An IETF Traffic Engineering Overview

Haomian Zheng zhenghaomian@huawei.com
Adrian Farrel adrian@olddog.co.uk

IETF-103 : Bangkok : November 2018

Who Are We?

- Adrian
  - Served six years as Routing Area Director to March 2015
  - Co-chaired WGs in Routing, Operations, and Security
  - Technical Advisor to the TEAS WG
  - Just appointed as Independent Stream Editor

- Haomian
  - Active in all the right working groups
    - TEAS
    - PCE
    - CCAMP
  - Specialist in GMPLS and management for optical network

Warning – Information Overload

- There is a lot to say about TE in the IETF
- We only have one hour

- It's going to be fast and furious!
- Just a taster and lots of pointers
- If we leave out your favourite protocol, please forgive us

Menu

- What is Traffic Engineering?
  - Why do it?
  - What do we need from our protocols?
- Lots of work already done
  - An overview of IETF TE techniques
- What's new and up-and-coming?
  - What new TE tools and techniques is the IETF working on?
- References
  - We can't mention every RFC and draft
  - Just flag up a few key RFCs

What is the Point?

- Traffic Engineering (TE) is concerned with performance optimization of operational networks
- The application of technology and scientific principles to the measurement, modelling, characterization, and control of Internet traffic, and the application of such knowledge and techniques to achieve specific performance objectives

--- So says RFC 2702

- The purpose is to allow a network operator better control of their network to:
  - Provide more reliable traffic delivery
  - Offer advanced services
  - Make better use of network resources
  - Survive outages and planned maintenance
  - Make the network predictable

What Does That Really Mean?

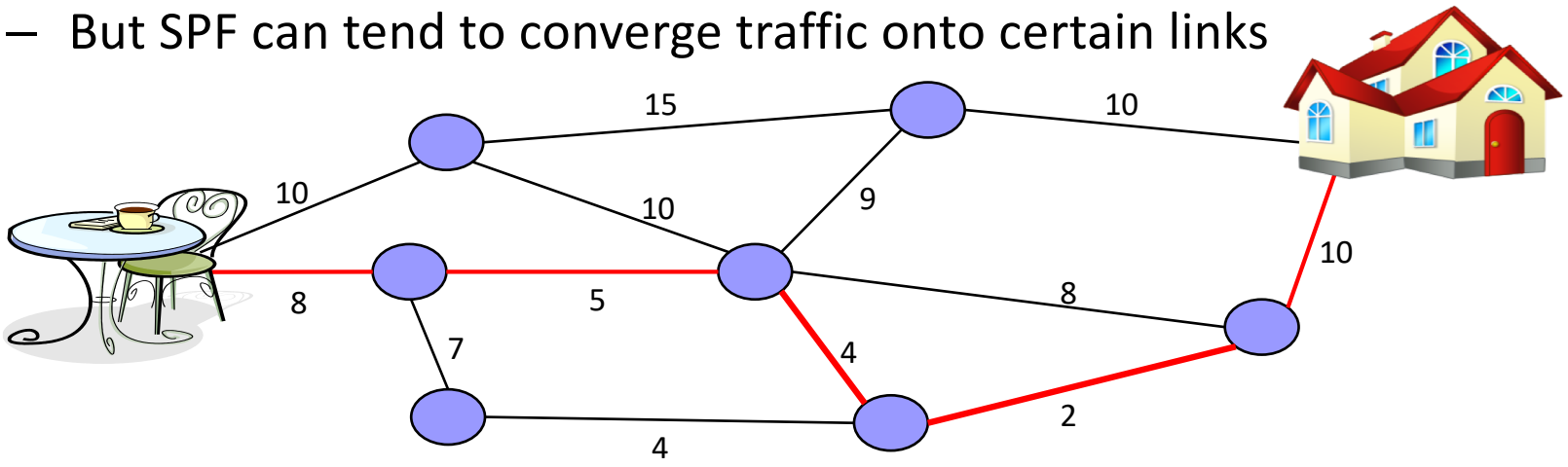
- First you have to know your network
  - Topology: nodes and links and connectivity
  - Capabilities: bandwidth, delay, metrics
  - Traffic: patters, demands, matrices
  - Errors and plans
- Then you have to control how the traffic flows
  - To do that you have to configure
    - The network
    - The traffic
    - The management tools

What Sort of Networks?

- Traffic Engineering applies to any data delivery network
  - Actually, applies to any commodity delivery network
    - Electricity, cars, water, sewage, etc.
- We focus on “layer 3 and below”
  - Packets/frames : IP, MPLS, Ethernet
  - Transport technologies : TDM, OTN, lambda, port

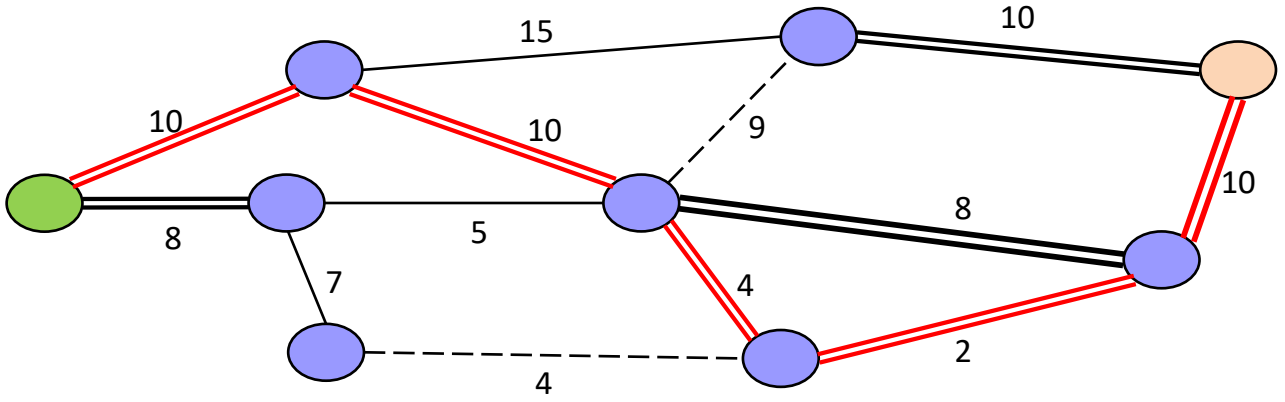
Shortest Path First

- In 1956 Edsger Dijkstra wanted to find the shortest way home from the coffee house
  - Least hops quickly leads to per-hop metrics
  - Dijkstra's algorithm is embedded in OSPF and IS-IS so that all nodes in the network make the same forwarding assumptions
  - SPF is quick to compute even in very large networks
  - But SPF can tend to converge traffic onto certain links

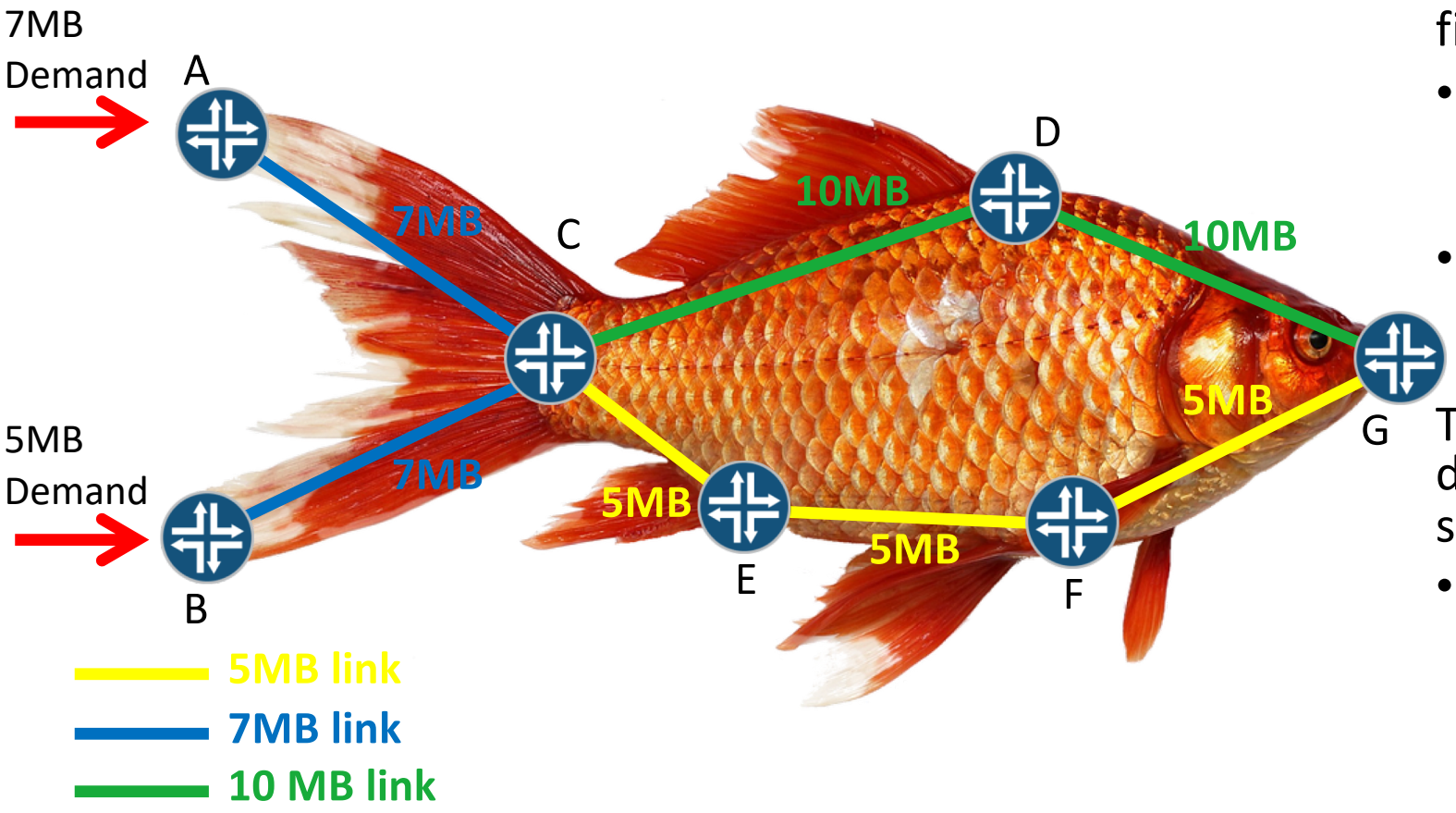


Constraint-based Shortest Path First (CSPF)

- Apply additional constraints to the SPF calculation
  - Constraints may be per-hop (for example, bandwidth or lambda continuity)
    - Processing is simple pruning of the graph before or during SPF
  - Constraints may be cumulative (for example, delay)
    - Processing is just like SPF with multiple counters
  - CSPF is quick to compute even on complex networks with multiple constraints
  - But can still tend to converge traffic on some links



It's All a Bit Fishy



Suppose a 5MB demand arrives first

- It gets placed on the shortest path
- Then the 7MB can't be placed

TE allows the 5MB demand to be steered to BCEFG

- Allows the network to support both demands

A Global View for An Optimal Network

- Solving the fish problem requires knowledge of:
  - The network topology
  - The status of all links
    - Up/down/wait-for-down
    - Available bandwidth
    - Other constraints
    - Share Risk Link Groups (SRLGs)
  - All demands
  - Any relationships between demands
- Adaptive TE places set of demands and tweaks the network

High Level View of the Toolkit

- In an abstract sense, we need:
  - Network management
    - Sees whole network and all/new demands
    - Computes how to place demands / modify network
    - Issues commands
  - Routing/discovery
  - Device configuration / signalling
  - Packet / flow marking
- You don't always use all the tools in any deployment

IETF Tools

- IP Source Routing
- IGP Metric Tweaking
- Coloured Graphs
- RSVP
- MPLS-TE
- GMPLS
- PCE
- BGP-LS and Network Aggregation

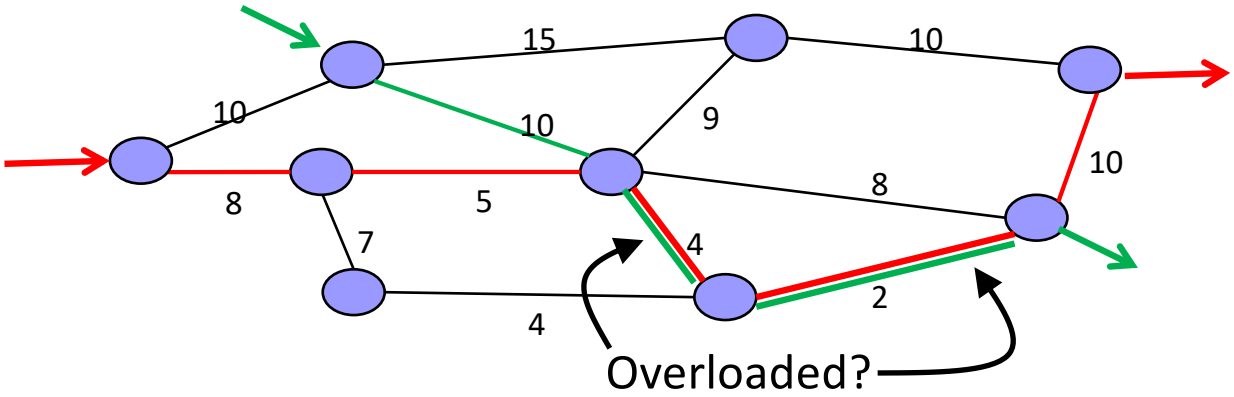
IP Source Routing

- Encode path information in the packet
  - Strict path
    - Packet header enumerates every node in the path
    - No path information stored in the network
    - E.g., IPv4 with Strict Source Routing Option
  - Loose path
    - Path is divided into segments
      - Segment contains one or more router hops
    - Packet header enumerates each segment that the packet traverses
      - But it does not necessarily enumerate every node
    - Network contains enough state to forward the packet through multi-node segments
      - Normal SPF routing
    - Examples
      - IPv4 Loose Source Routing Option
      - IPv6 Routing Extension Header

RFC 791

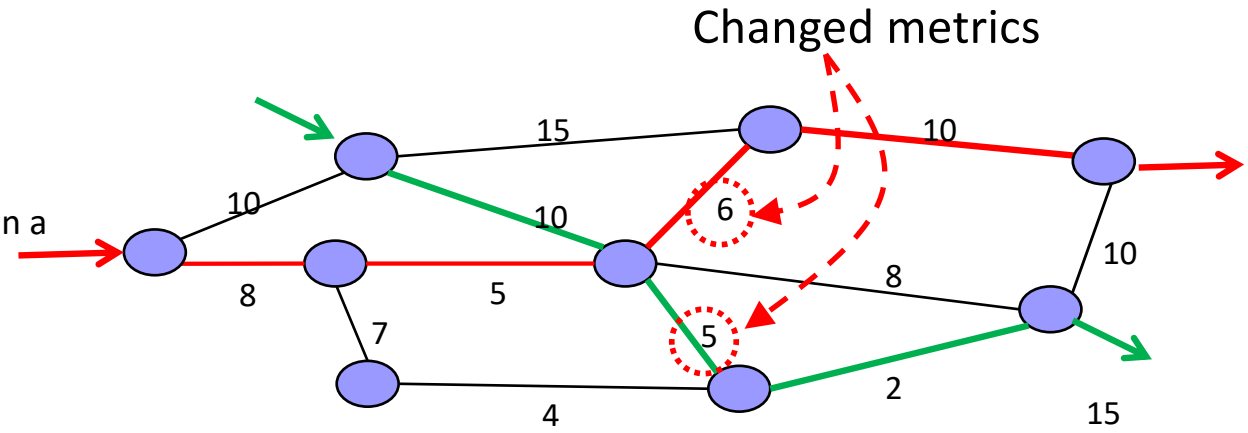
RFC 8200

Tweak Those Metrics



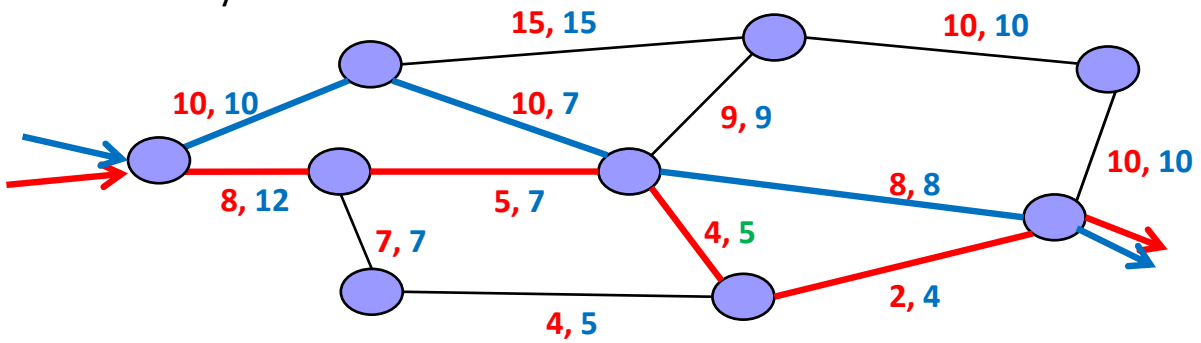
- Consider the SPF graph from before
- Add a second flow
- Two links may be overloaded

- Careful changing of metrics redistributes traffic
- Beware!
  - With many flows it is very complex
  - Unexpected consequences
  - Don't flap the network
  - Consider what happens with changes in a live network
- Use off-line central tool
- It is successfully deployed



Coloured Graphs

- IGP allows different metrics to be assigned to different code points on the same link
- Builds separate (coloured) topologies on the same physical underlay
- Entry point directs traffic to a topology
  - 5-tuple hash or policy configuration, etc.
  - Colour packets using DSCP
  - Use SPF on the coloured topology identified by the DSCP
- Main use cases
  - Resilience
    - Two diverse paths from source to destination, on in each graph
  - High priority versus best effort
    - One graph is deliberately under-used



Resource ReSerVation Protocol (RSVP)

- A control (signalling) protocol for packet neetworks
- Control packets
  - Follow flows (i.e., follows SPF) source to destination
    - Describe the flows (principally bandwidth)
  - Retrace their steps destination to source
    - Making bandwidth reservations
- State maintained in network
- Adaptive to:
  - Changes in SPF
  - Merging of flows
- Not widely deployed
  - But see MPLS-TE

RFC 2205

Multiprotocol Label Switching (MPLS-TE)

RFC 3031

- The MPLS data plane invented for fast forwarding
- The forwarding construct is the Label Switched Path (LSP)
  - Hop-by-hop forwarding state in the network
  - Indexed by a label on each packet
- MPLS-TE places LSPs in the network according to CSPF or central planning

RFC 2702

- Objectives:
  - Reduce the overall cost of operations by more efficient use of bandwidth resources
  - Ensure the most desirable/appropriate path for certain traffic types based on certain policies
  - Reserve network resources for traffic flows
  - Rapid recovery by routing (steering) around failures
- The ultimate goal is cost saving

MPLS-TE Protocol Family

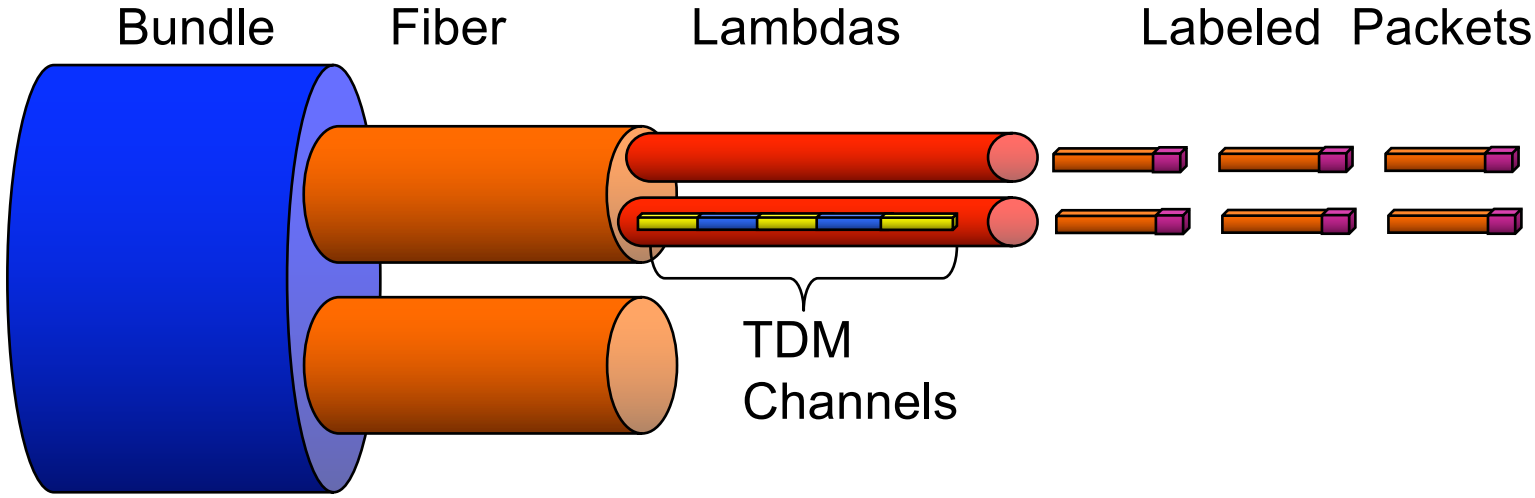
RFC 4202

- Routing dissemination
  - Extensions to the IGPs (OSPF-TE, ISIS-TE)
  - Add information/status about TE qualities of links

RFC 3209

- Signalling
  - Extensions to RSVP (RSVP-TE)
  - Now follow explicit path in preference to SPF
  - LSP setup, modification, teardown

GMPLS : A Label Hierarchy



RFC 3945

- Observe that MPLS-TE is a circuit switching technology based on labels
  - We can generalize the concept to any switching technology
  - Labels move from additions to the packet (headers) to physical identifiers
- Generalized MPLS (GMPLS)
  - MPLS control plane extended for circuits, lambdas, fiber and ports
  - OSPF-TE (and ISIS-TE), RSVP-TE
  - New protocol
    - Link Management Protocol (LMP) to coordinate physical links

RFC 3473

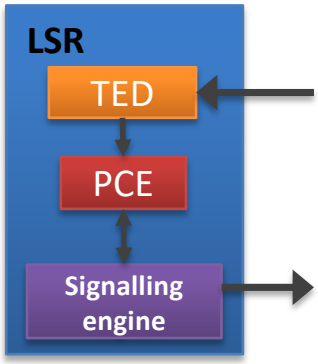
Path Computation Element (PCE)

- *PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints* – RFC 4655
  - This does not say it is a dedicated server
  - It can be embedded in a router
  - It can be embedded in ***every*** router
- For virtual PoP use case
  - PCE function in head-end LSR for local domain
  - PCE function in remote ASBR accessed through remote call

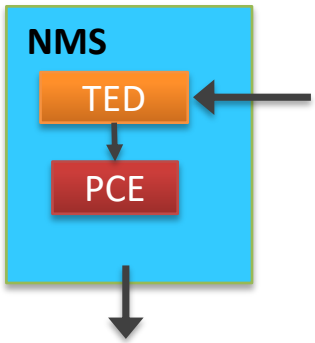
RFC 4655

Realisations of the PCE Architecture

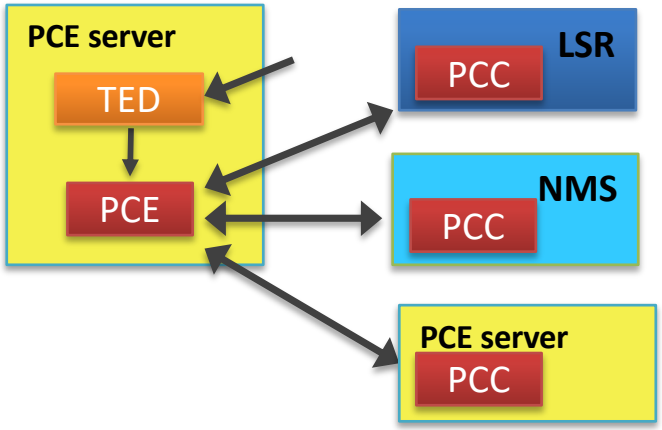
- Historically, head-end LSRs did path computation
  - They included a PCE component
- Historically , the NMS determined paths and instructed the network
  - It included a PCE component
- The PCE architecture recognises these and allows PCE to be externally visible perhaps on a dedicated server



PCE co-located in the LSR



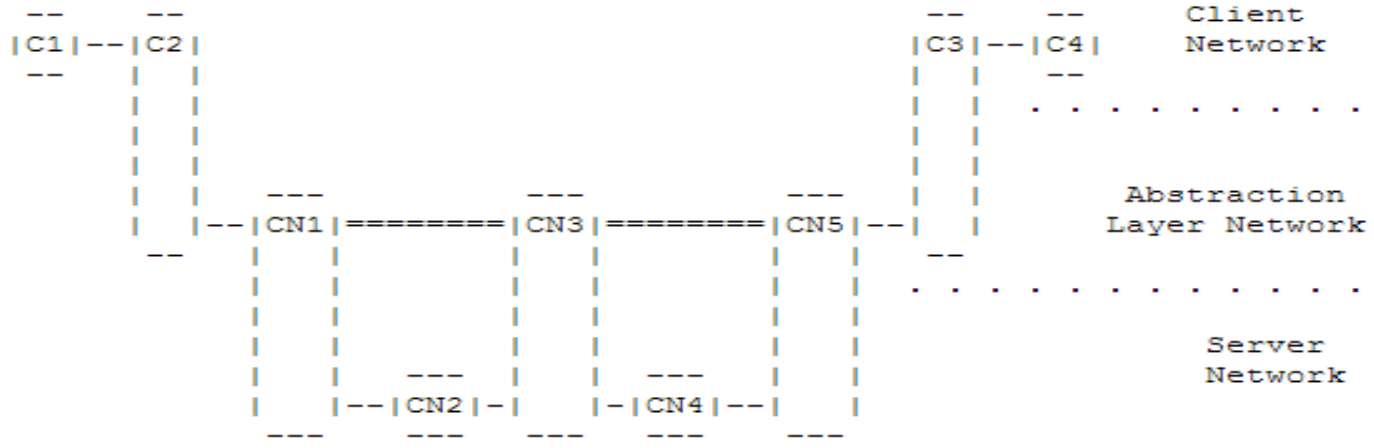
PCE in the NMS



PCE in a dedicated server

Topology Aggregation

- Abstraction Layer Network



RFC 7926

Client layer resources: C1, C2, C3, C4

Server layer resources: CN1, CN2, CN3, CN4, CN5

Abstraction layer resources:

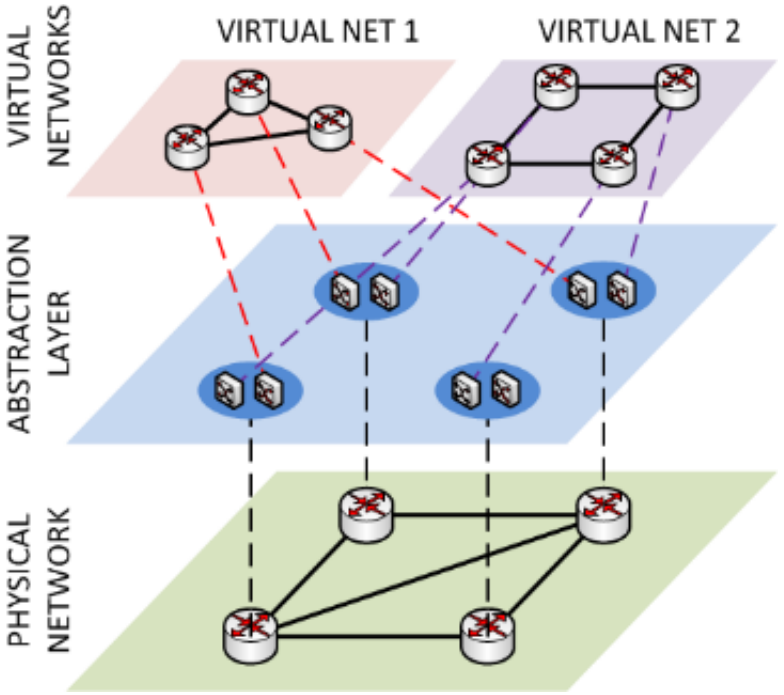
Nodes: C2, CN1, CN3, CN5, C3

Physical links: C2-CN1, CN5-C3

Abstract links: CN1-CN3, CN3-CN5

Abstraction Leads to Virtualization

- Abstraction is about providing a summarised topology of potential connectivity
- Policy-based
  - Policies set by one network with knowledge of the other networks
  - Overcome issues of scaling, stability, confidentiality, and misinformation found in aggregation
    - Hint: virtual node representations may struggle
- Apply policy to the available TE information within a domain, to produce selective information that represents the potential ability to connect across the domain
  - Don't necessarily offer all possible connectivity options
  - Present a general view of potential connectivity
  - Consider commercial and operational realities
- Retain as much useful information as possible while removing the data that is not needed
- Can be further filtered to provide different views for different consumers



Topology Export (BGP-LS)

- Gather topology information via the IGP
- Abstract it using policy
- Now you need to tell someone about it
  - Essentially you have a topology you want to tell someone about
  - The topology might change over time
- Why BGP?
  - BGP implementations are good at applying policy to routing information
  - BGP is good at exporting bulk data
  - Export is from “key points” that often already have BGP
  - Route Reflectors enable multicast of information
- Link State BGP (BGP-LS) is the encodings to do this
- Note well: work in hand to use YANG models for this
- Note also: proposals in hand to use PCEP extensions for this

RFC 7752

IETF Work In Progress

- Lots of tweaks to existing tools
- And some new stuff
  - IGP Flooding Modifications
  - Deterministic Networking
  - Segment Routing
  - PCE as a Central Controller – PCECC
  - Abstraction and Control of TE Networks – ACTN
  - YANG models

Improving IGP Flooding

- Long-term idea in IGPs
  - To reduce the load (on network and device) of flooding
  - May improve convergence in a busy network
- Why is this interesting for TE?
  - Not all nodes need to know about all TE link changes
- Why might it be unnecessary?
  - TE link updates are recommended to be damped, anyway
    - Configured to only re-advertise when change is more than a percentage of residual bandwidth
  - TE state may be relatively static
  - Arrival rates and setup times don't need such rapid convergence
- Nevertheless, may be some interesting work for TE
  - And care is needed to not break any TE functions!

Deterministic Networking (DetNet)

draft-ietf-detnet-architecture

- Deterministic data paths (predictable)
- Operate over Layer 2 bridged and Layer 3 routed segments
- Paths can provide:
  - Bounds on latency
  - Bounds on loss
  - Bounds on packet delay variation (jitter)
  - High reliability
- Work-load split
  - IEEE802.1 Time Sensitive Networking (TSN) responsible for Layer 2 operations
  - IETF DetNet Working Group responsible for Layer 3 operations
- Data plane encapsulations for MPLS and IP
  - It is simple, but not trivial
  - Tunnel-based approach
  - MPLS tunnels may be placed with knowledge of network behaviour
  - All tunnel packets may be marked to help achieve desired qualities

Segment Routing (SR)

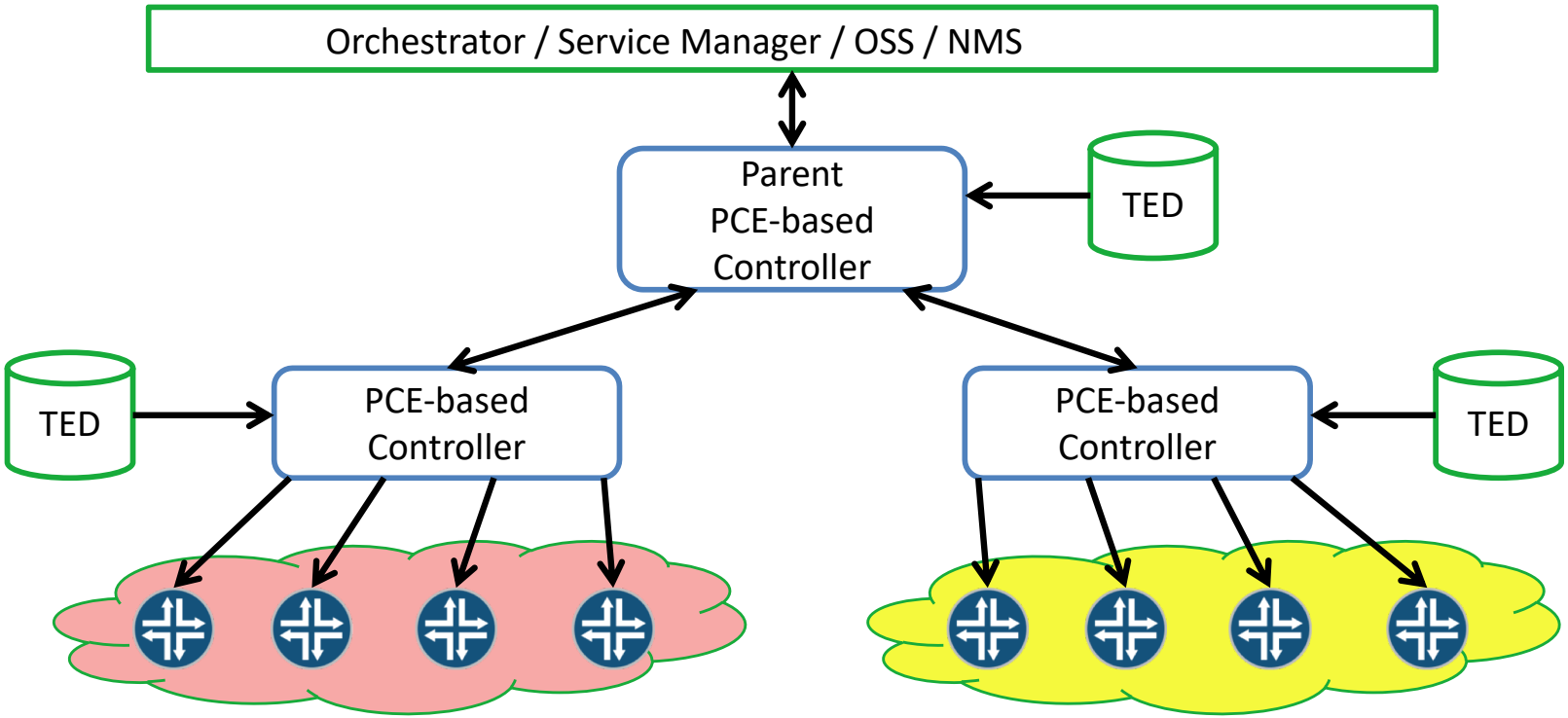
- A tunneling technology
  - Encapsulates a packet within a header
  - Forwards packet based upon encapsulating header
  - Compare and contrast with IP source routing
- A Traffic Engineering (TE) technology
  - Allows a router to steer traffic along an SR path
  - Path can be different from the least cost path
- Maybe more?
  - Innovative new applications to be discovered
- Control plane
  - Signaling removed from the network
  - Routing protocols augmented a little
- Forwarding planes
  - MPLS
  - IPv6
  - ***NOT*** IPv4

RFC 7855

PCE as a Central Controller (PCE-CC)

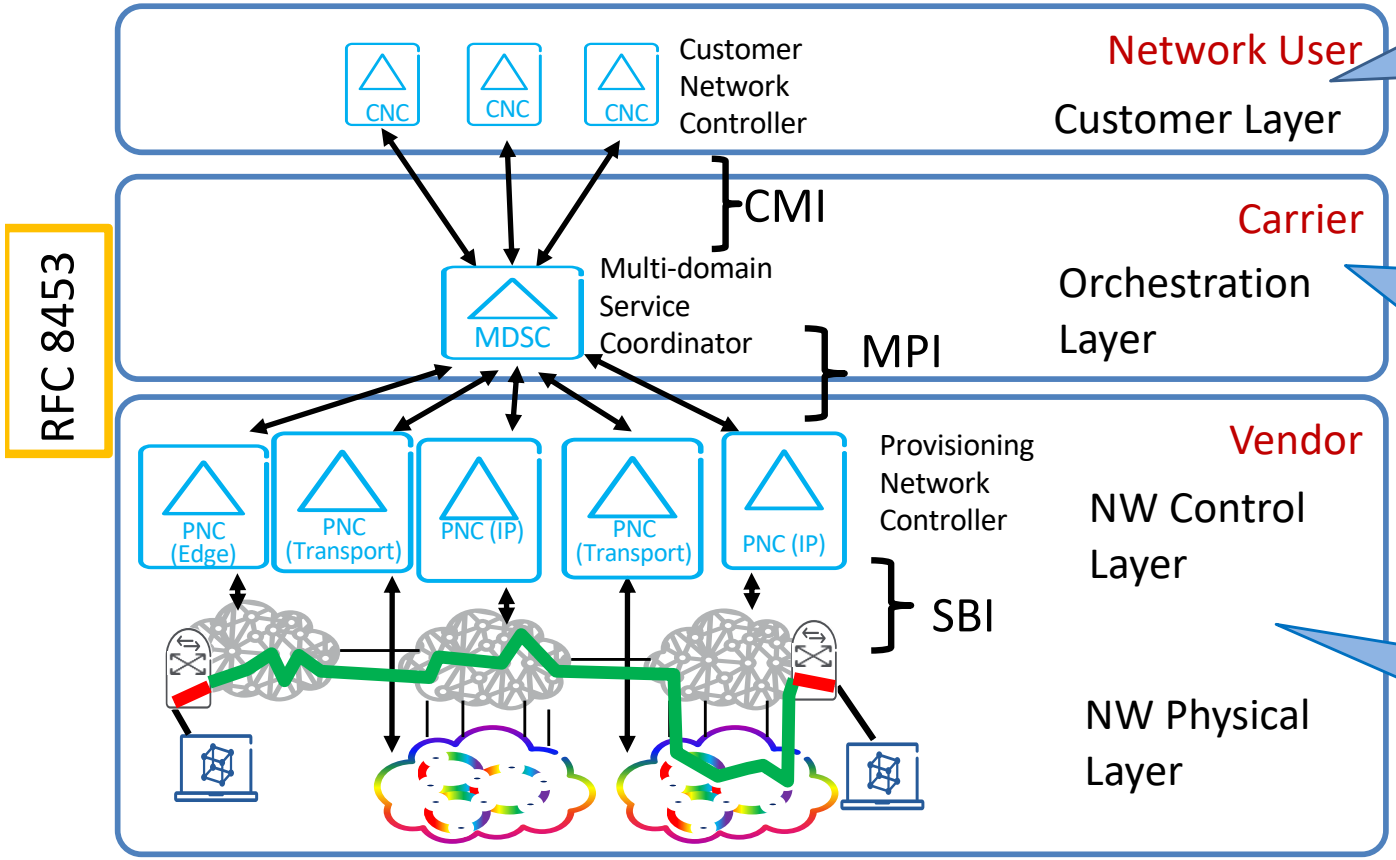
- Integrating PCE into an SDN architecture
  - All southbound exchanges use PCEP
  - Control may be single node
  - Applications proposed in MPLS, non-packet, and IP environments

RFC 8283



Abstraction and Control of TE Networks (ACTN)

Key Idea: introducing SDN controller hierarchies, and make use of abstraction techniques to provide multi-vendor, multi-domain solution

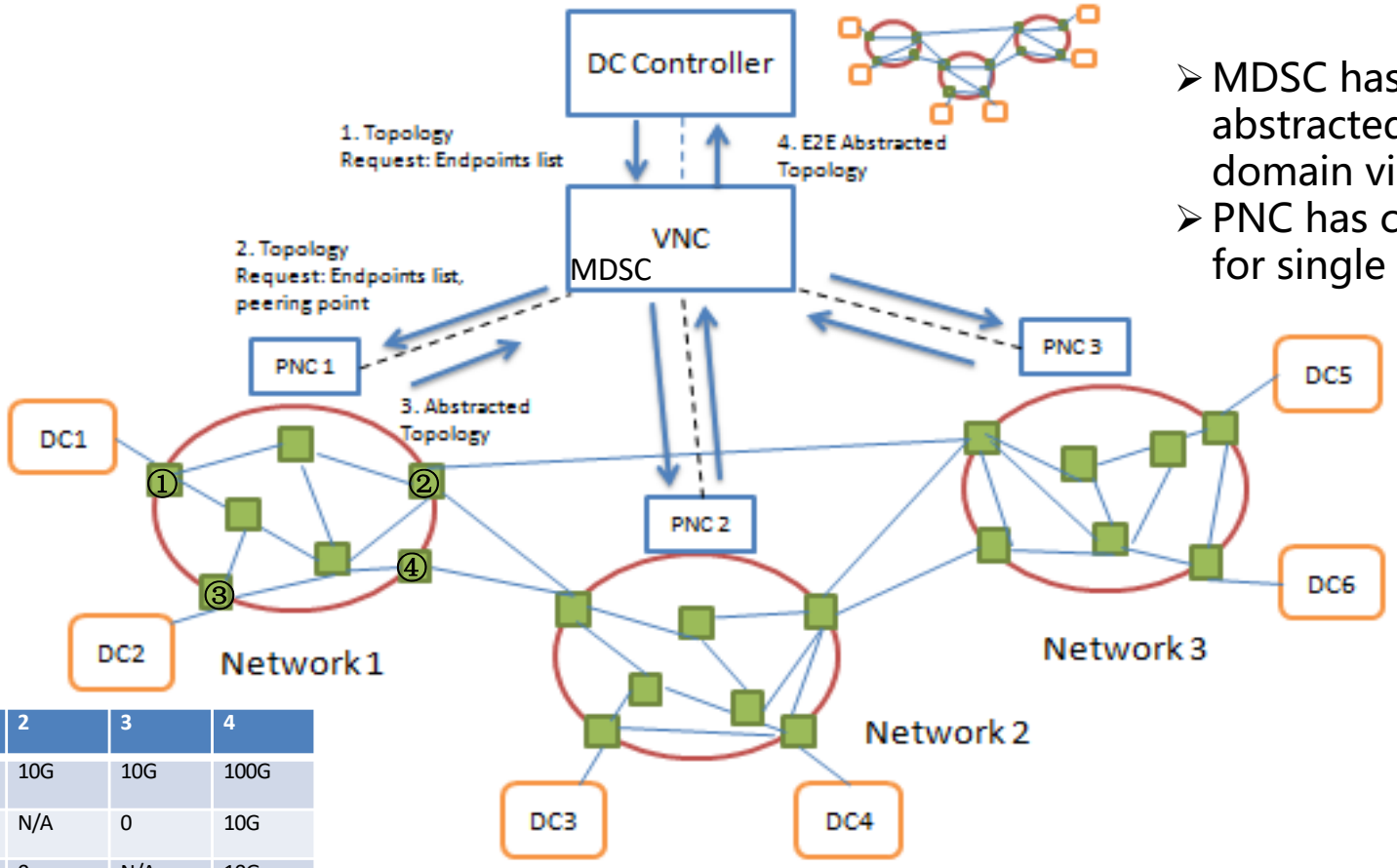


CNC: customer w/o network knowledge; representing application and service, to be understood by operators.

MDSC: bridges user to network.
- ✓ Customer Mapping/Translation;
- ✓ Virtual Service Coordination;
- ✓ Multi-domain Coordination;
- ✓ Abstraction/Virtualization;

PNC: Configures the Network.
- ✓ Control/Manage the NE;
- ✓ Monitoring the topology;

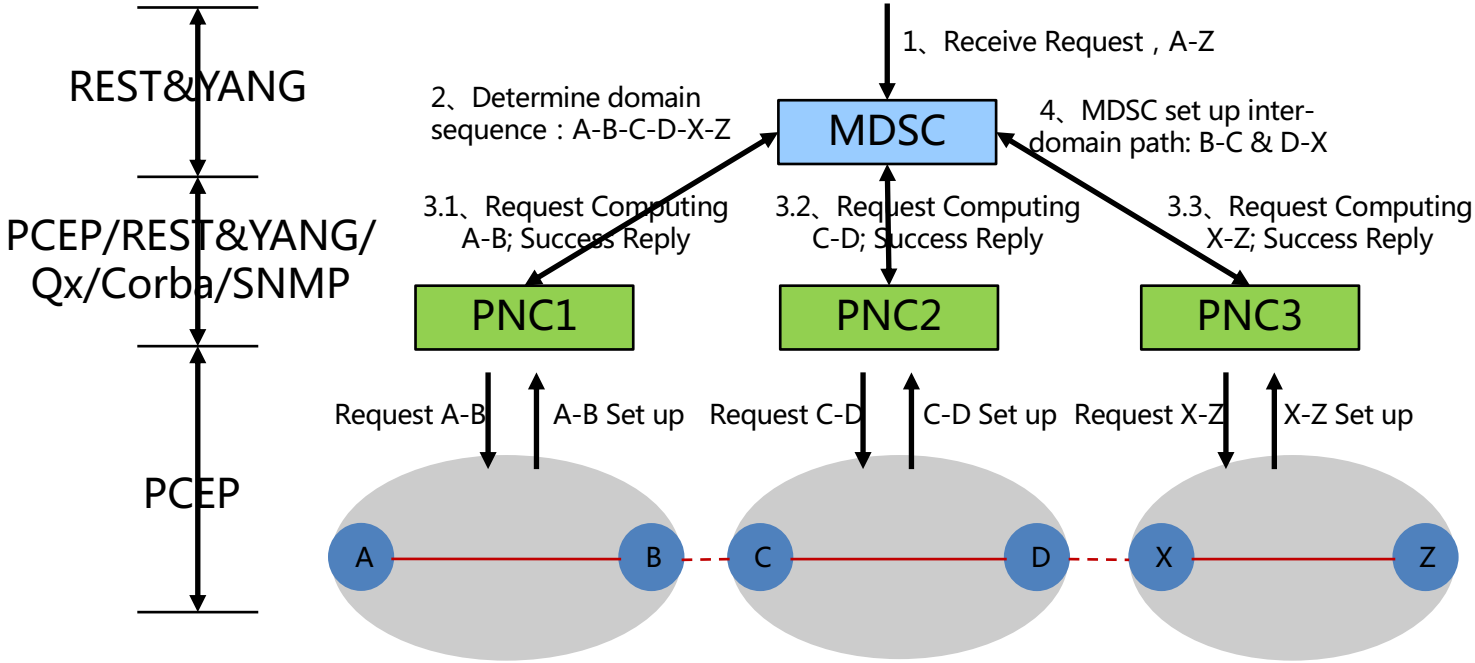
ACTN Scenario 1 : Topology Discovery and Abstraction



- MDSC has an abstracted cross-domain view
- PNC has complete view for single domain

	1	2	3	4
1	N/A	10G	10G	100G
2	10G	N/A	0	10G
3	10G	0	N/A	10G
4	100G	10G	10G	N/A

ACTN Scenario 2：End-to-End Path Computation

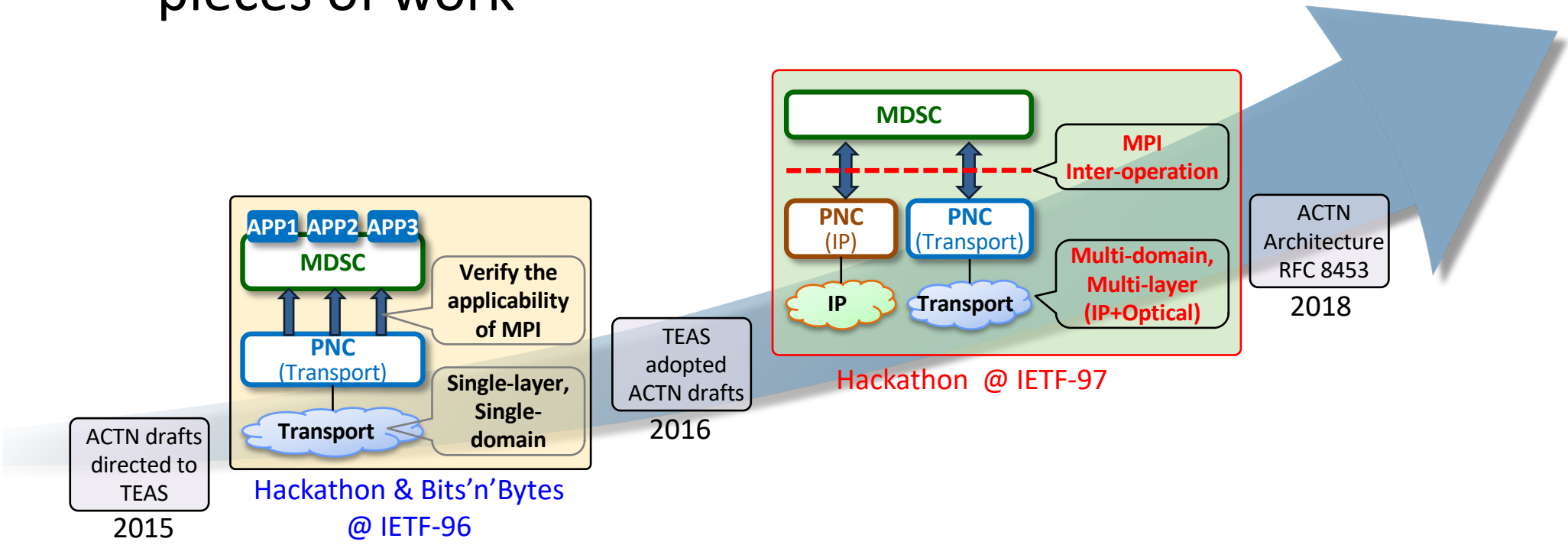


Modeling Requirement：

- ✓ MDSC global view (abstracted)；
- ✓ Unified models on MPI for path computation；
- ✓ Policy & constraint supported；
- ✓ E2E Tunnel Set up after path computation；

ACTN Progress in the IETF

- Demonstrates the time-line for developing significant pieces of work

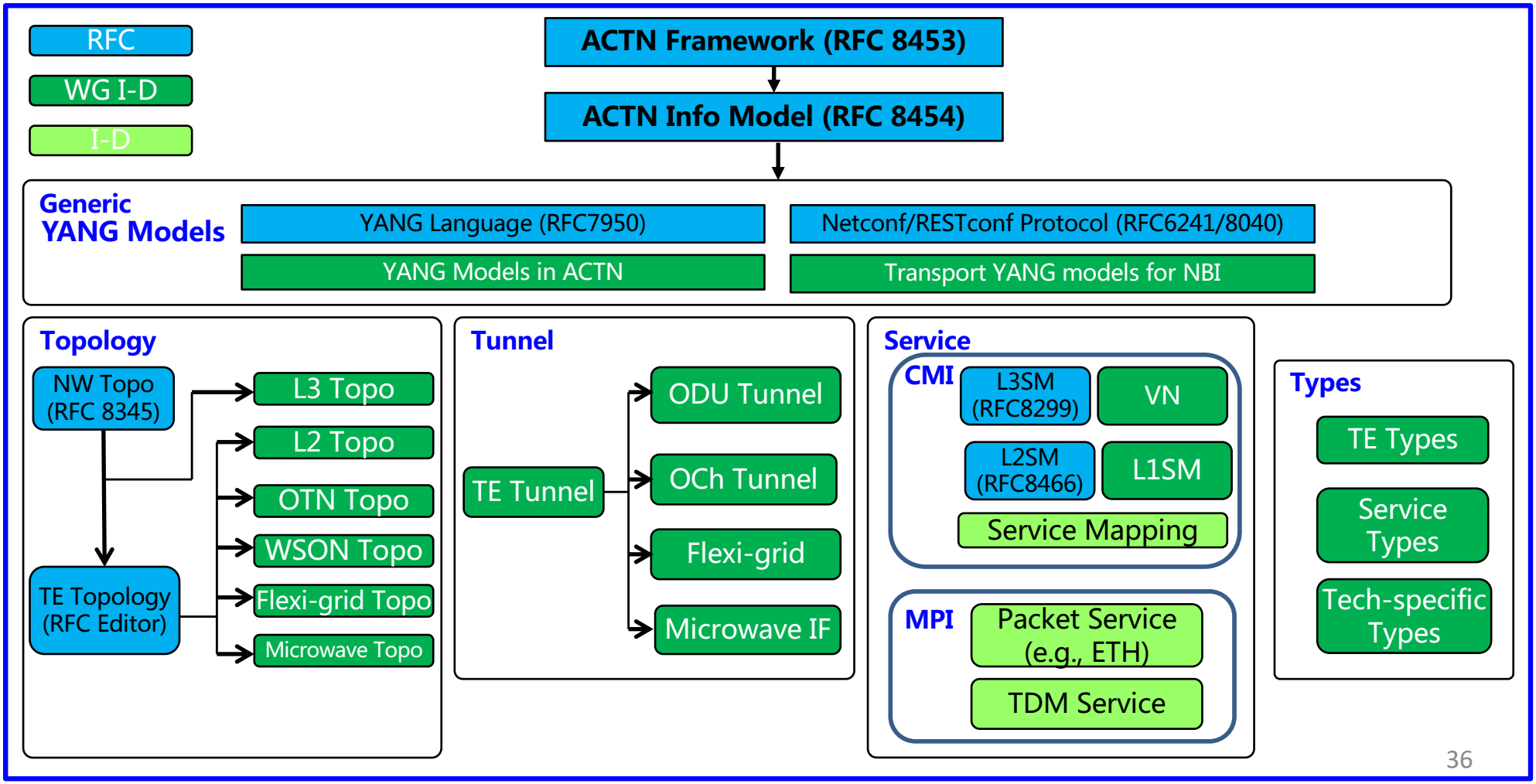


YANG Models

- A Data Model is a formal representation of the data elements that describe and control some element of protocol or networking
  - For example, and OSPF instance, an interface, a link, a topology
- Data models are written in human readable / machine parsable language
  - The IETF's language of choice is YANG
- On the wire YANG is encoded as
  - XML
  - JSON
- The IETF has several hundred YANG models
  - The arrival rate of YANG RFCs is increasing rapidly
- YANG models form a key component of SDN systems
  - Refer back to ACTN where YANG is used between components

RFC 7950

YANG Models for ACTN and TE



Summary

- TE is a valuable tool for operating networks
  - Enables the delivery of enhanced services
  - Allows an operator to get more out of their network
- There is a patchwork of IETF tools and techniques
  - Some of them are widely deployed
- More tools and techniques are being developed
  - To meet new requirements
  - Address evolving architectural models
- Lots of opportunity to get involved

Pointers

- Far too many references to list
- A few key RFCs are called out on specific slides
  - These are only a starting point for each topic
  - Look for other documents they reference
  - Look for other documents that reference them
- Some important working groups for new work
  - TEAS : Anchor for all new TE work
  - MPLS : MPLS-TE when not generic
  - CCAMP : Technology specific work
  - PCE : All PCEP work
  - SPRING : Segment Routing work
  - IDR : BGP and especially BGP-LS
  - NETMOD : YANG protocol and foundation models
  - LSR : Extensions to the IGPs
  - DETNET : Deterministic Networking

Questions and Follow-up

zhenghaomian@huawei.com

adrian@olddog.co.uk

Please provide feedback by taking this very brief survey

https://www.surveymonkey.com/r/trafficengineering