TELE 9301 Switching System Design

Handouts

Please pick up a copy of the handout as you enter. Handouts are located in the front row by the aisles. Either

• one copy with “double-sided” written at the top
  or

• one copy with nothing written at the top + one copy with “flipside of single-sided copies” written at the top

All slides from lectures will be available on the course web page in PDF format.
TELE 9301
Switching System Design

Session 1 2006
Course Coordinator: Tim Moors
Outline (for today)

Administration
Introduction
  Motivation for (& definition of) switching
  Evolution of networks
  Network elements
  Layering
  Switching at various layers
Types of information in a network
User characteristics/requirements
  Spatial aspects
  Volume aspects
  Temporal aspects
Initiation
  “Quality of Service”
Other aspects
What do you want to learn from this course?
Focus & form of the course

Objectives:

Appreciate the reasons for switching†, and the relative merits of the possible switching modes.

Understand the principles of the internal design and operation of communication switches†.

Understand the essence of the key protocols that are used with switched networks.

Emphasis on understanding concepts rather than nitty-gritty detail of every protocol

Mainly descriptive ☹; no labs; no switch configuration / sales

† Yet to be defined. Includes products marketed as “routers” as well as “switches”.
Course outline

Administration
Introduction to Switching
Fabrics
  Switching fabrics: Overview
  Switching fabrics: Time-division, Space-division
  Optical switching
Functions surrounding fabrics
  Packet classification
  Forwarding modes and buffer management
  Traffic management and scheduling
Distributed switching and protocols
  Bridges
  Prioritisation, Virtual LANs, Virtual Private Networks
  Signalling principles
  SS7 signalling and public telephone networks
  Early public data networks: X.25, End-to-end arguments, Frame Relay
  Asynchronous Transfer Mode / MPLS
  Integrated Services (RSVP) and Differentiated Services
  2006?: Caching?
Resources

Course presentation
• Lectures
• The lecturer

Electronic resources

Text book

Finding other material
Lectures

Slides will be available after the lecture on the course web page in PDF.

- “after”:
  - Notes from last year are available on the course web page now, well before lectures & are very similar to this year’s notes.
  - This year’s notes are being revised until just before the lecture, so I can’t guarantee availability before the lecture.
- “PDF”: This allows people to view them without buying Microsoft products, but sacrifices Powerpoint’s timing information for animations.
- “Video” of lectures (audio + screen capture) will also be available online.
  - Bulky ⇒ only accessible within UNSW

Slide numbers

- Use these to relate notes taken in lectures to slides printed out after lectures.
- May skip certain numbers (slides that are still “under construction”)
Participation

You don’t have to attend lectures, but:

• If you do attend, then don’t distract others (be quiet).
  • If you have to talk, do it outside.
• All material covered in lectures is examinable.
• Bonus marks may be awarded for participation.
Questions

Hopefully your exposure to the material in this course will raise many questions, and you are encouraged to seek answers to these questions.

Questions about course subject matter:
1. Ask another student.
2. Look through the textbook to try to find an answer.
3. Ask the question in the appropriate webCT discussion forum & wait a while for someone to answer.
4. Ask the lecturer verbally during consultation time.

Questions about the course administration (e.g. your marks for the test) to the lecturer...
The lecturer

He talks too fast...

**Do:** Best way to contact:
1. During consultation times
   in 2006: Fridays 3-4pm in office 341 of building G17
2. By email to
tele9301admin AT ee.unsw.edu.au
   include “TELE9301” in subject

**Don’t try:**
- Email to other addresses
- Telephone
- Visiting office outside consultation times

These avenues simply don’t scale well enough to deal with hundreds of students p.a.
Electronic resources

Course web page: http://subjects.ee.unsw.edu.au/tele9301/
• Lecture notes
• Logistics for the mid-session test
• Recommended Reading List

webCT: http://www.webct.unsw.edu.au and
WebCT Vista http://vista.elearning.unsw.edu.au
• Discussion forums
• Viewing marks
• Mid-session test and preview
Resources: Textbook

Textbook:


This text is *recommended*, not required. You *may* find this useful if you want to read a single text that touches on many (but not all) of the topics treated in this course.

1997 publication date: Fundamentals of networks that apply to switched networks have not changed substantially since the date of publication.

The author: S. Keshav
Switching-specific books


All available in the UNSW library.
Finding other material

**Publications** (technical papers):
Free sources: Citeseer, Google
Commercial databases: IEEE Xplore, ACM Digital Library ...

**Standards:**
Internet Engineering Taskforce (IETF) Requests for Comments (RFCs)
ITU recommendations
IEEE standards
...
Links: http://uluru.ee.unsw.edu.au/~tim/misc/standards.html
Terminology and inconsistency

Networking terminology tends to be inconsistent:
• Immature field
• Little formal checking of prose (unlike mathematics)
• Marketing hype
  e.g. router vs. switch; definitions of circuit switching

Course must balance range of terminology (for familiarity) with consistency (to avoid confusion).

Pointers to several glossaries of terms at:
Lecture shorthand

Some abbreviations that you may see in lectures:
standard mathematics: $\Rightarrow$ implies, $\sim \equiv \infty$
$\exists$ there exists, $\forall$ for all
$\uparrow\downarrow$ increases/decreases
$\checkmark \times$ advantages/disadvantages
c.f. compare with
s.t. such that
wrt with respect to
aka: also known as
a la: in the manner of
b bits, B bytes, k 1000, K 1024
Assessment

70% exam

30% mid-session test using webCT Apr. 13 2006, 6-7pm or 7-8pm

Schedule:
Next week: Sample questions online
Week 5: Deadline to email lecturer about scheduling constraints
Week 6: Announcement of allocations of students to slots & labs
Before week 7: Suggest test questions!
Week 7: Test
Week 8: Return of marks & diagnostic feedback information

Bonus marks: Up to 5% more for participation, e.g. for:
- attendance
- asking thought-provoking questions
- giving insightful answers to questions
- suggesting test/exam questions
“The mind is not a vessel to be filled, but a fire to be ignited.” – Plutarch

**Question everything**

**Query individuals:** Lecturers and authors make as many mistakes as students, if not more. The only difference is that the grading isn’t as explicit!

**Query whole fields:** Networking is a developing field; if you don’t understand a concept, then it might be because *it* is fundamentally wrong (not you).

*Don’t just accept what is, think about what should be.*
“uncertainty can be a guiding light”

- Just believing incongruous material is the easy way out
- Try to mark it as incongruous, rather than assimilating it, and find out why. You might be able to contribute a new solution that advances the field.
- **Interact with others.** That’s the advantage of university over self-study using books & web sites.
Advertisement

*Some* projects on network reliability available.
Outline

Motivation for (& definition of) switching
Non-switched networks
   Full mesh
   Broadcast and select
Switched networks
   “Switch” defined
Full mesh networks

• Each terminal directly connects to every other terminal (that it communicates with)

✗ **Uneconomical:** Large number ($N(N-1)$, e.g. 30) of poorly utilized connections

✗ **Unreliable:** Single path between endpoints (unless nodes are willing to forward for others)

✗ **Insecure:** Endpoints control who can access their node. No capacity for partitioning or centrally managed policy.
Broadcast and select networks

- Each terminal connects to a common shared medium. Sources broadcast information.
- Destinations select appropriate information.
- Poor scalability: Shared medium is a bottleneck.
  - As # of nodes ↑, transmission time spent arbitrating access (e.g. Ethernet collisions) also ↑.
- Poor security: Information is visible to all nodes. Endpoint control as per mesh.
- Poor reliability: Single failure point.
- Poor upgrade: Backward compatibility baggage, unless upgrade is universal.
Switched networks

Most traffic is directed (broadcast=bad) and bursty (mesh=bad)

Switches
- Forward traffic only towards its destination(s)
- Multiplex traffic from multiple sources

Advantages:
- Economical for large scale, e.g. 9 connections
- Smaller collision domains; less time spent arbitrating access
- Relatively secure
- Reliable, e.g. choice of path
- Simple to upgrade ⇒ supports heterogeneity

Caveats:
- Switches cost
- Switches may get congested or “block”
- Switches introduce delay
Functional definition of “Switch”

“Switch”: Any device with multiple ports that aims to direct unicast traffic only to one output port that leads to the destination.

Notes:
“functional definition” – not a marketing “definition”
“multiple ports” – multiple input ports alone would be a multiplexer; multiple output ports alone a demultiplexer. Ports are aka interfaces.
Multiple is best thought-of as 3 or more, in which case the switch must decide which output port to send traffic to. A switch with just 2 ports (many home “routers” are just that) is effectively a filter.
“unicast traffic” – multicast traffic may be sent to multiple output ports leading to multiple destinations.
“aims to” – bridges may not be able when they are yet to learn the destination’s location
“one output port” rather than “the output port” – there might be choices; which port is the best is a routing decision.
Relatives of “switches”

A multi-port device that directs input traffic to all ports isn’t a switch. Call it a hub, combiner, etc.

A router is a type of switch that deals with network layer headers. “a type of switch” ⇒ switch functions (fabrics, packet classification, scheduling, buffer management etc) are used in routers.

We’ll consider detailed definitions of types of switches (routers, bridges, etc) next week.
Outline

Evolution of networks
Communication networks of the 1970s

Master-slave relationship between dumb “terminals” and sophisticated mainframe. Terminals often directly connected to mainframe (“star” topology).
Communication networks of the 1980s

Spread of microprocessors raises the sophistication of end-user systems.

Distinction between:
- **clients** (initiate communication, usually low-performance devices) and
- **servers** (respond, usually high-performance devices)

LANs deployed to interconnect clients and servers. Some LAN interconnection.
Communication networks of the 1990s

Heterogeneous terminals
Wireless links and mobility
Communication networks of the 21st C

Merger between terminals and network elements:

• **Content Distribution Networks** locate services within the network, near clients
• **Active networks**: Network elements transform payload as it propagates
• **Peer-to-peer systems** employ computers in the distribution process
• **Mobile ad-hoc networks** employ hosts to forward traffic for others (to concatenate wireless links & save power)
• **Sensor networks** low power sensor devices collect and amalgamate data.
Outline

Network elements
Names of network entities

3 Primary entities:

- **Nodes:**
  - Hosts (aka† end-systems, terminals, computers, stations)
  - Source (aka origin, transmitter)
  - Destination (aka sink, receiver)

- **Intermediate systems**
  - Routers, “switches”‡, bridges, gateways, access points, caches, middleboxes

- **Links:**
  - Concatenated to form paths, routes

- **Networks:**
  - Subnetworks, internetworks, The Internet; areas, domains, regions; clouds

Node/link/network terminology comes from graph theory

† “aka” = “also known as”
‡ as if routers, bridges, access points etc do no switching

Copyright © 2006 Tim Moors
Symbols used by one vendor (Cisco)

Interconnection devices

- Router
- ATM Switch
- Switch
- Processor
- Workgroup Switch
- Bridge
- Gateway
- ISDN Switch
- Multilayer Switch
- Small Hub
- Repeater

Nodes & networks

- Network Cloud
- PC
- Laptop
- File Server

Hybrid interconnection devices

- ATM/FastGb Etherswitch
- Route/Switch Processor
- ATM Tag Switch Router
- ATM Router
Outline

Layering
Layering of communication systems

Decompose complicated communication system into modules.
Stack one module (“layer”) on top of another to refine its service to be more appropriate to the end-user ⇒ “Protocol stacks”

Layers insulate upper layers from changes below.
Layers insulate lower layers from changes above.
Layering theory vs implementation

Layering describes the required *functionality* of the system, but not how the system should be *implemented*.

Layered implementations can be highly inefficient (e.g. buffer double-handling).

In practice: sublayering and overlay networks produce layers ad infinitum
3 essential layers for internetworks

Example protocols

- HTTP / SMTP
- IP / BGP / OSPF
- Ethernet / PPP
# 7 layer theoretical ISO Reference Model

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td><strong>Application</strong></td>
</tr>
<tr>
<td>6</td>
<td><strong>Presentation</strong></td>
</tr>
<tr>
<td>5</td>
<td><strong>Session</strong></td>
</tr>
<tr>
<td>4</td>
<td><strong>Transport</strong></td>
</tr>
<tr>
<td>3</td>
<td><strong>Network</strong></td>
</tr>
<tr>
<td>2</td>
<td><strong>Data link</strong></td>
</tr>
<tr>
<td>1</td>
<td><strong>Physical</strong></td>
</tr>
</tbody>
</table>

*“layer 4 switching” refers to awareness of transport protocol.*
## Network terminology

<table>
<thead>
<tr>
<th></th>
<th>Data unit</th>
<th>Interconnection device</th>
<th>Maximum extent (in the LAN context)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td></td>
<td>Gateway</td>
<td></td>
</tr>
<tr>
<td>Presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>Segment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>Packet</td>
<td>Router</td>
<td>Internetwork or internet</td>
</tr>
<tr>
<td>Data link</td>
<td>Frame</td>
<td>Bridge</td>
<td>Catenet or Bridged LAN</td>
</tr>
<tr>
<td>Physical</td>
<td>Symbol</td>
<td>Repeater</td>
<td>Local Area Network</td>
</tr>
</tbody>
</table>

The term “switch” doesn’t appear anywhere in this table! They can appear in multiple layers. Based on Table 1.2 of R. Seifert: *The Switch Book*, p. 60

Copyright © 2006 Tim Moors
Outline

Switching at various layers
   Lower layer switching
   Higher layer switching
      Transport layer switching
   Application layer switching
Lower Layer Switching

**Physical**: all-optical networks: Wavelength Division Multiplexing, MicroElectroMechanical Systems (MEMS)

**Data link**: bridging

**Network**: routing

\{ \text{Most common layers for switching} \}
Higher-layer (4+) switching

The switches that we’ve considered so far implement all functions of the layers that they use for switching:
- Layer 2 (link): MAC & framing
- Layer 3 (network): routing

Another type of switch (common at higher layers) only implements a subset (possibly null) of the functions of a layer, but is influenced by the information sent by that layer.

i.e. it depends on what protocol is used at that layer, but it doesn’t implement all of the functions of that protocol.

TCP ports identify software processes, and are different from switch ports which are hardware entities.
Transport layer switching

**Strict interpretation:** Transport layer fields affect *direction of propagation*. Switching above network layer processing. Switching between *processes*, e.g. for load balancing on a web server: might construct what clients perceive as a singular “server” by placing a switch between the Internet & a server farm.

might use the source port number to determine which machine receives the request: odd $\rightarrow$ machine 1, even $\rightarrow$ machine 2

(Strictly, you could argue that *end-systems* implement a form of layer 4 switching because they forward segments to the appropriate process, as indicated by their port numbers.)

**Loose interpretation:** Transport layer fields only affect *type of service* through switch. Lower layer fields alone may determine direction. e.g. Network layer switch (IP address $\Rightarrow$ direction) that gives telnet (TCP port 23) priority over FTP data (TCP port 20)
Application layer switching

e.g. consider a web service, handling HTTP GET requests

1. Could use cookies (identifiers included in requests) that identify users to direct them to a specific machine (helps in providing consistent state between consecutive requests)

2. Could direct GET requests for different information to specialised machines (less content each ⇒ higher cache hit rates etc):
   - image requests (file with .JPG extension) to one machine
   - HTTPS to machine with crypto hardware
   - cgi-bin/ to another
   - ...

Figure from W. Mangione-Smith and G. Memik: “Network Processor Technologies Tutorial”
Outline

Types of information in a network
Types of information in a network

**Payload**: Content that end-users want to exchange
(the *load* for which users are willing to *pay* to be carried)

**Control information**: Specifies *how* end-users want to exchange payload

  e.g. *destination* address and specify service requirements (e.g. low delay)

**Overheads**: Introduced by the communication system to provide communication.

  **Examples:**
  - protocol headers, congestion control signals, topology updates for routing
  - *source* address – e.g. for demultiplexing, replies and error messages, charging, flow identification
Encapsulation

Payload from the application tends to be progressively encapsulated as it descends the protocol stack. Reverse on ascent.

Placing overheads around payload ensures contiguous storage, simplifying buffering. Lower layers can’t distinguish between higher layer control and payload.
“Ethernet” frame format

**Type encapsulation** – used for IP, IPX

<table>
<thead>
<tr>
<th>Field</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble</td>
<td>8B</td>
</tr>
<tr>
<td>Destination Address</td>
<td>6B</td>
</tr>
<tr>
<td>Source Address</td>
<td>6B</td>
</tr>
<tr>
<td>Type</td>
<td>2B</td>
</tr>
<tr>
<td>Data</td>
<td>46-1500B</td>
</tr>
<tr>
<td>CRC</td>
<td>4B</td>
</tr>
</tbody>
</table>

Transmission order: ➔
“Ethernet” frame format

**Length encapsulation** – used for Appletalk, NetBIOS

<table>
<thead>
<tr>
<th>Preamble</th>
<th>Destination Address</th>
<th>Source Address</th>
<th>Length</th>
<th>Data</th>
<th>CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>8B</td>
<td>6B</td>
<td>6B</td>
<td>2B</td>
<td>46–1500B</td>
<td>4B</td>
</tr>
</tbody>
</table>

DA in header s.t. can be processed while waiting for remainder.
We’ll consider addressing in detail later, in the context of packet classification.

Type/protocol in header for distribution to appropriate process

CRC in trailer s.t. can be generated as remainder is transmitted

MAC requires minimum Data+Pad of 46B (short frames are “runts”)
Bandwidth vs data rate

Shannon’s capacity theorem shows that the data rate (bits per second) of a channel increases with the channel’s:

- bandwidth (Hz)
- signal to noise ratio

In networking, the term “bandwidth” is often loosely used to refer to the data rate, eg:

- “802.11b has a bandwidth of 11Mb/s”
- Broadband ISDN (Mb/s) vs narrowband ISDN (kb/s)
- Broadband cable vs narrowband dial-up

Risks:

- Data rate can increase independently of true bandwidth (e.g. by raising SNR)
  e.g. 100Mb/s Ethernet provided by both 100Base-T4 (100MHz) and 100baseTX (125MHz) through different coding.
- Bandwidth can be measured instantaneously, whereas data rate/transmission capacity/throughput needs to be measured over some interval.
  (It takes non-zero time to transmit any number of bits.)
  ⇒ Caution in interpreting graphs of bandwidth over time.

“There is a difference in kind between eating an ice cream cone every day, and eating 365 ice cream cones on your birthday” [T., 2nd edition, p. 86]
Outline

User characteristics/requirements
User characteristics/requirements

Spatial aspects
- Directivity
- Spatial distribution

Volume aspects
- Symmetry of data flow

Temporal aspects
- Variability of data flow
- Duration of transfer

Initiation
“Quality of Service”
- Timeliness
- Reliability

Other aspects
- Security, privacy
- Economy, accountability
“Flows”

Flows, conversations (Keshav), streams, sessions, connections, calls, ...
All describe a sequence of data units being sent between common endpoints
Often using common service type, though sometimes may have sub-flows, e.g. lecture consists of flow of PPT slides + flow of audio

Reasons for flows:
• **Segmentation** of large data, e.g. file >> 1500 B limit for Ethernet
• Source **progressively produces** information, e.g. telephone call
• **Repeated interactions** with same node, e.g. with a web server:
  connect, send request, download page, think, repeat, disconnect

How to determine when a flow ends?
• Connection-oriented: Explicit setup & release mark beginning and end.
• Connectionless: Timeout after prolonged inactivity
Directivity of traffic flow

<table>
<thead>
<tr>
<th>Name</th>
<th>From</th>
<th>To</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>unicast</td>
<td>one</td>
<td>one</td>
<td>most</td>
</tr>
<tr>
<td>anycast</td>
<td>one</td>
<td>any</td>
<td>when multiple targets provide identical service, e.g. telephone operators</td>
</tr>
<tr>
<td>multicast</td>
<td>one</td>
<td>many</td>
<td>distribution, e.g. advertising, lectures(!)</td>
</tr>
<tr>
<td>broadcast</td>
<td>one</td>
<td>all</td>
<td>used for multicast apps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>broadcast itself is never really justified</td>
</tr>
<tr>
<td>concat</td>
<td>many</td>
<td>one</td>
<td>gathering information, e.g. monitoring weather sensors; multi-source file tx</td>
</tr>
<tr>
<td>Multi-access</td>
<td>many</td>
<td>many</td>
<td>bidirectional distribution, e.g. videoconference</td>
</tr>
</tbody>
</table>

Importance:
• Multicast allows sources to heavily load switches, affecting performance & favoring shared-media switches.
• Anycast reduces chances of switch blocking
Constructing one type of flow from another

Multicast by iterated unicast
Burdens the source and links
Distributed expansion (left) can increase throughput

Multiaccess from multiple multicast
Requires many identifiers within the network
Spatial distribution

Traditionally, networks have been designed assuming “80/20” or “90/10” “rules”, suggesting that of the traffic originating in a subnetwork:
- 80%/90% is destined to other nodes in that subnetwork
- 20%/10% is destined to nodes in other networks

Justification?: Originated in telephony, but perpetuated through folklore (e.g. analogous to Pareto distribution)

Actual numbers depend greatly on the technology, e.g.
- **Phone network**:
  - Rarely call someone in an adjacent office
  - Frequently call colleague on another floor of building and others in same city
  - Cost & time zones inhibit international calls
- **Web browsing**: Little care for location; origin sources might be uniformly distributed around the world (though may have local caches)

Hard to generalise, but worth considering – some traffic will be local.
Symmetry of data flow

Most flows are *asymmetrical*:

- Short requests to server elicit long replies
- Participants in group interactions are usually silent

Importance:

- At periphery of network, more traffic propagates towards the edges ("downstream") than towards the network centre ("upstream").
  UNSW stats from 2004: campus: 40Mb/s outbound, 60Mb/s inbound
- Acknowledgements often can’t be piggybacked with payload ⇒ large number of short packets

Important exceptions:

- Interactive media – telephony, videoconference, etc
- Modern file sharing – discourage freeloaders; encourage content providers

† related terms: “download”/“upload”; “downlink”/“uplink” (satcom)
Outline

Temporal aspects
Variability of data flow

Network traffic tends to be *bursty*:
- Users download information, pause to think, download more
- Video and other streaming media that can be compressed & degree of compression depends on content.

Strong motivation for switching:
Multiple users sharing a channel can exploit each other’s idleness ⇒ higher end-to-end throughput from same capacity channel.

Burstiness can be smoothed by buffering

More consequences later on “Statistical multiplexing” slide
Variability of different applications


Lower figure(s) from R. O. Onvural: *Asynchronous Transfer Mode Networks: Performance Issues*, Artech House, 1995
Self-similarity of network traffic

Figure from Leland94
UNSW dial-up system load

Multiplexing points affect burstiness
Predictability of data rate

“Constant Bit Rate (CBR): for sources with continuous traffic or requiring guaranteed bandwidth† (uncompressed voice, video)
Variable Bit Rate (VBR): for sources with some traffic variation and ability to tolerate some data loss (compressed voice, video)
Available Bit Rate (ABR): for bursty traffic sources requiring low loss and ability to adapt rate based on feedback
Unspecified Bit Rate (UBR): for bursty traffic requiring just best effort cell delivery with no guarantees”

† bandwidth here = data rate

For elaboration, see M. de Prycker: Asynchronous Transfer Mode: Solution for Broadband ISDN, Prentice Hall, pp. 302-40