OSPF Traffic Engineering (TE) Express Path

draft-giacalone-ospf-te-express-path-00.txt

Spencer Giacalone, Alia Atlas, John Drake, Dave Ward

Agenda

- Introduction to OSPF TE Express Path
  - Background
  - Problem
  - Protocol overview
  - Encodings
  - Next steps
  - Questions

The Scenario

- Financial networks have changed:
  - Orientation towards machine (“algo”) trading
    - Arbitrage
    - Real time data: Low latency (LL) and ultra low latency (ULL)
  - Milliseconds and *(increasingly)* microseconds count
- High rate flows
- Not able to gap (drop) packets
- Out of SLA is out of service!

Problem

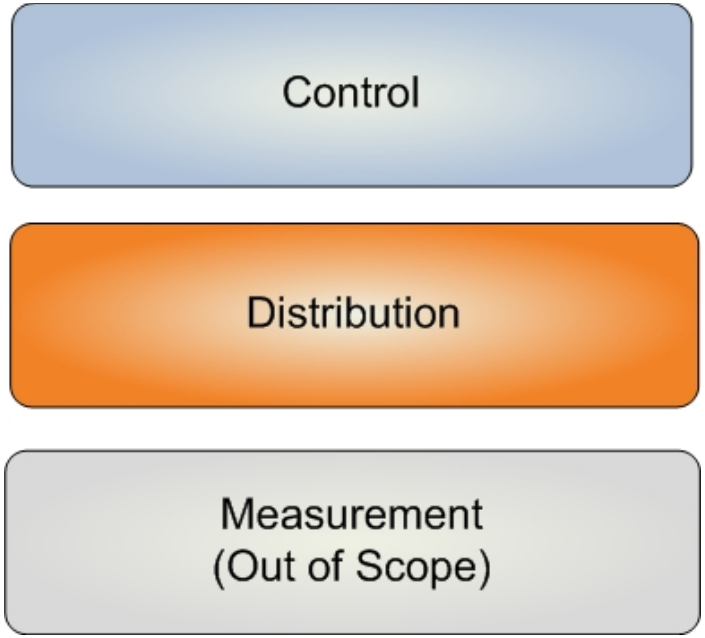
- We need to *guarantee* delivery of *large quantities* of data with the *lowest latency*- not lowest cost, etc
  - *In certain richly interconnected networks, interface cost is becoming generally irrelevant. Performance Is King. This is a real and current need.*
- We have high redundancy and bandwidth, but managing performance flows is difficult:
  - Overall path lengths vary
  - We act as a service provider, but are not one in the classical sense
    - Dependant on underlying transport services
    - Segments not always dark fiber
    - Full transport service “stack” not visible

Static Costing Is A Problem

- Difficult to capture latency, loss, and bandwidth in single static metric
- Performance changes- latency, loss, etc
  - Path protection
  - Flaps, drops, etc
  - I need to know the *current* values at LSP compute time
- Complicated, error prone, time consuming

Overview

- OSPF TE Express Path simplifies these issues
- Automatically distributes performance data
  - Allows control plane manipulation
    - To permit MPLS tunnel setup, failover, fail back
    - Based on network performance
  - Intentionally independent from measurement protocols
    - E.g. MPLS TP, PTP, etc
  - Also, intentionally independent from “applications”
    - Routing/MPLS enhancements
    - Weighted ECMP
    - Others
- Modular and extensible



What About Stability??

- Aimed at MPLS TE
- Averaged values
- Controlled announcement
- Does not define how control plane reacts- just distributes data
- Not having a déjà vu

Protocol Architecture

- Extends RFC 3630
- Two Main Types of Sub-TLV
  - Nominal (Routine) Sub-TLVs
    - Steady state path selection according to performance
    - Initial tunnel build
    - Fail over path selection and monitoring (Non SLA compliant best path may not be desirable for fail over use)
    - Possibly also general monitoring
      - Alternative method- topology database
        - » Link by link or path
  - Anomalous (Significant) Sub-TLVs
    - Can trigger re-computation when performance SLAs are violated
    - Fail back
- Different announcement scheduling and averaging periods
- Individually configured
- Intentionally kept separate to simplify implementations

Sub-TLVs

- Five New Sub-TLVs (Currently)
  1. Nominal Unidirectional Link Delay
  2. Nominal Unidirectional Delay Variation
  3. Nominal Unidirectional Link Loss
  4. Anomalous Unidirectional Link Delay
  5. Anomalous Unidirectional Link Loss
- Next version will include additional sub-TLVs for items like residual bandwidth

Encoding

- Types: TBD
- Length: 4 Bytes
- Values:
  - Latency or jitter as rolling average, to remote peer, floating point
  - Loss as packet percentage
  - Sent when threshold exceeded
    - Different thresholds for different sub-TLVs
    - Configurable
    - And when reuse threshold passed (Anomalous sub-TLVs only at this time)

Simple MPLS TE Example

- Nominal TLVs used to calculate CSPF- Initial state for path selection
- Upon SLA violation, Anomalous sub-TLV prompts CSPF
  - And fail over to secondary
- CSPF uses Nominal sub-TLVs ensure secondary path is conformant

Next Steps- Short/Mid Term

TLVs aimed at weighted ECMP

- Residual bandwidth
- Available bandwidth

– Modifications based on feedback

WG adoption

for cross pollination

- Modeling
- Interworking/requirements with MPLS TP Loss/Delay

ECMP

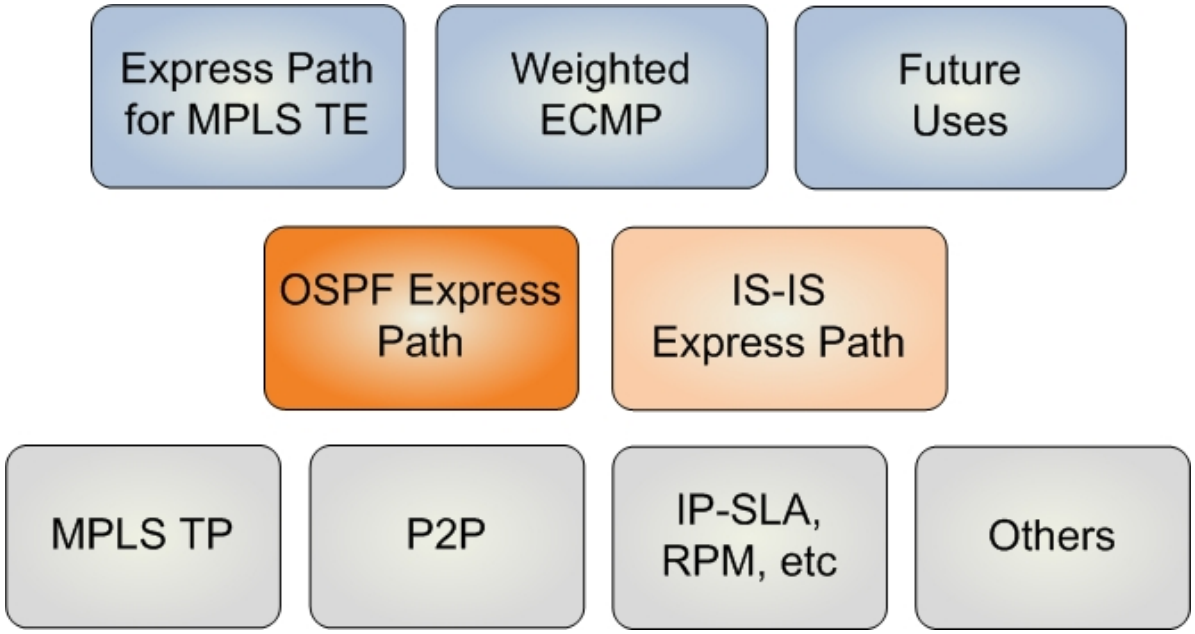
- Residual bandwidth
- Available bandwidth

– Modifications based on feedback

- Modeling
- Interworking/requirements with MPLS TP Loss/Delay

Next Steps- Longer Term

Longer term plans include IS-IS TE Express Path, and drafts related to “applications” such as MPLS TE control plan Express Path, Weighted ECMP, and possibly others



Questions