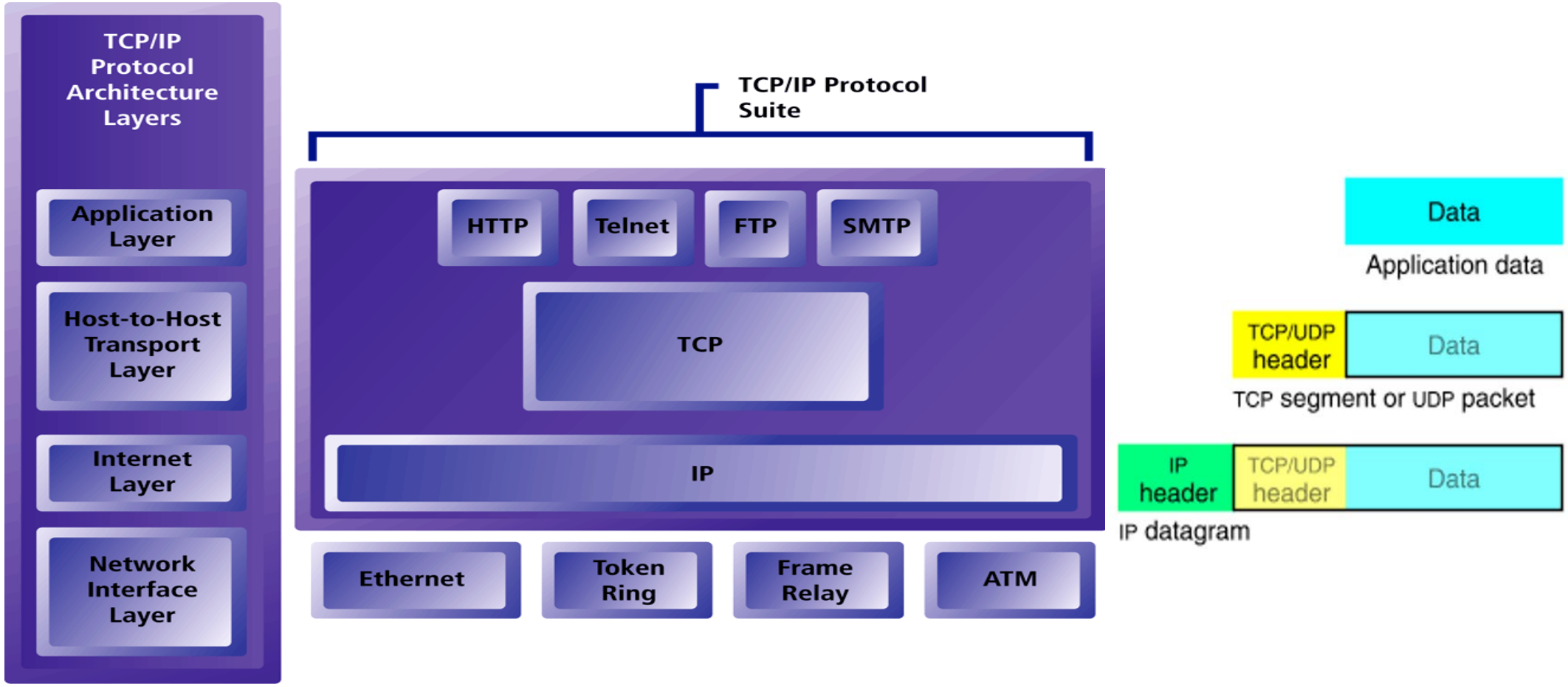
- Example:
  - The capacity of the connection between route *I* and *III* is 5000 packets.
  - Router *I* is receiving 3000 packets from host *A* that it must send to router *III*.
  - Router *I* is also receiving 3000 packets from host *B* that is must send to router *III*.
  - Unfortunately, the *I-III* connection can only handle 5000…

Overrun

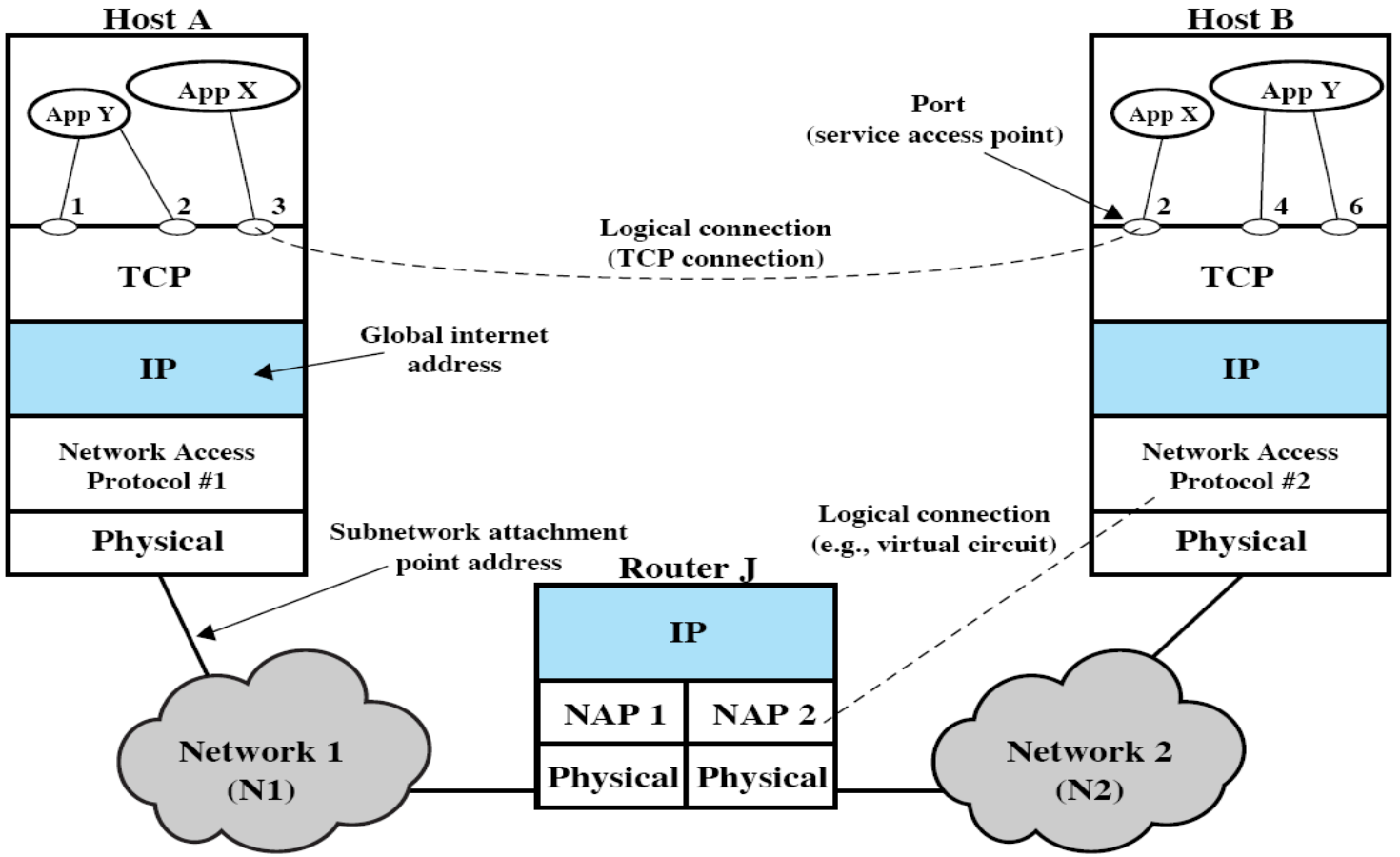
- What happens, a traffic jam?
  - Electrical signals are not like cars!
- **Routers discard (dump) the extra packets.**
  - The loss must be detected and the packets resent.



The TCP/IP Architecture and Protocol Suite



TCP/IP Concepts

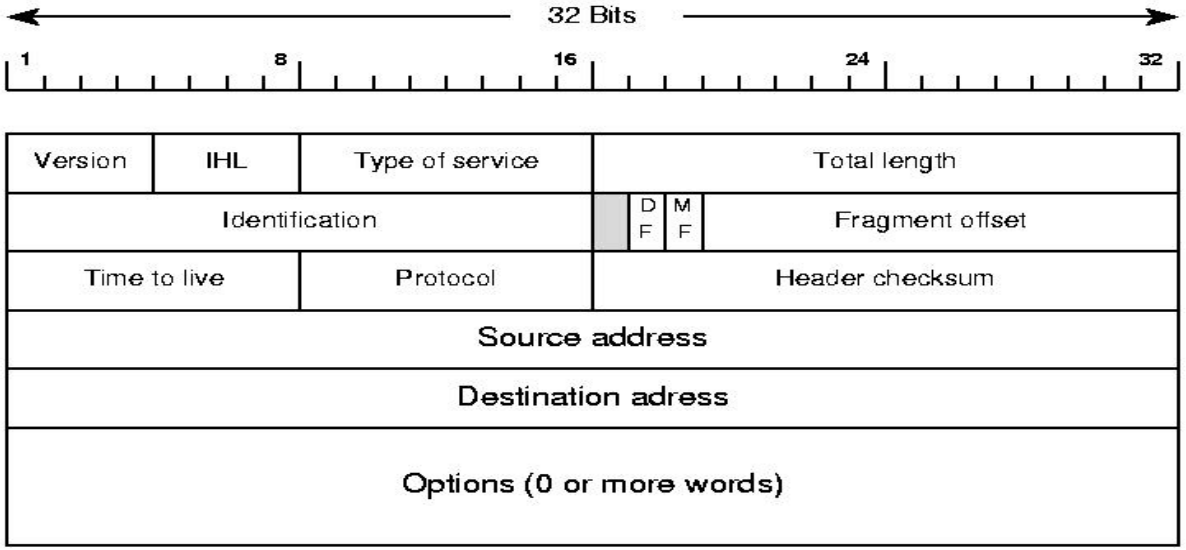


Internet Protocol (IP)

- **Rules** for how packets are routed across the Internet:
  - How to specify source and destination.
  - IP determines “how to get from A to B”.

- A **packet switching** protocol used on the Internet
  - Sends data in individual packets
    - A packet conforming to the IP = “IP datagram”
  - Maximum size of packet is determined by the networks
    - Fragmented if too large
  - Unreliable
    - Packets might be lost, corrupted, or delivered out of order

Internet Protocol (IPv4) datagram header



Field	Purpose
Vers	**IP version number**
Len	Length of IP header
TOS	Type of Service
Total Len	Length of entire datagram
Ident.	IP datagram ID (for frag/reassembly)
Flags	Don't/More fragments
Frag Off	Fragment Offset

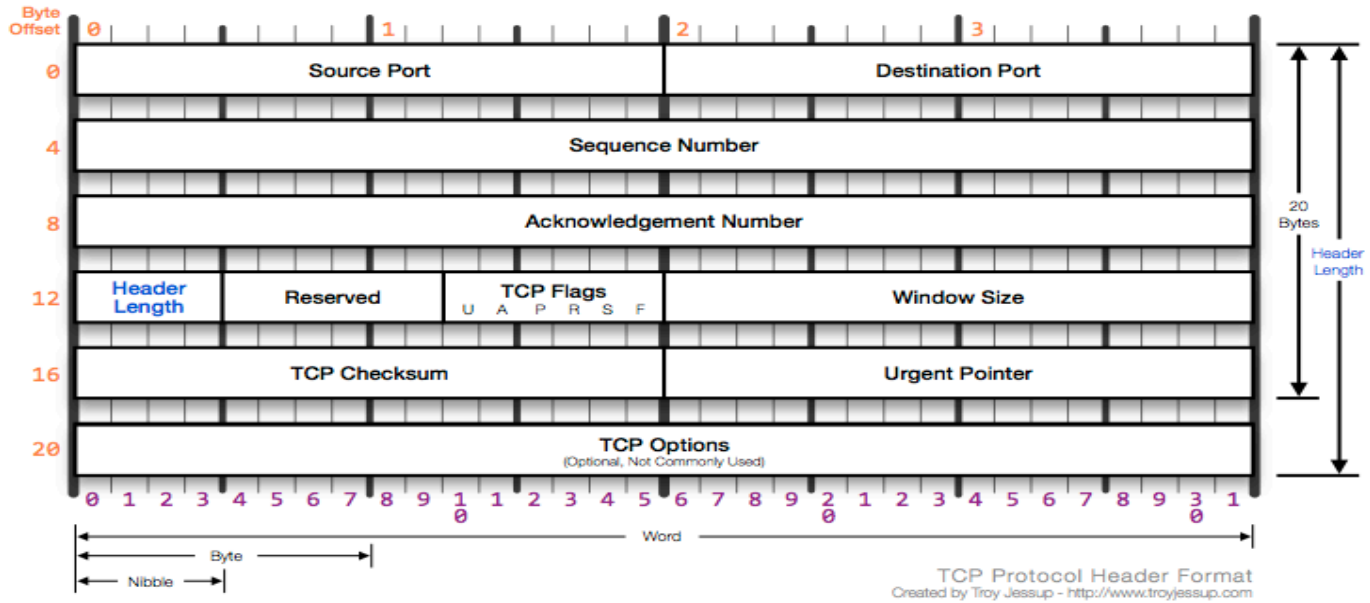
Field	Purpose
TTL	**Time To Live - Max # of hops**
Protocol	Higher level protocol (1=ICMP, 6=TCP, 17=UDP)
Checksum	**Checksum for error detecting**
Source IA	**Originator's Internet Address**
Dest. IA	**Final Destination Internet Address**
Options	Source route, time stamp, etc.
Data...	Higher level protocol data

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Transmission Control Protocol (TCP)

- **TCP provides for reliable communication over packet-switched networks.**
- TCP enables two computers (hosts) to establish a connection and exchange streams of data.
- TCP breaks up the data into packets
- TCP guarantees delivery of data in the same order in which they were sent.
- TCP provides control for rate of transmission.
- TCP tells how big a datagram it can handle during the connection creation

Transmission Control Protocol (TCP) Header



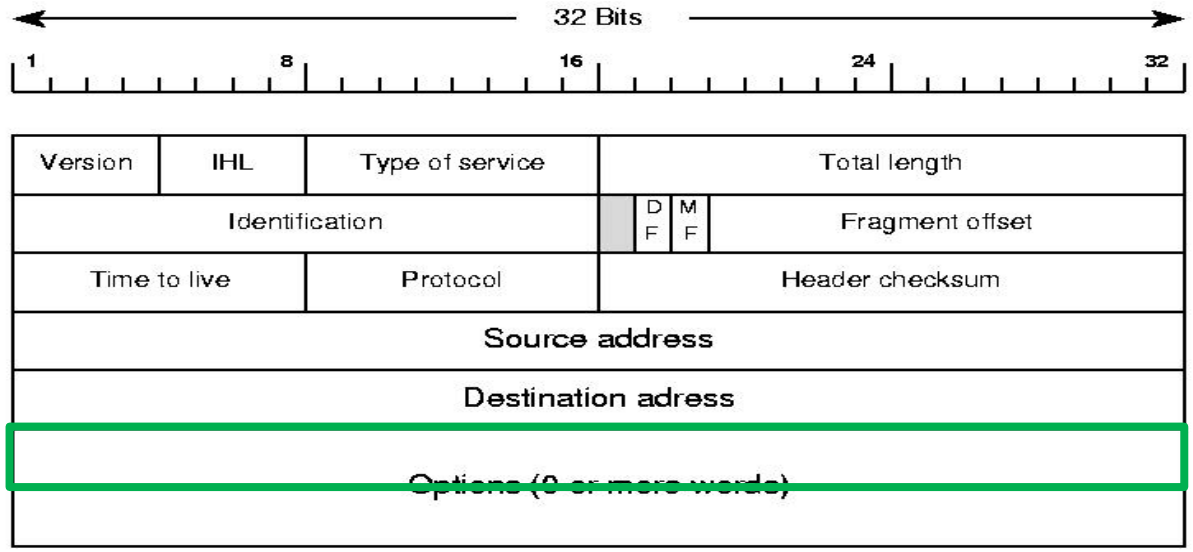
- Ports and Protocols
  - 20 FTP
  - 22 SSL
  - 23 Telnet
  - 25 SMTP
  - 53 DNS
  - 80 HTTP
  - 8008, 8080 alternate ports for HTTP

Field	Purpose
Source Port	Identifies originating application
Destination Port	Identifies destination application
Sequence Number	used by receiver to sort datagrams into the right order
Acknowledgment #	if the sender doesn't get this within a reasonable amount of time, it sends the data again
Header Len	Length of TCP header
Flags	TCP flags: SYN, FIN, RST, PSH, ACK, URG
Window	Control how much data can be in transit at one time
Checksum	**"sum" of headers bytes – used to detect errors.**
Urgent Pointer	eg., Ctrl-C
Options	Special TCP options such as MSS and Window Scale

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

IP Addressing

- IP address = a 32-bit number that identifies each sender or receiver of the packets sent across the Internet.
  - Each host on the Internet has an IP.
  - IP addresses are usually expressed as four period-separated decimal numbers (each 8 bits). E.g., 134.148.96.137



IP Addressing

- An IP address has **two parts**:

  Part1: An Internet network number;

  Part2: The local number of a particular device (server, workstation, printer, etc.) within that network.
- There are four classes of address:
  - Class A: (0) Network (7bits) Local Address (24bits)
  - Class B: (10) Network (14bits) Local Address (16bits)
  - **Class C: (110) Network (21bits) Local Address (8bits)**
  - Class D: (1110) Multicast Address (28bits)
  - Class E: (1111) Reserved for Experiments (28bits)
  - Class C is most common and is fast running out.
- 2 level structure is wasteful
  - Can we have 2^{32} (the full 4 billion addresses) different IP addresses? Only in theory – some numbers may never be used.

Too many computers!

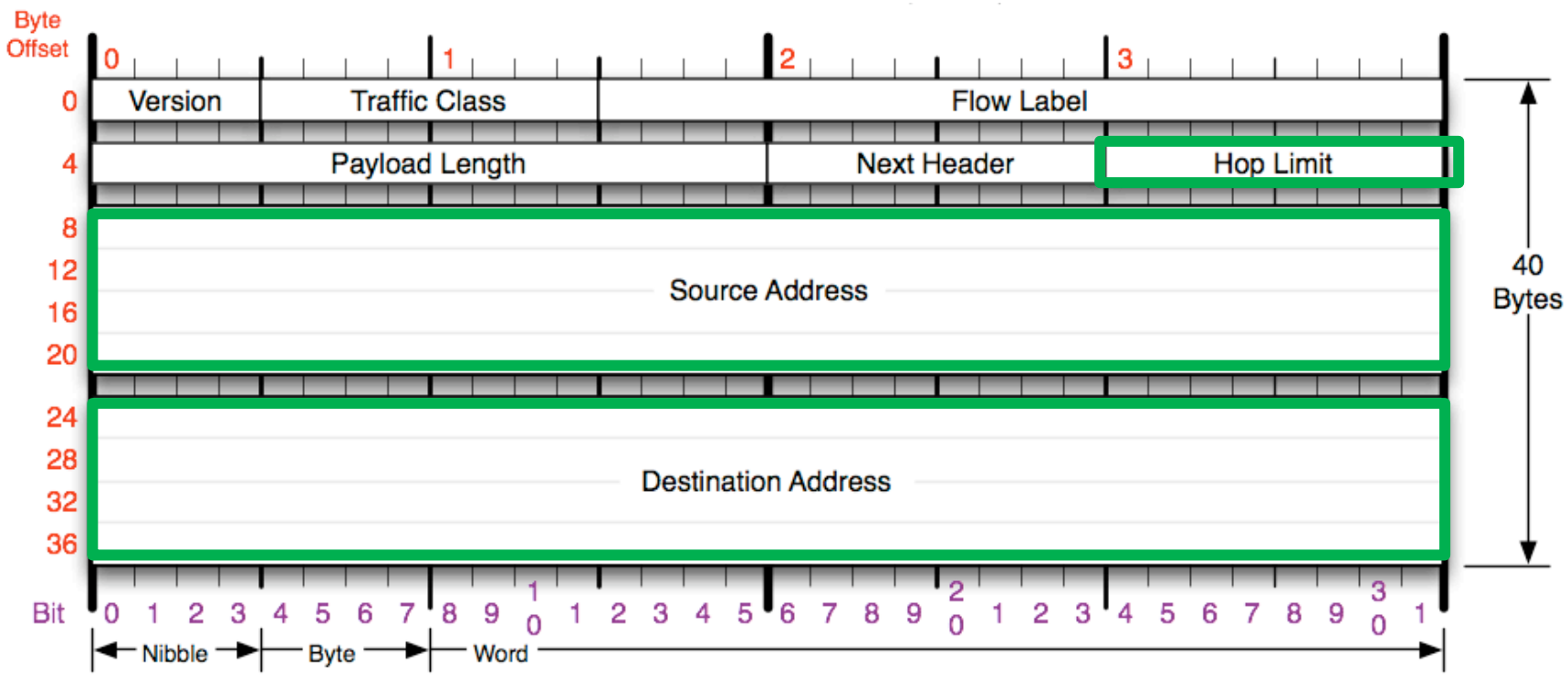
- Problem: What happens when we run out of IP addresses?!
  - This was identified many years ago
  - Adjustments were made to the IPv4 protocol to allow for an *endless* number of computers
  - The new standard is called IPv6, and uses **128-bits** to identify each computer instead of 32-bits
  - IPv4 and IPv6 can work together
  - Users are slowly converting from IPv4 to IPv6, and will be the default in upcoming operating systems (LINUX, Windows Vista, etc.)

IPv6

- At 128 bits provides a maximum 2^{128} hosts.
- Roughly 3.4 x 10^{38}
- *“Even if every human now living were to have a personal network, with a billion nodes on each network, the IPv6 address space is large enough to support (at least theoretically) roughly another 50 billion billion similarly wired planets”*

  » Pete Loshin – ‘TCP/IP Clearly Explained’, Morgan Kaufman, 1999

IPv6 Header

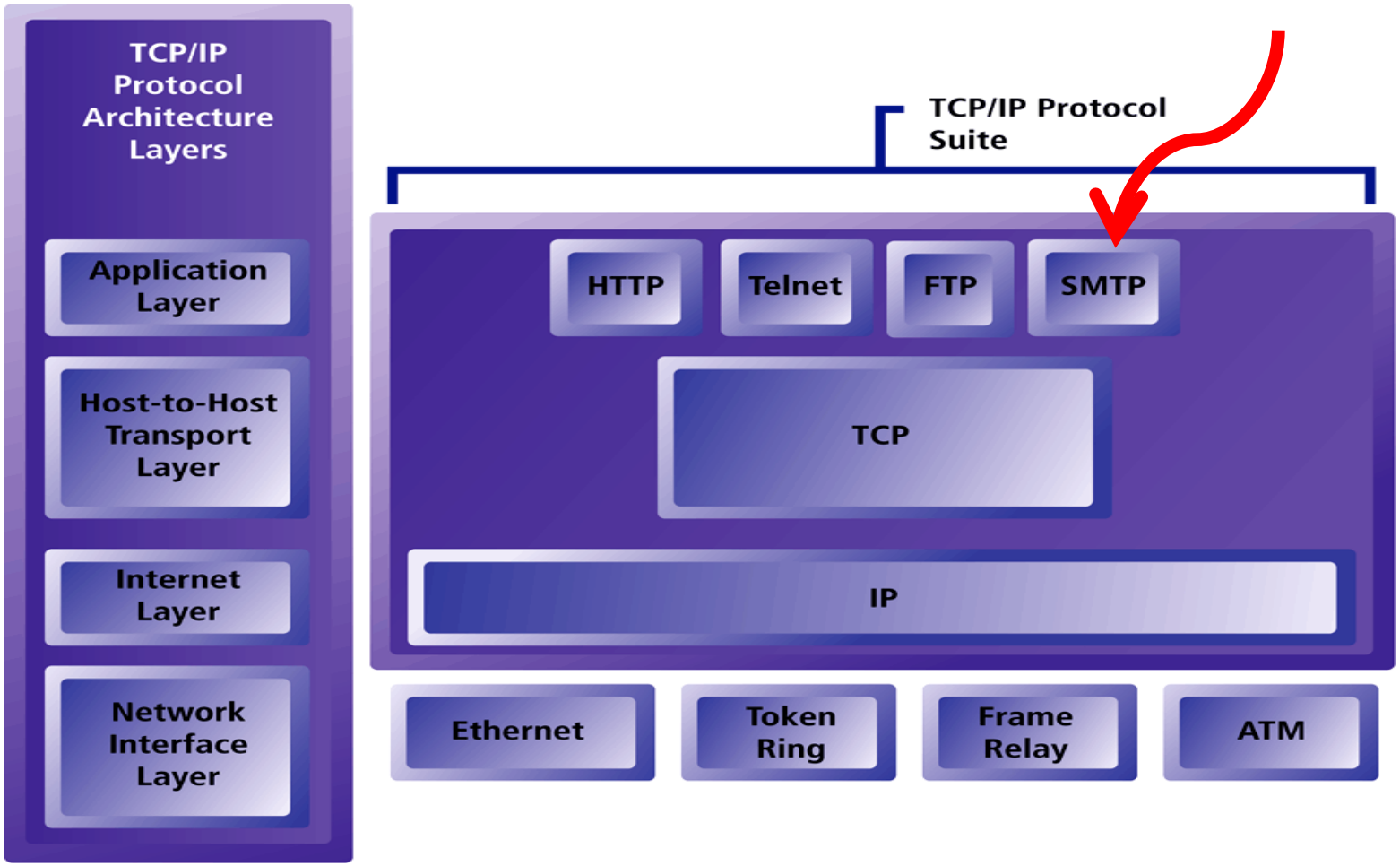


- IPv6 fixed header (40 bytes).
- RFC 2460 - http://www.faqs.org/rfcs/rfc2460.html

IP Addressing – Domain Name System (DNS)

- DNS provides a symbolic equivalent for numeric IP address
  - 134.148.96.137 = lily.newcastle.edu.au
  - Each computer knows (at least) one local DNS server
- Each local server knows the IP number ↔ name mappings for its local network
- If the requested mapping is outside of the **local domain**, then the local server asks a "root server", which relays the to the appropriate local server

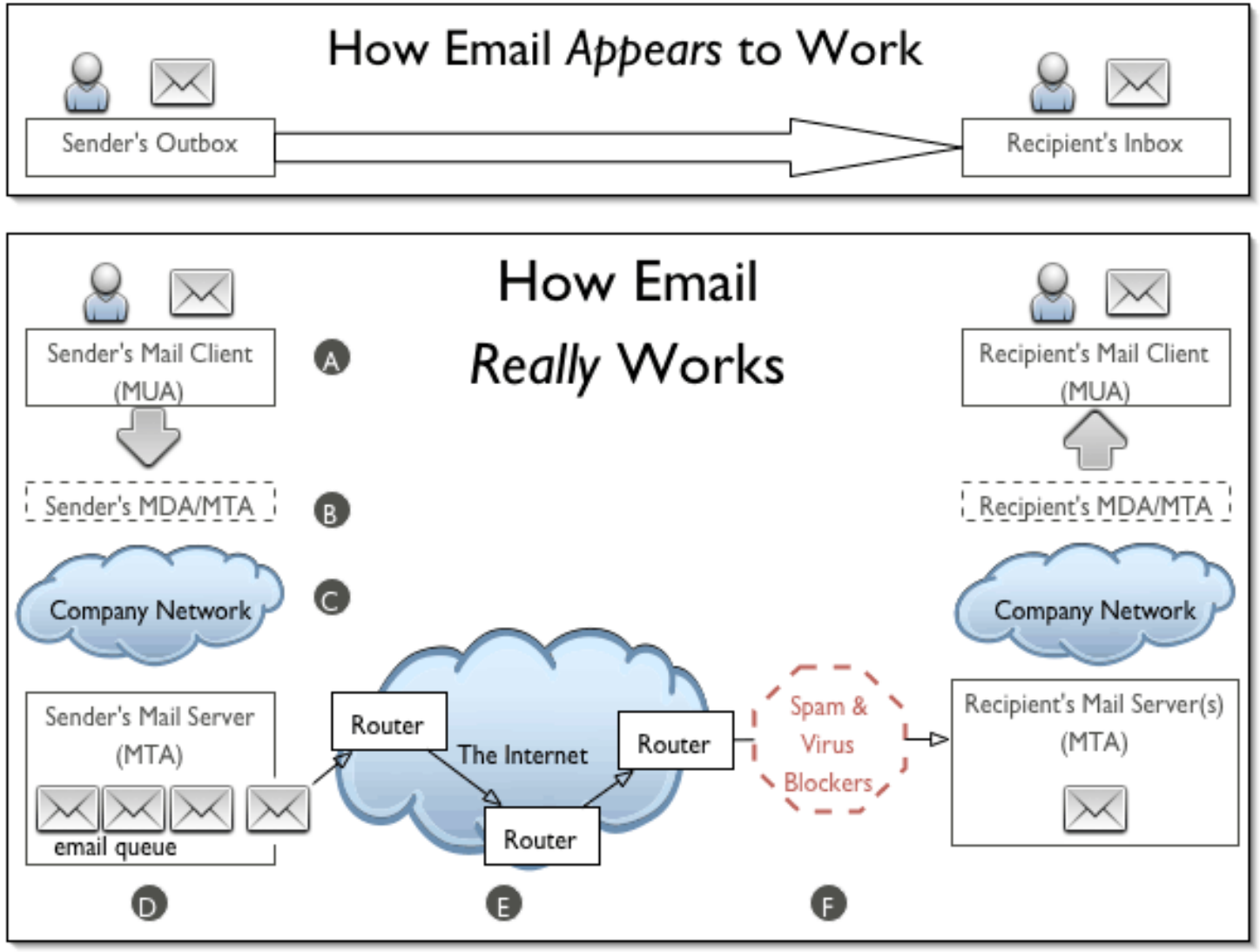
The TCP/IP Architecture and Protocol Suite



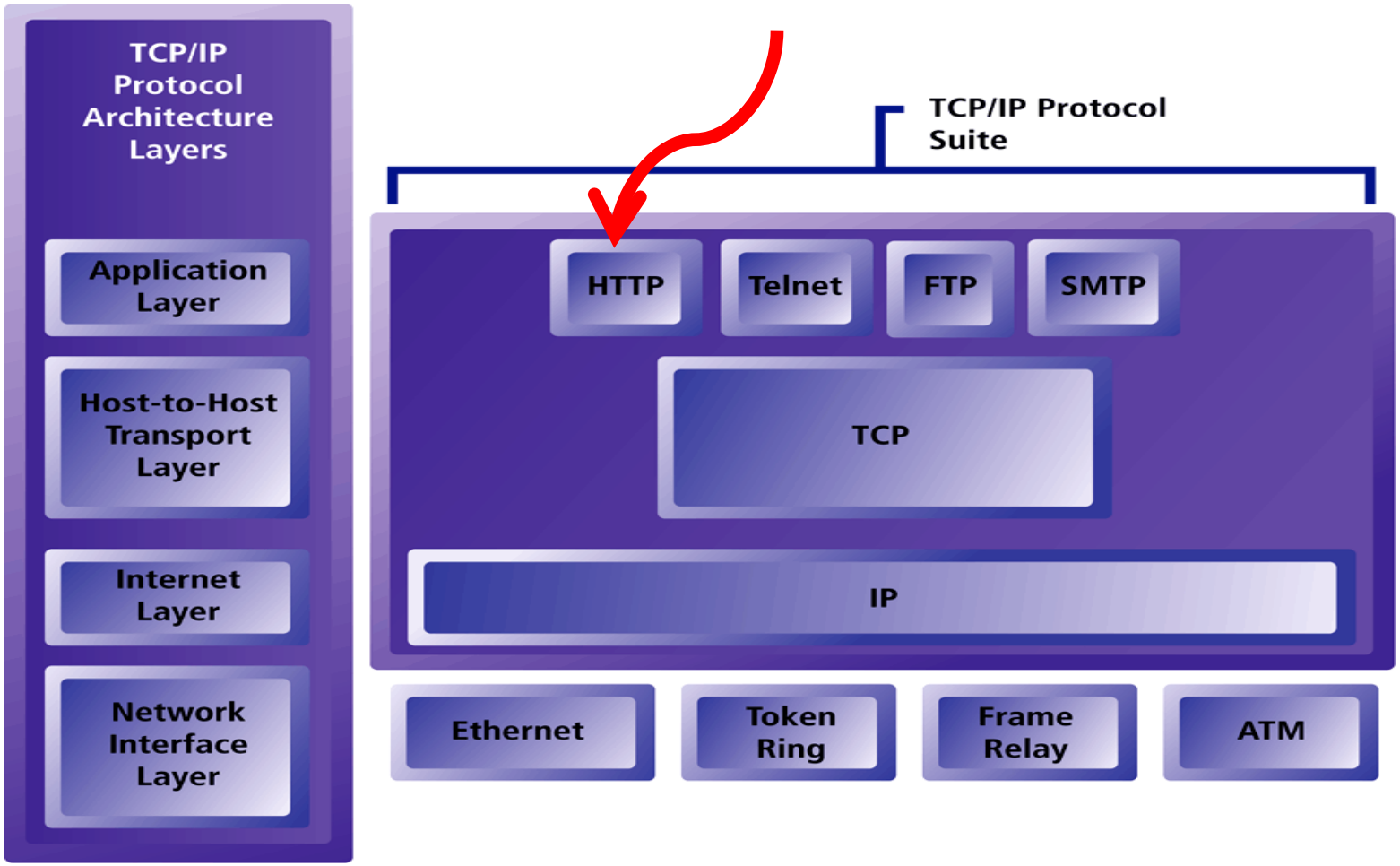
Email

- E-mail is the most popular Internet application (?)
  - The first e-mail was sent on in 1971 by Ray Tomlinson
- **Simple Mail Transfer Protocol (SMTP)** is a protocol for sending e-mails over the Internet
  - SMTP conforms to TCP/IP (packets, etc.)
  - Delivery time under SMTP is not guaranteed – for urgent matters use instant messaging or a phone!
- There are many different e-mail programs
  - You have access to StudentMail, (you@uon.edu.au) but you can "forward" your mail to another service if you like

What happens to Email?



The TCP/IP Architecture and Protocol Suite

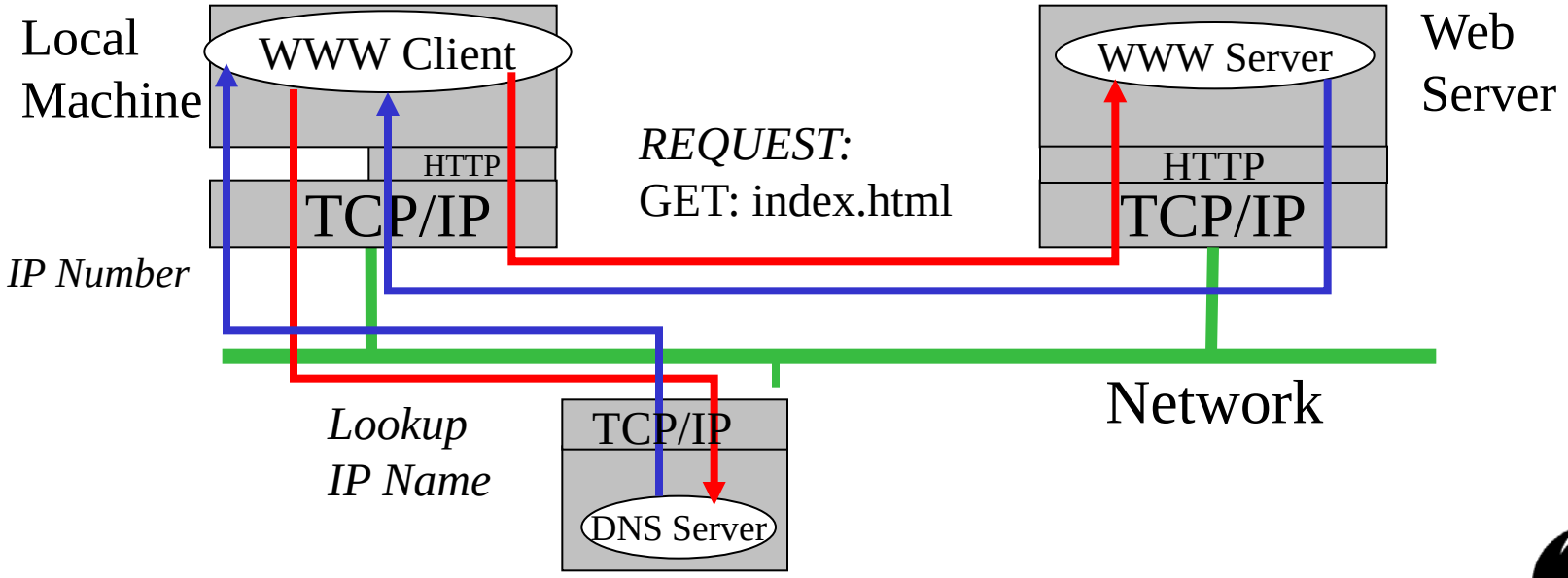


Hyper Text Transfer Protocol – HTTP

- Based on a *request* / *response* paradigm
- "*the lightness and speed necessary for distributed, collaborative, hypermedia information systems*"
- Generally run over TCP/IP
  - Remember, WWW came along *after* the Internet
- Requests addressed using URIs

HTTP Over TCP

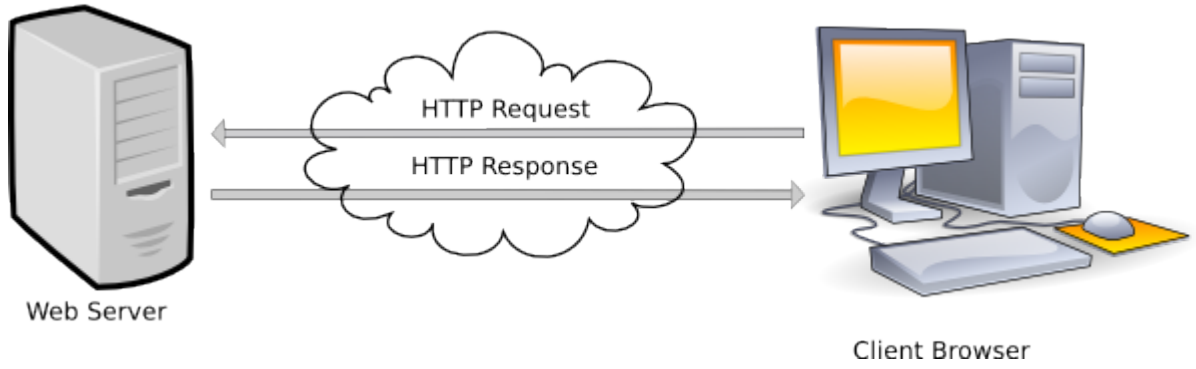
- Client extracts IP address from URL
- Client performs DNS lookup of IP address
- Client initiates TCP connection to port 80 on IP address
  - Port 80 is the default port for WWW connections



HTTP/1.0

- Used in hypermedia systems distributed across network
- Message types: **requests** and **responses**
  - **Requests** are made to get information
  - **Responses** are replies to *Request messages*

HTTP-message =
Simple-Request | *Simple-Response* (HTTP/0.9, no header)
| *Full-Request* | *Full-Response* (HTTP/1.0 or HTTP/1.1)



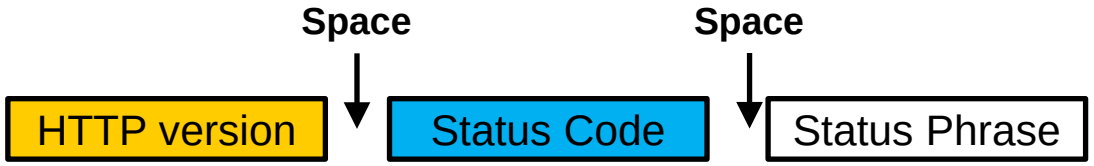
HTTP/1.0: Request Line

Example: `GET /path/to/file/index.html HTTP/1.0`

- Kinds of requests
  - **GET** – returns whatever information is requested
    - *Simple-Request* = `GET` *Request-URI*
  - **POST** – includes a body message
    - E.g., used to send “forms” to the server
  - **HEAD** – returns the *Request-URI* header information
    - This has *many* uses
  - Also **PUT**, **DELETE**, **TRACE**, **OPTIONS**

HTTP/1.0: Response

- First line of *Full-Response* is the `Status-Line`



Example: `HTTP/1.0 200 OK`

- **1***xx*: Informational
  - Not used, but reserved for future use (HTTP/1.1)
- **2***xx*: Success
  - The action was successfully received, understood, and accepted
- **3***xx*: Redirection
  - Further action must be taken to complete the request

HTTP/1.0: Status Line

- **4***xx*: Client Error
  - The request contains bad syntax or cannot be fulfilled
  - `401 Unauthorized` – you need a password
  - `403 Forbidden` – chmod!
  - `404 Not Found` – files doesn't exist (or 403!)
- **5***xx*: Server Error
  - The server failed to fulfill an apparently valid request
  - `500 Internal Server Error` – e.g., bad script!

HTTP/1.0

- HTTP/1.0 transfers **1 file at a time!**

- E.g., To view HTML file with two `<img />` tags
  - The HTML file is transferred as a HTTP message
  - Then the 1st image is transferred as a HTTP message
  - Then the 2nd image is transferred as a HTTP message

- HTTP uses one TCP connection per file transfer
  - Why? HTTP is a **stateless protocol**
  - Involves setting up and bringing down the TCP connection for each file transferred
  - HTTP/1.1 addresses this…

HTTP/1.1

- http://www.faqs.org/rfcs/rfc2068.html
- Enhances HTTP/1.0
- Generic Protocol – can run SMTP, FTP, etc. over HTTP

1. Persistent Connections
2. Supports Caching
3. Supports Proxy Servers

HTTP/1.1 – Persistent Connections

- E.g., To view HTML file with two `<img/ >` tags
  - Setup the TCP connection
  - The HTML file is transferred, then the 1st image is transferred, and the 2nd image is transferred
  - (The TCP connection is closed or times out)

- Creating/closing a TCP connection takes time
  - TCP was designed to work most efficiently for large transfers – not for small HTML and GIF files
  - Reusing the connection saves CPU time for the client, the server and on routers, and reduces network traffic

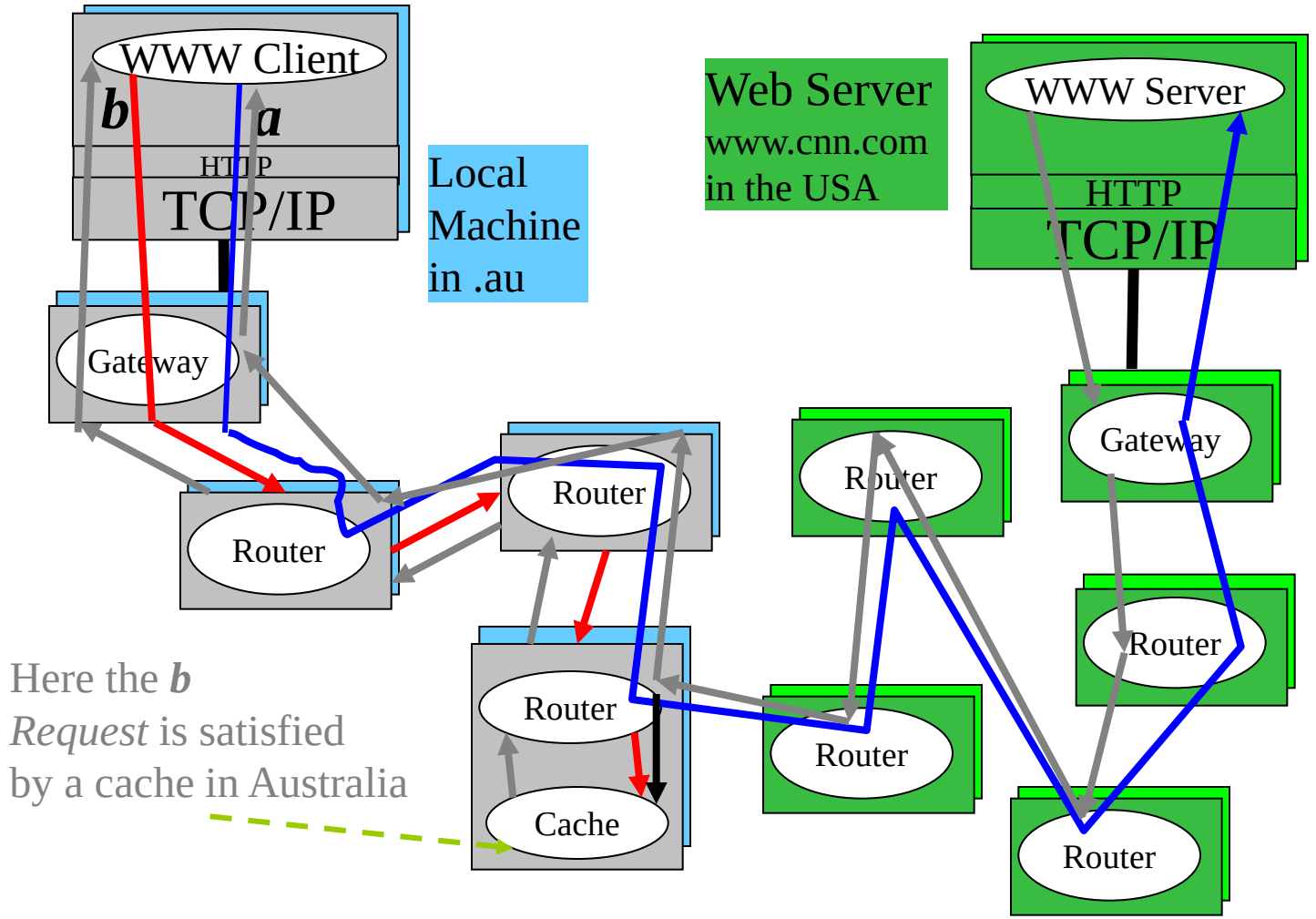
- However, malicious programs can tie-up TCP connections = Denial of Service attack

HTTP/1.1 – Caching

- Some browsers cache recently viewed pages on the local disk
  - You may need to click on "**reload**" to see the new version of a page, if the cache holds the old version

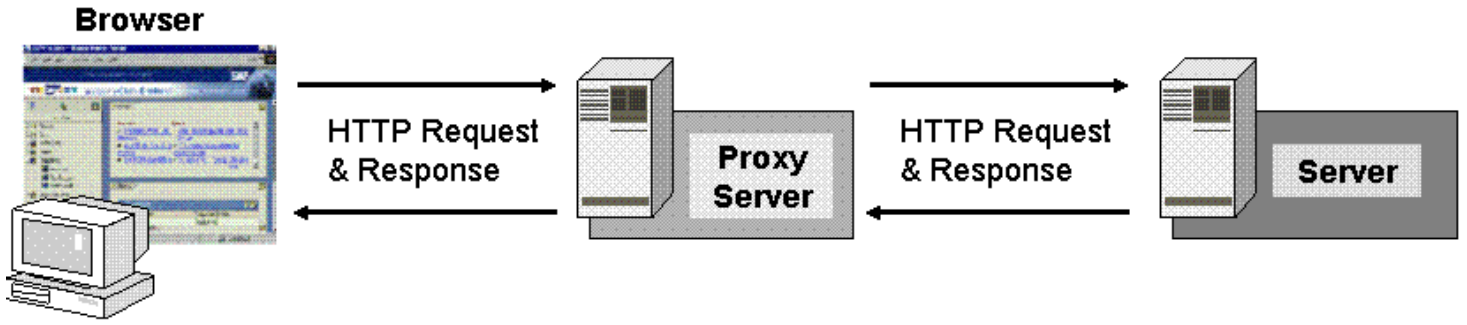
- Deciding what, where and when to cache is a big area of on-going research
  - **Post-fetch**: Caches the file when it is first requested
  - **Pre-fetch**: Attempts to cache a file **BEFORE** it is requested

HTTP/1.1 – Caching Example



HTTP/1.1 – Proxy Server

- Acts as a buffer between your intranet and the Internet
  - Allows central control of local Internet traffic
  - Proxies send requests to / receive responses from Internet servers on behalf of machines on its network
  - Usually does caching for entire intranet
  - Can block messages to/from proscribed Internet sites
    - E.g., http://an-ad-site.com/
    - E.g., http://a-porn-site.com/
  - Can filter messages for viruses



Summary

- Different types of computer networks (LAN, WAN)
- Network topologies (Bus, Ring, Mesh, Star, Tree)
- Internet and WWW
- Grid, Clouds
- The difference between circuit switching and packet switching
- Layers of TCP/IP model
- How TCP and IP works together to transfer data
- IP addressing system
- Role of Domain Name System
- How Email works
- The working principle of HTTP

References

- “Data Communications and Networking” by Behrouz A. Forouzan, (4th Ed), (library)
  - [Sec 8.1~8.3 (P213-P227); Sec 2.4 (P42-P45); Sec 20.2 (P582-P588); Sec 23.3 (P721-P724); Sec 19.1 (P549-P554); Sec 19.2 (P566-P568); Sec 25.1~25.2 (P797-P801); Sec 26.2 (P824-P828); Sec 27.3 (P861-P867)]
- “Internet FAQ Archive”
  - http://www.faqs.org/faqs/
  - Includes RFCs http://www.faqs.org/rfcs/