A Brief History of Router Architecture

Tony Li

NLNOG Day 2022

0



Driven by Experience

Disclaimer

- The point of this talk is education, not blame. No names will be given. Everyone was guilty.
- This is from non-ECC protected memory. Mistakes and omissions are all mine.

In the beginning (1969), there was a bus





JUNIPer.

3

... and it was not good

- Every byte passes through the CPU twice.
- Capacity limited by bus bandwidth, memory bandwidth, and CPU speed.
- Infrastructure costs (bus, CPU, memory) are maximized, regardless of bandwidth in use.
- It just doesn't scale.





... but not much

- ALU is a Digital Signal Processor, which gives fast, cheap CPU cycles
- More bandwidth, more cycles, but the fundamental scalability issue is unchanged.







... helps, but we are confused.

- Proof of concept: you CAN forward packets using silicon.
 - Maybe we should build routers using ASICs?
 - Management: No! CPUs will save us!
 - Time lost while we do the wrong thing.
- Improving the lookup helps but moves the bottleneck to bandwidth.
- To be scalable, processing **and** bandwidth must be distributed.

The Web changes everything

- Carriers buy up the NSFnet regionals.
- Real money becomes available.
- Competition blossoms.
- ASICs become credible.





Progress, but...

- This gives us distributed processing and bandwidth.
- Scheduling is hard. Contention for output ports means that switch lanes need more bandwidth than an output (speedup).
- You can't always drain inputs, so you end up with Head Of Line Blocking.
- You'd like a queue per output on each input.
- It still doesn't scale well enough.



Distributed cell memory and a full mesh? (1999) NIC NIC NIC NIC NIC NIC JUNIPER

12

Yes, but...

- Distributed processing and bandwidth without scheduling.
- All NICs are dependent on all other NICs.
 - Hotswap of one card drops all traffic.
- Bandwidth of the system is a fixed multiple of the bandwidth of the card.
- Still not scalable enough.



NLNOG starts (2002)



What about a Torus? (2003)



Juniper Business Use Only

Donuts make lousy routers

- Traffic patterns can cause congestion in the fabric.
- Links are finite.
- Bandwidth needs to be distributed and **uniform**.







Definitely not

- Can't be a single bank of memory
- Many banks, many memory controllers
- Now you need an interconnect for memory controllers
- Still doesn't scale







JUNIPEG. 20

Finally, something that scales

- Bandwidth is distributed, uniform, and redundant
- Scalability is still a challenge:
 - Still need a queue per output
 - Cell addressing is finite: fabric can only be so big
 - Cell fabric is proprietary
 - Technology upgrades are difficult to integrate
 - Change the fabric? Forklift upgrade.
 - Vendor lock-in and chipset lock-in are issues
- Thank you Charles Clos! Work from 1952 takes root.







The future...

- Elastic scaling
 - Still a Clos topology
 - Scalable in two dimensions: add more spines or add another pair of stages
 - Easy incremental upgrades
- Fully vendor independent
 - Standardized links (Ethernet)
- Packet based, not cell based. Needs a little more bandwidth.
- Missing today: management plane abstraction









How is that different?

- Identical architecture, now with white box hardware
- Same underlying chipset
 - Chipset lock-in
 - Software lock-in



Questions?

THANK YOU



Driven by Experience

...........