An Introduction to Packet Switching for Sync Engineers

4th International Telecommunications Synchronization Forum

November 14th, 2006

Stewart Bryant
stbryant@cisco.com

Note that this is a revised version of the slides on the CD.

Introduction

- As you will discover over the next two days there is an emerging need to deliver frequency, phase and time synchronization over packet networks.
- The purpose of this introduction is to highlight the characteristics of packet networks that are of particular interest to the synchronization engineer that needs to design, deploy or manage a synchronisation services over a packet network.

What are Packets

- Packets are variable length data units (strings of octets) that are used to carry information from a source equipment to a destination equipment.
- Each packet includes the address to which it is to be delivered.
- An equipment may use the following addresses to deliver the packet:
  - –The physical address of the next hop to deliver the packet (A LAN switch)
  - –The network layer address (endpoint logical address) to deliver the packet (An IP Router)
  - –A nickname for the network layer address that changes at each hop (An MPLS Label Switched Router)

Packet Formats – Ethernet

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Ethernet dest addr (0..31)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Ethernet dest addr (32..47)   |  Ethernet src addr (0..15)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Ethernet src addr (16..47)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type code           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Packet Formats – IPv4

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Time to Live |    Protocol   |         Header Checksum       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Address                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Destination Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Note that the IPv4 packet is preceded by the datalink header, for example an Ethernet header

Packet Formats – MPLS

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ Label
|                Label                  | Exp |S|       TTL     | Stack
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ Entry

 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  | Exp |0|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  | Exp |1|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

LSE is preceded by a datalink header (e.g. Ethernet) and followed by either an IP header, or a pseudowire payload

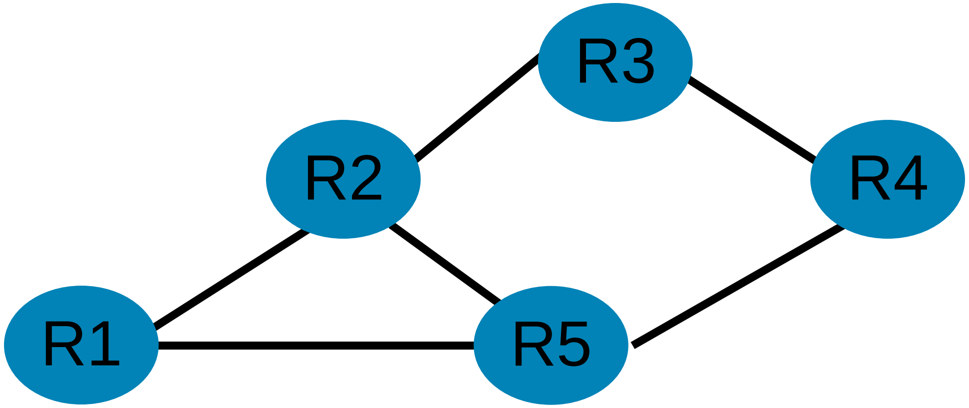
What is a Packet Switched Network

- A PSN consists of a collection of
  - –Endpoints (hosts)
  - –Links
  - –Nodes (Switches or routers or LSRs)
- In most packet switched networks the nodes figure out the best path to a destination automatically as a result of information exchanged with their neighbours.
- In LANs, a single tree is formed.
- In Routed networks, every node has a shortest path tree routed at itself.
- In label switched networks special paths may be installed to engineer the paths used.

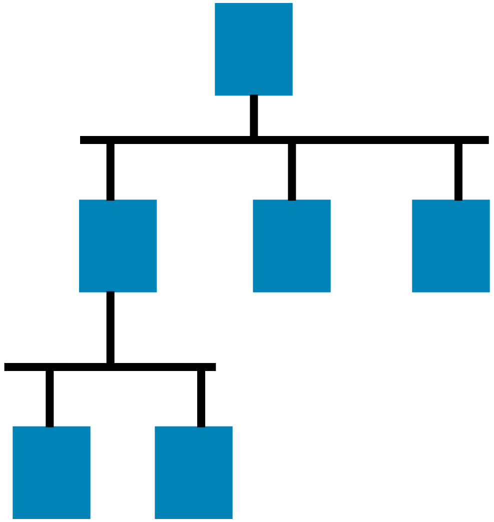
Topology

- The nodes in packet networks normally run a protocol that allows them to determine who their neighbours are.
- They then exchange topology information with their neighbours either:
  - –Their current best estimate of cost to every other node (or one special node), or
  - –Every nodes report of who its neighbours are, or
  - –The path list of the best to each node.
- The nodes then each run an algorithm to deduce the best next hop to any destination node.
- Provided the information exchanged and the algorithm is correct and the information is consistent, the result will be a set of loop-free paths from any node to any other.
- If there is any danger that a packet will loop (even for a short while) the packet header includes a hop-count that causes the packet die if it transits too many hops.

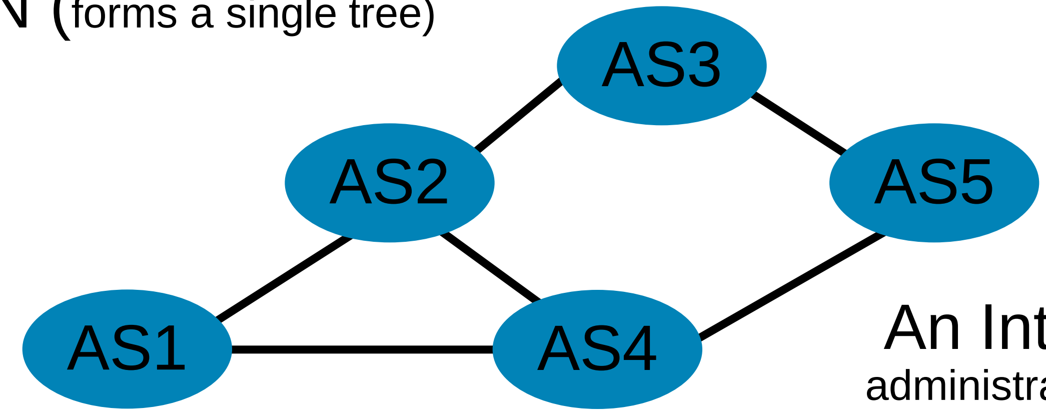
Topology Scope



A Routed network
(Run by a single administrator. All resources available to all nodes)



LAN (forms a single tree)



An Internet (Run by many administrators that apply policy to resource usage and hide information)

Topology Change

- When the topology of the network changes for any reason, the network needs to enter a process of convergence.
- During this time there may be destinations that are unreachable.
- There may also be inconsistency between routers as to the correct next hop. This leads to micro-looping.
- A micro-looping packet may die of old age, or it may get delivered much later than normal.
- Micro-looping introduces congestion that disrupts the delivery of packets along network paths that are unaffected by the topology change.
- When the network has re-converged, some set of paths will have changed their path length with corresponding effect on the time of flight.
- The time for a network to re-converge may range from 200ms to several minutes.

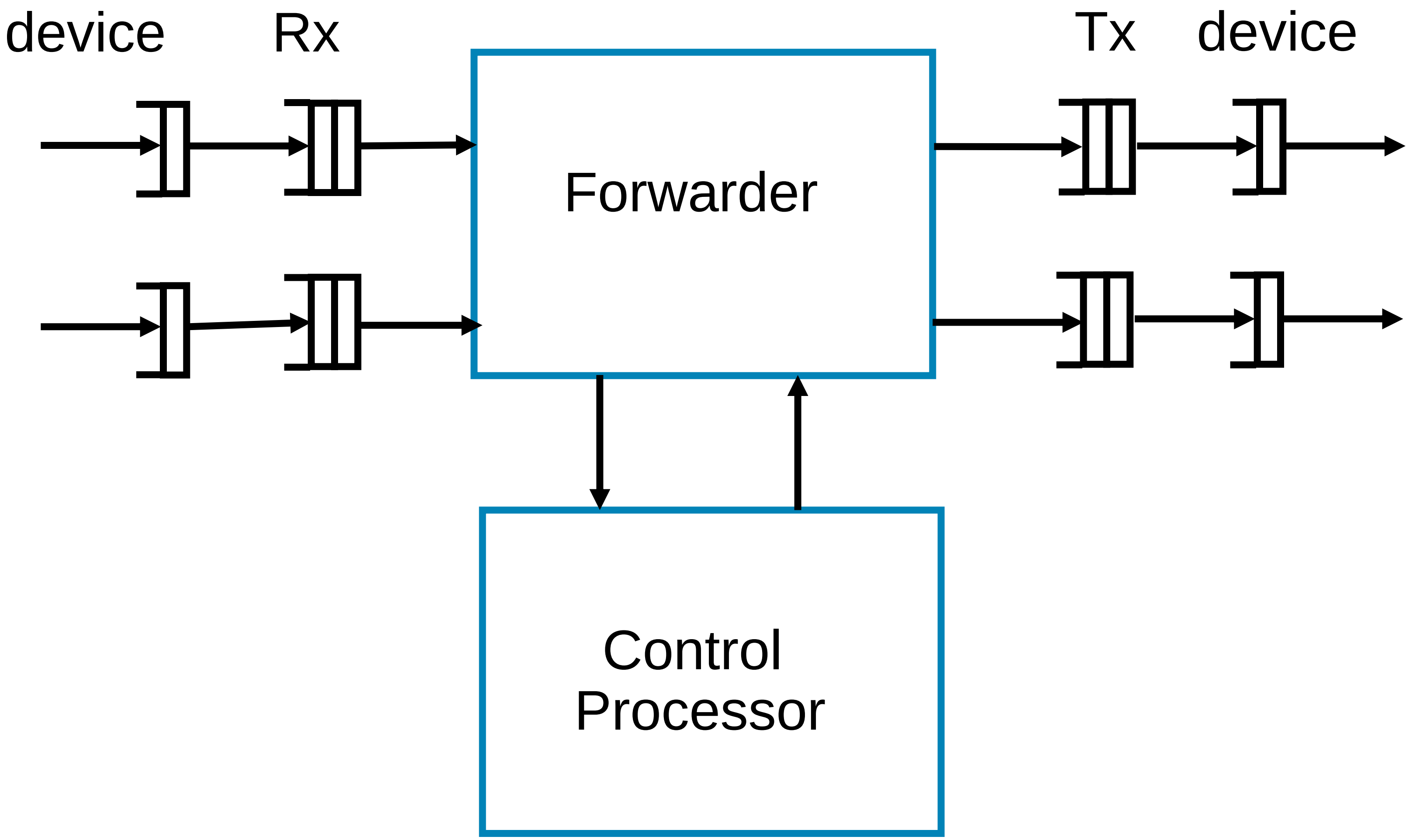
Topology Layering

- Consider two adjacent network layer nodes A and B.
- The path from A to B may also consist of zero or more LAN switches. Due to network layering, neither A nor B will normally be aware of the number of switch hops that separate them.
- Sometimes the network designer may wish to force the path from A to B to go via some node C. To do this they encapsulate packet in another packet addressed to C. This is known as tunnelling.
- A and B may well be unaware that their packets are going via C (and maybe other intermediate nodes).
- Sometimes a network operator may replace a physical link, with a logical emulation of that link (known as a pseudowire). Thus what appears to a pair of nodes as a point to point links may be a whole network.
- This process of layering and virtualization of the network is a very powerful tool that minimises the complexity of the network.
- However this topology hiding has significant consequences for any synchronisation plan, because a virtual link has different delay characteristics to that of a simple physical link.

Jitter

- Packet networks achieve communication economy by applying statistical multiplexing to the use of all resources - links and routers.
- Packets are variable size, self describing data units usually between 64 and 1500 bytes long.
- The characteristic of packet networks that is of most importance to the synchronisation engineer is jitter and wander in the packet delivery time.
- Jitter and wander arise from the packet by packet variation in transit time from a source to destination (i.e. the apparent variation in path length).

A Typical Packet Switch



Queuing Jitter

- A packet is not released from the device until it is complete and its checksum is verified.
- The packet is then placed in a receive queue to wait for the forwarder.
- When the forwarder is ready, it determines the packet's next hop
- The forwarder places the packet in a transmit queue to wait its turn at the output device.
- When the output device is available it takes the next packet from the transmit queue and sends it.
- The purpose of these queues is to smooth out the fluctuations introduced by the statistical multiplexing nature of packet switching, and to make sure that ALL components always have work waiting for them.
- The queues are fundamental to achieving high performance
- These queues are the most fundamental and most significant source of jitter in a packet switch.

Quality of Service

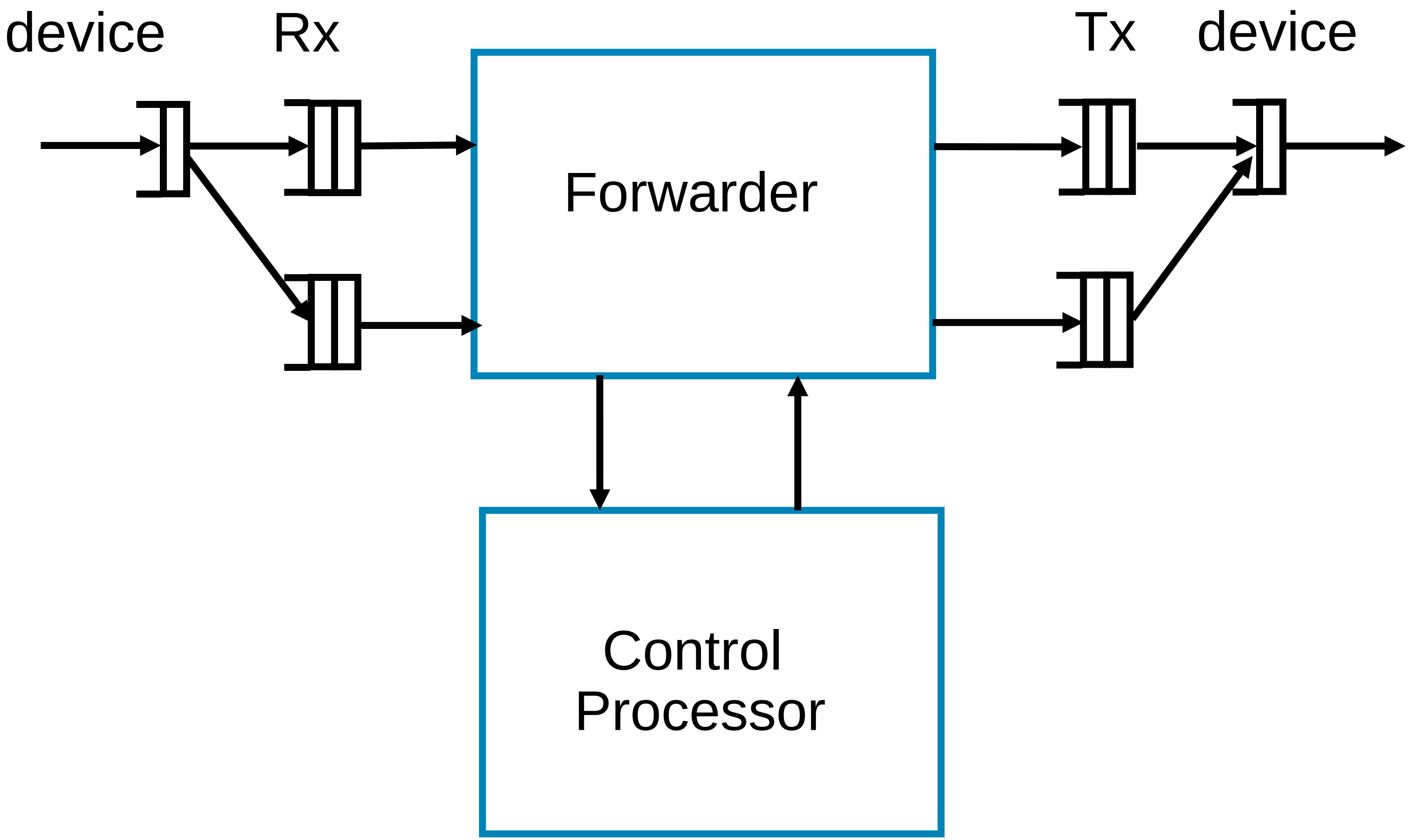
- Quality of service is a way to prioritise “important” packets.

Priority is an attempt to implement the principles of jealousy and envy in computer networks. —Tony Lauck.

Quality of Service -1

- QoS is a very useful tool, but unfortunately from a synchronisation perspective there are problems:
  - –QoS does not create bandwidth, it manages it
  - –The are many candidates for “most important” packet
  - –There is a limit to the number of priorities supportable
  - –Some queues are essential to performance
  - –You can’t predict when a packet will arrive
  - –It’s a bad idea to take back a packet that you have started to send (principle of conservation of work)
  - –If used, QoS needs to be applied at many physical and virtual interfaces that a packet traverses when passing through an switching equipment.

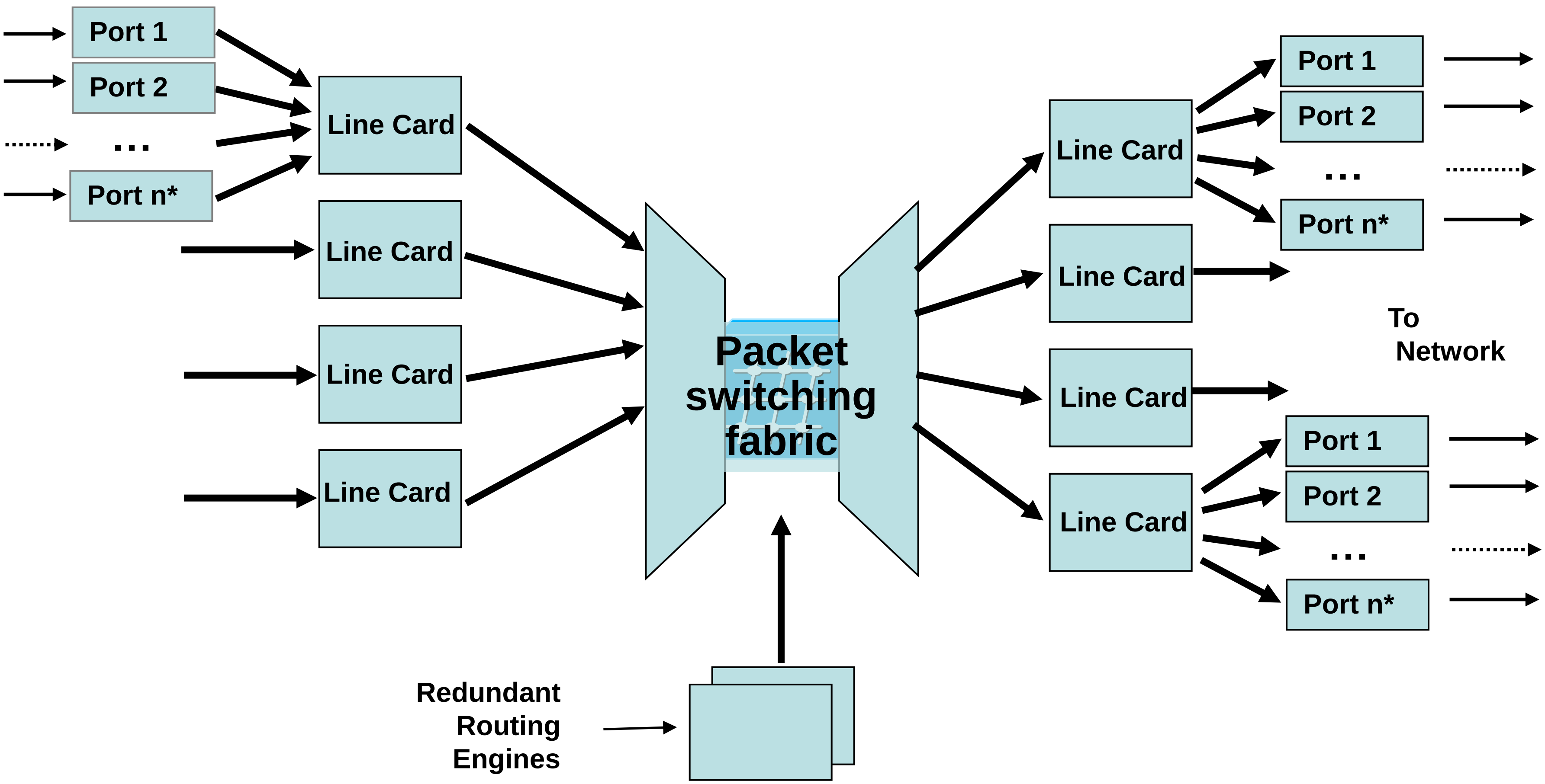
A Typical Packet Switch – with QoS



Forwarding Jitter

- The forwarder is the function that parses the received packet, determines the next hop, and sends the packet to that next hop.
- This may be a centralised function, or a distributed function.
- This may be executed in hardware, in microcode, or in software.
- Hardware forwarders introduce the lowest level of jitter. Hardware forwarding is only deployed where the forwarding features are relatively simple and well characterised. This is why bridges often have hardware forwarders and are perceived as having the lowest forwarding jitter.
- Microcode forwarders are more expensive than hardware forwarders, but are much more flexible. Some of the fastest forwarders are microcoded. There is always some scheduling jitter and so the level of jitter is usually higher than for a hardware forwarder.
- Software forwarders running on a general purpose processor (maybe the management processor) are the most flexible, but introduce by far the most forwarding jitter.
- The forwarder is often shared between a number of interfaces making forwarding delay load dependent.

Typical Multistage Switch/Router



Congestion

- Because packet switched networks are statistical in nature there will always be times when the offered load exceeds the capacity.
- When this occurs the network has no choice but to discard packets.
- Ideally packet discard takes place on the basis of priority, but this is not always possible. It is not therefore possible to guarantee the arrival of any particular packet.
- Fortunately the majority of the traffic uses a congestion sensitive transport protocol (TCP) which dramatically cuts the offered low for its traffic flow – this protects the "network community" at the expense of some (randomly chosen) set of individual flows.
- The majority of time sensitive traffic is not sent over TCP, and whilst this is a minority traffic flow nothing bad happens, but there are concerns about what happens as the proportion of such traffic rises.

Conclusion

- Packet switching is a very powerful flexible technology that provides large economy of scale.
- Minimum packet delay, has been a design consideration since day one.
- Packet jitter was never a consideration in any of the widely deployed packet switching components and hence by definition in the currently deployed packet switching networks.
- Addressing this issue is the focus of interesting work by both the packet switching and synchronisation engineering community – as you will see over the next two days.

Reading

Internetworking Design Basics
http://www.cisco.com/en/US/tech/tk1330/technologies_design_guide_chapter09186a00806666f1.html

Designing Packet Service Internetworks
http://www.cisco.com/en/US/tech/tk1330/technologies_design_guide_chapter09186a00804f83c0.html

LAN Switching
http://www.cisco.com/en/US/tech/tk1330/technologies_design_guide_chapter09186a008066670e.html

Internetworking Technology Handbook
http://www.cisco.com/en/US/tech/tk1330/tsd_technology_support_technical_reference_book09186a00807594e5.html

MPLS in The Internet Protocol Journal - Volume 4, Number 3
http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_4-3/mpls.html

Measuring IP Network Performance in Internet Protocol Journal - Volume 6, Number 1
http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_6-1/measuring_ip.html

High Availability in Routing in The Internet Protocol Journal - Volume 7, Number 1
http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_7-1/high_availability_routing.html

Caveats in Testing Routing Protocol Convergence in The Internet Protocol Journal - Volume 8, Number 4 http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_8-4/testing_routing.html

Further Reading

- **Tutorial on Bridges, Routers, Switches, Oh My!** http://www.ietf.org/proceedings/06jul/slides/rbst-0.pdf
- **Interconnections: Bridges and Routers** by Radia Perlman
- http://www.ietf.org

Q and A