The Sub-IP Area and Optical Networking at the IETF

iGRID
25 September 2002

Scott Bradner
Harvard University
sob@harvard.edu

Syllabus

- the IETF
- the IETF Sub-IP area
- IETF & optical networks

The IETF

- The Internet Engineering Task Force
- standards development for the Internet
- since 1986
- international
  - most recent meeting - July in Yokohama
- individuals not organizations
- no defined membership
- scale: about 2,000 attendees in Yokohama
  - thousands more on mailing lists (from 100s of companies)
- under umbrella of the Internet Society (ISOC)

The IETF Organization

- most work done on mailing lists
- 3 times a year face-to-face meetings
- individuals or groups request BOFs
  - exploratory meeting - may lead to working group
- working groups for specific projects
  - about 135 working groups
  - restrictive charters with milestones
  - working groups closed when their work is done
- working groups gathered together into Areas
  - each area has 1 or 2 Area Directors - managers

IETF Areas

- Applications Area
- General Area
- Internet Area
- Operations and Management Area
- Routing Area
- Security Area
- Sub-IP Area
- Transport Area

IETF Standards Process

- "rough consensus and running code"
  - rough consensus required not unanimity
  - interoperable implementations needed to advance standard
- multi-stage standards process
  - Proposed Standard: good idea, no known problems
  - Draft Standard: multiple interoperable implementations
  - Standard: market acceptance

Above and Below

- traditionally the IETF has been:
  - "above the wire and below the application"
  - not (often) defining user interfaces
  - not defining physical wire types
- while doing "IP over foo"
  - "foo" has been types of networks
    - Ethernet, Token Ring, ATM, SONET/SDH, ...
  - but foo has been changing

IP over "Trails" "Circuits" "Paths" ...

- what looks like wires to IP may not be physical wires
  - may instead be something where paths can be configured
  - where a path looks like a wire to IP
    - e.g. ATM VCs & optical networks
  - might also be routed datagrams another layer down
    - e.g. IPsec tunnels
- and then there is MPLS
  - a progressively more important "foo"

Layer Violations

- there is another complexity when the sub-IP technology is configurable
  - e.g. MPLS, ATM, Frame Relay, ...
- how should the sub-IP technology be controlled?
  - what information should be taken into account?
  - question may be "could a new path exist with certain characteristics"
  - not just "can (or does) a path exist?"

A New IETF Area

- a systematic approach to sub-IP issues would be nice
  - but exact scope is not clear
- IESG created a temporary area for sub-IP
  - like what was done for IPng
- to be short lived (1-2 years)
  - 2 current ADs were appointed to run the area
  - Bert Wijnen & Scott Bradner
  - looks like 2ish years

Non-Objectives

- the IETF is not expanding into standards for physical or virtual circuit technologies
  - no new circuit switch architecture from IETF
  - e.g., the IETF is not working on optical switches
  - leave them to others
- need to communicate with other standards organizations on what we are actually doing

A Crowded Field

- many other standards organizations working in this area
  - ITU-T, MPLS Forum, IEEE, ATM Forum, ...
- need to work out ways to coordinate and cooperate
  - bi-lateral arrangements to minimize redundant efforts
    - but they will not be eliminated
- IETF needs to know what not to do
  - at the same time it and others need to know what it needs to have a hand in

Sub-IP Area Work

- "Layer 2.5" protocol: MPLS
- protocols that monitor, manage or effect logical circuit technology
  - e.g. IP Over Optical, Traffic Engineering, Common Control and Management Protocols
- protocols that create logical circuits over IP
  - e.g. Provider Provisioned VPNs
- protocols that interface to forwarding hardware
  - General Switch Management Protocol

Working Groups in Sub-IP Area

- Internet Traffic Engineering (tewg)
  - principles, techniques, and mechanisms for traffic engineering in the internet
- Common Control and Management Protocols (ccamp)
  - measurement & control planes for ISP core tunnels
  - info collection via. link state or management protocols
    - e.g. OSPF, IS-IS, SNMP
  - protocol independent metrics to describe sub-IP links
  - signaling mechanisms for path protection

Sub-IP Area WGs, contd.

- Multiprotocol Label Switching (mpls)
  - label switching technology
  - RSVP & CR-LDP signaling to establish LS paths
  - MPLS-specific recovery mechanisms
- Provider Provisioned Virtual Private Networks (ppvpn)
  - detail requirements for ppvpn technologies
  - define the common components and pieces that are needed to build and deploy a PPVPN
  - BGP-VPNs, virtual router VPNs, port-based VPNs (L2)
  - security

Sub-IP Area WGs, contd.

- IP over Optics (ipo)
  - framing methods for IP over optical dataplane and control channels
  - identify characteristics of the optical transport network
  - define use of ccamp protocols for optical networks
- General Switch Management Protocol (gsmp)
  - label switch configuration control and reporting

Sub-IP ex Working Group

- IP over Resilient Packet Rings (iporpr)
  - input to the IEEE RPRSG to help it formulate its requirements
  - moved to Internet Area

What's In and Out?

- boundaries of IETF work have been blurry
  - the sub-IP area did not help clarify this
- basic concept:
  - the IETF works on IP-related technology
  - if something does not have a relationship to IP networks then the work should be done elsewhere
- but since many networks (e.g. all-optical) carry IP control of those networks may be IP-related
  - but MPLS support for power distribution is out of bounds
    - see RFC 3251

Partitioning between WGs

- some overlap between working groups
  - e.g. tewg and ccamp and mpls
  - tewg explores the requirements for control of sub-IP networks
  - ccamp defines tools to control of sub-IP nets
  - some of the tools are mpls specific
- careful coordination required
  - main mission of the sub-IP directorate

Summary

- created temp area to coordinate IETF sub-IP work
  - area to last a year or two
- will reevaluate experience soon
- most work of the sub-IP WGs should be done by the time the area is closed
- any remaining working groups will be distributed to existing IETF areas
- above from when the Sub-IP area was formed
  - looks like it will be closed early next year (i.e., ~2 years)

Sub IP Conceptional Organization

```
Applications           +-------+  +-------+        (new) Hour glass
that use CCAMP: \      | TE-WG |  | PPVPN |  ...             /
                 \     +-------+  +-------+                 /
                  \      +----------------------+          /
                   \     |        CCAMP         |         /
                    \    |-----------+----------|        /
                    /    |   C       |    M   --|------ IGP LSA ext
                   /     | control   | measure  |        \
                  /      +----------------------+         \
Technologies to  / +-----+ +-----+ +-----+ +-----+ +-----+  \
measure/control:/  |MPLS|  |OPT |  |RPR |  |ATM |  | FR |...\
                   +-----+ +-----+ +-----+ +-----+ +-----+
```


IETF Sub-IP Basic Architecture

- for all sub-IP network types
  - not just pure optical nets
- two main components
  - topology discovery
  - control signaling
- development work being done in IETF ccamp working group

Traffic Engineering

- aim: combat congestion at a reasonable cost
  - networks w/o congestion are not a problem
  - other than speed of light issues
- decide paths through network rather than letting routing do its thing
  - paths could be in infrastructure: ATM PVCs, Frame relay PVCs, optical (SONET, Ethernet & other)
  - paths could be IP-ish: MPLS
- note - tail circuits a common congestion point
  - but can not be traffic engineered around
- same issue with servers

TE Requirement

- measurement system
  - you need to know what is wrong before you can fix it
- need to know where there are congestion problems
- hard to know what to measure
  - link utilization by itself is not enough
    - but may indicate trends
  - router drops (packets dropped for lack of resources)
    - tell you when there is a problem
  - harder if QoS in use (like diffserv)
    - router counters do not say what type of packet was dropped

TE Requirement, contd.

- reporting system for link utilization could be tricky
  - what sample period
  - what hysteresis algorithm should be used
    - too fast a reaction will cause churn
- reporting on a large network could be a problem
  - what propagation delay is OK
  - what information do you actually need?
  - too much information is a waste

TE Steps

- define control policies
  - what are you trying to achieve?
- measure
  - find out what’s going on now
  - “now” is a variable
- analyze
  - measurement results and requirements
- optimize
  - configure network to provide “best” service
  - may include restricting input

TE Assumptions

- TE assumes that the capacity of the net is not evenly distributed
  - i.e. some links are bigger than others
  - and some links are underutilized
- TE assumes that the load is not evenly distributed
- i.e. TE assumes that directing traffic in a way that routing would not has benefit
  - not the case where there is well distributed excess bandwidth
  - or where there is not an alternative path

TE Example

- UUnet used an underlying ATM network
- city-PoPs interconnected with ATM PVCs
  - full mesh
- PVCs configured for specific bandwidths
- PVCs configured to follow specific paths
- traffic stats recorded for each PVC
- VC bandwidths & paths recomputed occasionally
  - somewhere between daily & weekly
- new VCs installed when needed

TE and QoS

- initial TE work was directed at general QoS
  - i.e. aimed at reducing congestion
- not type of service specific
  - i.e. no per-service type TE
- but now QoS seems to be a great desire
- seen as a way to make datagram networks look like circuit-based networks QoS-wise
  - is that realistic?

Traffic Engineering Future

- alternatives coming along
- more bandwidth
  - bandwidth getting cheap
  - but not everywhere
    - e.g. international or in enterprise WANs
- link metric manipulation
  - configure the link metrics on IGP
  - can direct traffic along desired paths
  - but very complex software

Just do Routing

- some research that says you can do it all with link-state routing
  - adjust link metrics in link-state routing protocol
    - every link gets a computed metric
  - can balance traffic across net based on link size
    - i.e. make full use of resources where they exist
  - assumes load split across paths with equal metrics
    - assumes microflows are not split (no packet reordering)
    - does not deal with the case where a single micro flow is bigger than a link

Traffic Engineering Reminder

- most common points of congestion in the Internet are:
  - customer connections (tail-circuits)
  - servers
- ISP traffic engineering will not fix these problems
  - i.e. the user will still see poor “network” performance

MPLS

- Multiprotocol Label Switching
- basic functions:
  - direct packets in a way that routing would not have
    - but not required feature
  - enable packet forwarding based on things other than IP destination address
  - simplify network core (e.g., no routing needed)
  - aggregate traffic with some common characteristics
  - can provide traffic matrix data
  - apply QoS to specific traffic group

coin2002 - 33

MPLS, contd.

- not really routing (was in IETF routing area)
- circuit-based path setup
- original purposes:
  - traffic engineering & forwarding speed
- moving into QoS
  - circuit per QoS class -> circuit per flow
- some treating MPLS like packet-based ATM

coin2002 - 34

MPLS & Performance

- older IP routers were slower than switches
  - more processing required
- MPLS core network is a switch network
  - common assumption: MPLS switches would be easier (cheaper) to build and faster than IP routers
  - true at the time - no longer generally true
- most routers today use ASICs in the forwarding path
  - run at "wire speed" for very high speed wires
  - small (if any) cost difference compared to MPLS ASICs

coin2002 - 35

MPLS Overview

- at ingress: group traffic into **forwarding equivalence classes** (FECs)
  - traffic to be handled in the network in the same way
- ingress router uses whatever criteria it wants to
  - destination addr, source addr, protocol, router input port, diffserv class, etc
- label prepended to packet to specify FEC
- **label switch routers** (LSRs) in network use labels to select next hop: **label switched path** (LSP)
- label removed at egress

coin2002 - 36

MPLS, LSR Databases

- LSR has table of **Next Hop Label Forwarding Entries** (NHLFE)
  - entry includes output interface, next_hop IP address, label manipulation instructions
  - can also include new label
- **incoming label map** (ILM)
  - map from incoming labels to NHLFEs
- **FEC-to-NHLFE map**
  - map from incoming FECs to NHLFEs

MPLS, LSR Processing

- label from incoming packet mapped (using ILM) to NHLFE
- LSR processes label manipulation instructions e.g.
  - pop label
  - swap with new label
  - swap with new label and push a new label onto stack
- labels locally significant
  - no requirement for wide spread synchronization
- forward packet to next_hop
  - may need to change L2 encapsulation

MPLS, Label Stacks

- can have more than one label on a packet
  - "label stack"
- label stack can be used to implement trunking
  - many LSPs can be seen as one
    - as long as they are taking the same route
  - e.g. MPLS-enabled phone calls accumulated in a trunk
- exit LSR pops label and then uses L3 routing

MPLS, Path Installation

- path information installed in LSRs by:
  - manual configuration
  - RSVP-TE
  - Label Distribution Protocol (LDP)
    - uses destination address prefixes
  - Constraint-Based Label Distribution Protocol (CR-LDP)
- can follow underlying routing paths
- or path can be explicitly placed

MPLS, Original Purpose

- defining paths for large city-pair like trunks
  - i.e. Internet Service Provider traffic engineering
  - make up for unequal distribution of bandwidth vs. load
- in use at some large LSPs
- full mesh between core routers in pops
  - e.g. 20 pops
    - 2 core routers each = 40 routers
    - 780 LSPs ((40) * (40-1)) / 2
- class of service additions
  - N classes of service = N * 780 LSPs

MPLS, Imagined Uses

- MPLS now seen by some as a way to introduce circuits to the Internet
  - Virtual Private Networks (VPNs)
  - per-application path selection
  - generalized tunneling protocol
- label stacks to support scaling
  - many levels envisioned
- whatever ATM was thought to be good for
- "they are trying to replace IP"

MPLS, Example: VoMPLS

- VoMPLS phone does not run IP - runs MPLS instead
- call encapsulated in MPLS
- call setup sets a path through MPLS network to destination - e.g. with RSVP
  - could be another VoMPLS phone
  - or VoMPLS / PSTN gateway
- end-to-end LSPs run through trunks where possible
- local, regional, national & international trunks
  - i.e. multiple layers of lables

coin2002 - 43

MPLS, Issues

- scaling
  - state in LSRs
  - management
- other
  - multiple signaling options
  - inter-provider connections
  - rationale

  ATM-like assumed uses

coin2002 - 44

CCAMP

- 2 separate objectives
- measure current state of sub-IP links
  - the links that make up the IP-level links
  - e.g., the links between ATM or optical switches
- control (signaling) protocol to manage sub-IP network
  - manage with IP protocol
- 1st product: GMPLS

GMPLS

- generalized MPLS
- assumes sub-IP links can be controlled with tags
  - extension of MPLS concepts
- routing algorithms do not need to be standardized
  - can compute explicit routes
- can do link bundling for scaling
  - parallel links between switches can be treated as a bundle
- data and control planes do not need to be the same

Architecture

- separate control & data planes
  - out of band signaling (by definition)
  - do not need to use same media
- split control plane
  - signaling plane
  - routing plane
- extend MPLS to link technologies where forwarding plane can not see packet or cell boundaries
  - i.e., label refers to time slots, wavelengths or physical ports
- attempt to be link technology independent

Control for Multiple Link Types

- link types
  - (PS) packet switch: e.g., IP networks
    - (can be done with MPLS or GMPLS)
  - (L2S) layer-2 switch: e.g., ATM
  - (TDM) time-division mux: e.g., SDH/SONET
  - (LS) lambda switch: e.g., optical wavelength-based
  - (FS) fiber-switch: e.g., switch between physical fibers
- link bundling
  - group set of parallel links into a single logical link
  - e.g., multiple lambdas on a fiber
- supports link nesting

GMPLS Routing Plane

- uses link-state routing protocol between switches to report on link status, characteristics & constraints
  - note, below the IP layer
- can use OSPF or IS-IS with TE extensions
- can do path determination with routing protocol or using explicit routing

coin2002 - 49

GMPLS Signaling

- GMPLS extends RSVP-TE & CR-LDP
  - up to vendor to decide which to use
  - most vendors use RSVP-TE
- uses Link Management Protocol (LMP)
  - runs between data-plane-adjacent nodes
  - manages bundled links
    - maintain control connectivity, verify physical connectivity of data links, correlate link characteristics, manage link failures
  - link technology independent

coin2002 - 50

GMPLS Signaling Building Blocks

- new generic label request format
- Generalized Label to support TDM, LS & FS
- waveband switching support
- label suggestion by upstream
- label restriction by upstream
- bi-directional LSP establishment
- rapid failure notification
- protection information
- explicit routing with explicit label control
- per technology traffic parameters
- LSP administrative status handling

Optical UNI & NNI

- GMPLS does not separately specify User Network (UNI) or Network-Network (NNI) interfaces
  - UNI: interface between user an network cloud
  - NNI: interface between two network clouds
- GMPLS can be used as a UNI or NNI but IETF not specifically defining how
- OIF has defined a UNI using GMPLS

GMPLS Status

- docs will soon be approved for RFC publication
- 22 implementations reported

Politics: IETF Optical Work

- technologies for Internet service providers (ISPs)
  - not necessarily anyone else - but may be useful to others
- i.e, IETF works on technology for the Internet (including private IP networks), the technology may be useful for networks not carrying IP but it's not a design goal
- ways to control optical networks from IP point of view
  - based on IETF traffic engineering technologies
  - i.e., intelligent IP-based control plane for optical networks

Technology: IETF & Optical Networks

- GMPLS
- IP Over Optical Working Group
  - framework for using IP on optical networks
  - framing for IP on optical networks
  - identifying characteristics of optical nets important to IP control
  - document control requirements
  - document the applicability of IP-based protocols for control of optical networks

The Internet & Optical Networks

- to the Internet a lambda switched optical network is another link layer
  - not an end-to-end circuit
- could be a point-to-point link between routers
- different case for optical packet switched networks
  - not "tomorrow" but I'd like to install some before I retire

coin2002 - 57