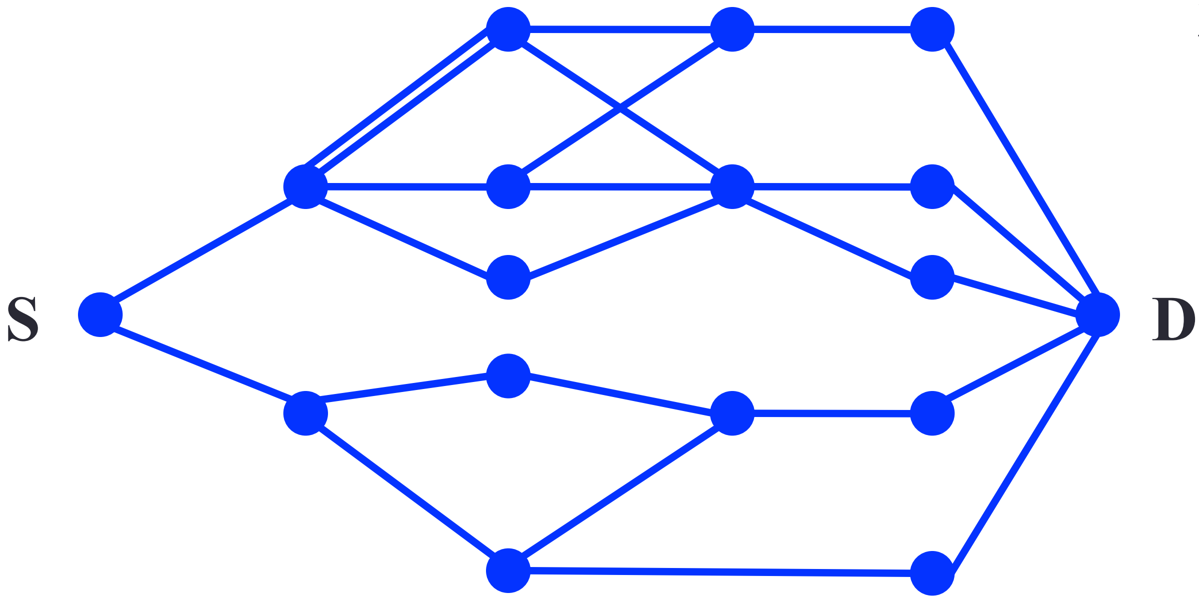
# MULTI-PATH RSVP-TE

KIREETI KOMPELLA
CONTRAIL SYSTEMS

# The Choice Today

14 ECMP paths from S to D



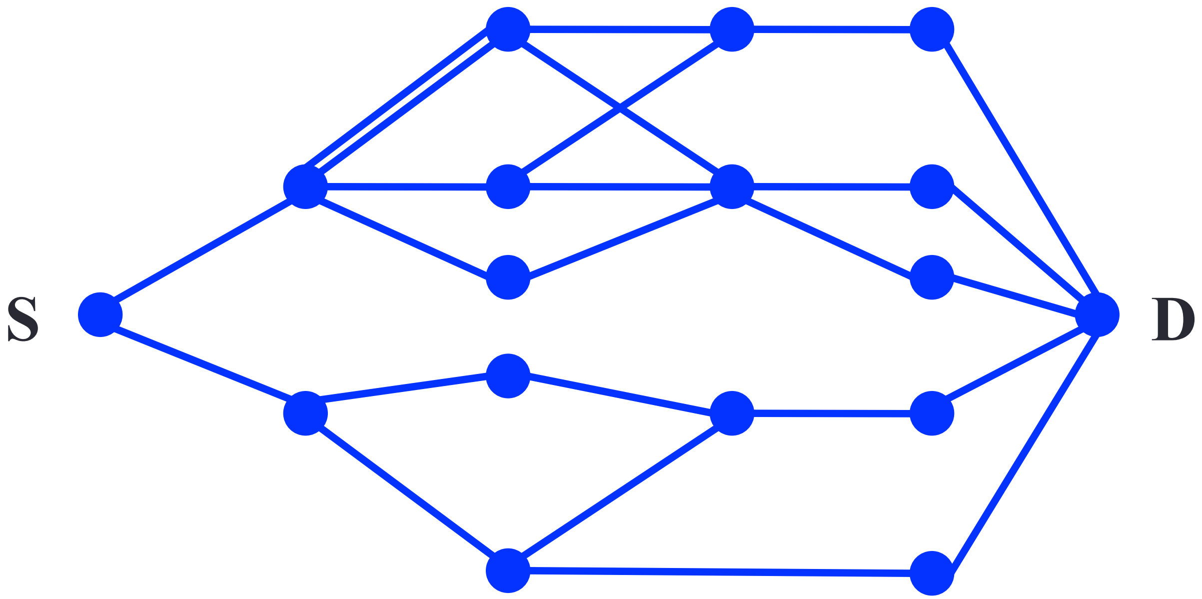

1. ECMP across all equal-cost paths using LDP
2. **OR** individual constrained TE paths using RSVP-TE

# Why Not Both ECMP and TE?

- Is there a fundamental reason that
  - **destination-based routing** (IP, LDP) is amenable to ECMP
- while
  - **traffic-engineered source routing** (ATM, RSVP-TE) is not?

# Can’t We Already Do ECMP With TE?

**Compute an LSP along each possible ECM-Path;**
**configure a TE LSP along each one;**
**tell S to load balance among these paths**




**Won’t that do the job?**

# Can’t We Already Do ECMP With TE?

- What that does is ECMP *followed by* TE. That is, we make a load balancing decision only at the ingress, S
- Thereafter, we simply make source routing decisions
  - That’s a fine approach, and may be sufficient in many cases.
- However, we still haven’t answered the question whether source routing and ECMP are fundamentally incompatible
- Furthermore, we haven’t explored the benefits of *true* TE with ECMP, where at each hop, both a load balancing decision and a source routing decision are made

# Reasons for Multiple RSVP-TE LSPS

- ECMP
- Bin packing of a single large bandwidth LSP
- Better FRR

- Having a single logical construct that takes in the requirements

  <ingress, egress, constraints, bandwidth>

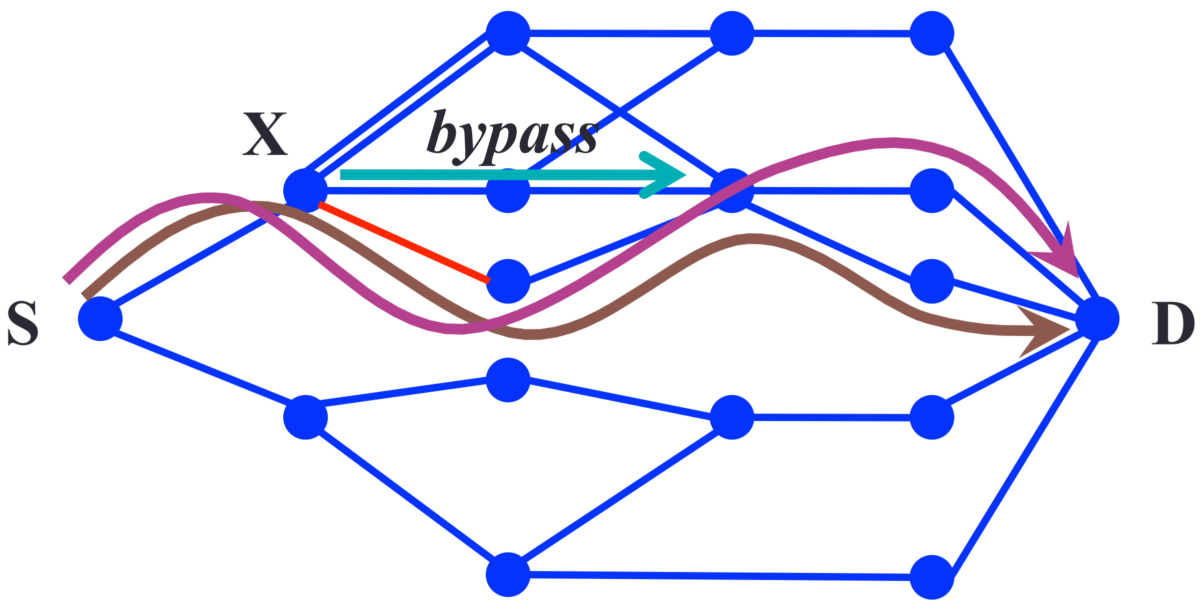
- and deals with the details of splitting up, placement and traffic allocation is interesting to many service providers

# Problems with Multiple RSVP-TE LSPs

1. You have to configure independent LSPs along all the ECMP paths (N configs instead of 1)
2. As these are independent LSPs, some paths may overlap; other paths may be missed
   - Even if laid out carefully, when the topology changes, the LSPs will get out of whack
   - Thus, path computation **must** be aware that ECMP is the goal
3. Assigning bandwidth to each LSP is non-trivial
   - If an independent LSP fails, its bandwidth is gone
   - If a path of an ECMP LSP fails, its bandwidth can be redistributed among the remaining paths
4. FRR across multiple RSVP LSPs is *highly* sub-optimal

# FRR With N Independent LSPs

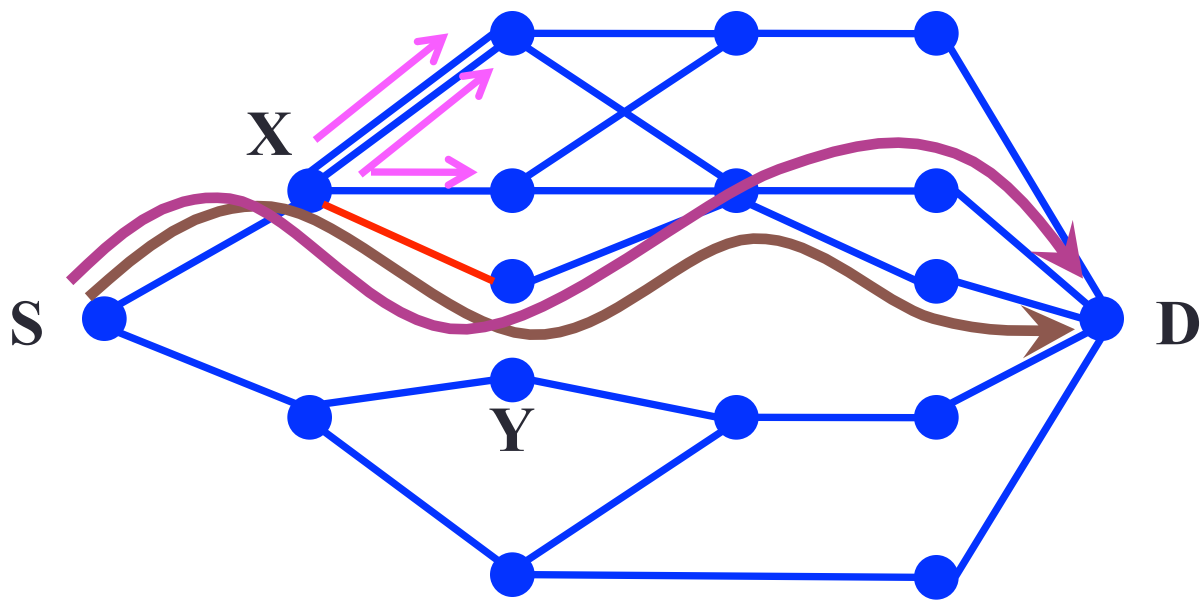
**There are 11 LSPs going through X, two through the red link. For recovery after the red link fails, those two LSPs have to be resignaled. Their new paths will overlap with existing LSPs.**




**If these two LSPs are not resignaled, their bandwidth is "lost": the overall provisioned bandwidth from S to D decreases**

# FRR With a Multi-path LSP

**Since this is ECMP, you don't need bypass or detour LSPs where paths split – each node just rebalances traffic**




**When the red link fails, X rebalances traffic from the two sub-LSPs over the remaining nine sub-LSPs**

**However, at Y, need "regular" FRR**

# Steps in ECMP RSVP-TE

1. Configure a *single multi-path entity* from **S** to **D** with constraints rather than N independent LSPs
2. Compute the paths of the constituent LSPs of the multi-path LSP in an ECMP-aware fashion
   - Split the aggregate *bandwidth* among the constituents
3. Signal the constituent LSPs as part of a single multi-path LSP
4. If a failure occurs, recover constituent LSPs knowing that they are part of a single multi-path LSP
5. If topology changes, go back to Step 2

# Signaling ECMP LSPs

Key question: how do transit nodes know that a set of LSPs are part of a multi-path construct?

1. Re-use the sub-LSP concept from P2MP RSVP-TE
   - with a new session object for “ECMP LSPs”
   - Current approach in draft
2. OR use “associated signaling” to associate LSPs
   - This may be easier for incremental deployment

# Bandwidth Management

- Given a single ECMP-TE LSP of some given bandwidth, how should one partition the bandwidth among the sub-LSPs?
- Many choices:
  - Partition the bandwidth equally among sub-LSPs
  - Partition incoming bandwidth equally among outgoing links (commonly used for IP and LDP LSPs)
  - Same as above, but proportional to link speeds
  - Weighted ECMP
  - Other ideas?

# Next Steps

- Update draft with signaling choices
- Add bandwidth management choices
- Add more details about FRR
- (Maybe) talk about state management