Data Driven Connectivity

Junda Liu, Aurojit Panda, Ankit Singla, Brighten Godfrey, Michael Schapira, Scott Shenker
Division of Concerns
Division of Concerns

- Routing is a **control plane** operation.
- Operates in the order of milliseconds.
Division of Concerns

- Routing is a **control plane** operation.
- Operates in the order of milliseconds.
- Packet forwarding is a **data plane** operation.
- Operates in the order of microseconds.
Link Failures Hard
Link Failures Hard

- Some users require low latency packet delivery.
Some users require low latency packet delivery.

Some users require high reliability.
Link Failures Hard

- Some users require low latency packet delivery.
- Some users require high reliability.
- Control Plane response to link failure is too slow.
Today’s Solution

- Rely on precomputed backup paths
Today’s Solution

- Rely on precomputed backup paths
- Typically support single link failures.
Today’s Solution

- Rely on precomputed backup paths
- Typically support single link failures.
- State grows exponentially for more links.
Today’s Solution

- Rely on precomputed backup paths
- Typically support single link failures
- State grows exponentially for more links
- Hard to generalize. Hard to configure
Routing is the Problem!

• Routing conflates two functions
  • Optimality - Use good paths
    • Inherently global, requires coordination.
  • Connectivity - Deliver packets
    • Can it be local?
Data Plane Connectivity
Data Plane Connectivity

- Can we push connectivity to the data plane?
Data Plane Connectivity

- Can we push connectivity to the data plane?
- What would it take?
Data Plane Connectivity

- Can we push connectivity to the data plane?
- What would it take?
- No FIB changes at packet rate.
Data Plane Connectivity

• Can we push connectivity to the data plane?
• What would it take?
  • No FIB changes at packet rate.
  • No additional data in packet header.
Data Plane Connectivity

- Can we push connectivity to the data plane?
- What would it take?
  - No FIB changes at packet rate.
  - No additional data in packet header.
- Impossible
Data Plane Connectivity

• Can we push connectivity to the data plane?
• What would it take?
  • No FIB changes at packet rate.
  • No additional data in packet header.
  • Impossible
Data Plane Connectivity

• Relax constraints
  • Change a few bits in FIB at packet rates.
• Clearly feasible, but is it enough?
Guaranteeing Connectivity

1. Take advantage of available redundancy.
Guaranteeing Connectivity

1. Take advantage of available redundancy.
2. Restore connectivity at data speeds.
Guaranteeing Connectivity

1. Take advantage of available redundancy.
2. Restore connectivity at data speeds.
3. Achieve optimality at control speeds.
Using Redundancy: DAGs

Destination
Using Redundancy: DAGs

• Current paths to a destination do not use all links
Using Redundancy: DAGs

- Current paths to a destination do not use all links
- Extend routing tables to increase redundancy.
Restoring Connectivity
Reverse to Reconnect
Reverse to Reconnect

- Link failure can disconnect a DAG.
Reverse to Reconnect

- Link failure can disconnect a DAG.
-Disconnected node reverses all links to point out.
Reverse to Reconnect

- Link failure can disconnect a DAG.
- Disconnected node reverses all links to point out.
- Finite set of reversals reconnect DAG.
Reversals in Data Plane

- Two challenges must be addressed
Reversals in Data Plane

• Two challenges must be addressed
• Notifications can be lost.
Reversals in Data Plane

• Two challenges must be addressed
  • Notifications can be lost.
  • Notifications can be delayed.
Walk Through
Walk Through
Walk Through
Create an OUT Link
Create an OUT Link

Local Sequence
Create an OUT Link

Local Sequence
Remote Sequence
Create an OUT Link

Local Sequence
Remote Sequence
→ Reversible
Create an OUT Link

- Reverse link direction

Local Sequence
Remote Sequence
→ Reversible
Create an OUT Link

- Reverse link direction
- Increment Local Sequence

Local Sequence
Remote Sequence
→ Reversible
Create an OUT Link

- Reverse link direction
- Increment Local Sequence
- Forward packet

Local Sequence
Remote Sequence
→ Reversible
Dealing with Notifications

Local Sequence
Remote Sequence
→ Reversible
Dealing with Notifications

- Receive on link pointing OUT

Local Sequence
Remote Sequence
→ Reversible
Dealing with Notifications

• Receive on link pointing OUT
• Compare sequence numbers

Local Sequence
Remote Sequence
→ Reversible
Dealing with Notifications

- Receive on link pointing OUT
- Compare sequence numbers
- See if anything changed

Local Sequence
Remote Sequence
Reversible
Dealing with Notifications

- Receive on link pointing OUT
- Compare sequence numbers
- See if anything changed
- Reverse link

Local Sequence
Remote Sequence
→ Reversible

Remote Sequence
Local Sequence
Zooming Out
Zooming Out
Zooming Out
What about Optimality?
Safe Control Plane

- Cannot interfere with data plane.
Safe Control Plane

- Cannot interfere with data plane.
- Build a safe primitive
Safe Control Plane

- Cannot interfere with data plane.
- Build a safe primitive
- Set all edges of a node to point out
Safe Control Plane

- Cannot interfere with data plane.
- Build a safe primitive
- Set all edges of a node to point out
- Described in paper
Evaluation
Evaluation Overview

- Test on WAN and datacenter topologies
- Stretch, Throughput, Latency
- Effect of FIB update delays
- On latency and throughput
- End-to-end benefits of using DDC.
Evaluation Overview

- Test on WAN and datacenter topologies
- Stretch, Throughput, Latency
- Effect of FIB update delays
- On latency and throughput
- End-to-end benefits of using DDC.
End-to-End Test

- 8 Pod FatTree
- Partition aggregate workload
- 5 link failures
- Simulated effect for 550 seconds
Requests Fulfilled

- Bucketed 10 second intervals.
- Percentage requests satisfied.
Request Latency

Cumulative Probability

Time (ms)
FIB Update Delay

- What is the impact of delayed FIB changes?
- On packet latency?
- Three link failure: all traffic in test affected.
- Focus on behavior before convergence.
FIB Update Delay

Overall ~99% of packets in under 3 ms.
No packets get dropped, just long tail.
FIB Update Delay

• What is the impact of delayed FIB changes
  • On TCP throughput?
  • Use a WAN topology (AS 2914)
  • 1 Gbps links
• 5 link failures
FIB Update Delay

Cumulative Probability

Throughput (Gbps)

Delay = 0.0 ms
Delay = 1.0 ms
In the Same Vein...

- FCP (SIGCOMM ’07)
- Unbounded bits in header
- Extensive FIB changes on failure packet
- Packet Re-Cycling (HotNets ’10)
- First solve an NP-Complete problem.
- log(network diameter) bits in header.
- DDC is simpler.
Potential Impact

- ASICs implement DDC
- Connectivity guaranteed by the data plane.
- Control Plane focuses on optimality/functionality.
Questions?