ONTC PRISM Newsletter

Dear ONTC Members,

A warm welcome from the Editorial desk of the ONTC's newsletter "Prism". It is our pleasure to bring to you this third edition of PRISM – ONTC's quarterly newsletter. The industry is moving towards two interesting trends, of consolidation and new business models. Providers and vendors are both adopting new business models as well as new technologies that will have a significant impact

The current month focuses on carrier class technologies for packet based telecommunication networks. We have in this issue a good fit between academia and industry, covering various aspects of packet backbones that are making up much of transport networks. The impact of transport networks is important from the larger telecommunication business perspective.

The industry just concluded the IEEE 802 Standards' Plenary in San Diego. The next 802 plenary will be held in Dallas, in November. It will be encouraging to see more academics participate in the .1 and .3 sessions. The ITU study group meets next on Oct 18-22 in Geneva.

In this issue of Prism we bring to you 3 articles on a wide range of topics. The first article is by Nurit Spreecher and Yacov Weingarten from Nokia Siemens Networks who describe in detail the working of the emerging Ethernet standard of MPLS-TP. The MPLS Transport Profile technology is being heralded by vendors and providers as the new way to transport Ethernet across large networks achieving low-cost and making much use of the existing MPLS infrastructure. The authors describe the work in the standards, the technological manifestations and operational issues pertaining to MPLS-TP

The second article describes the other alternate technology for Carrier Ethernet transport. The article written by David Allan, (Ericcson) Peter Ashwood-Smith, (Huawei) Nigel Bragg, (Ciena) János Farkas, (Ericcson) Don Fedyk, (Alcatel-Lucent) Mick Seaman, Paul Unbehagen (Alcatel-Lucent) talks for the new 802.1as emerging standard on shortest path bridging. The Shortest Path Bridging or SPB technology gives a significant edge to PBB-TE networking and has been explained in sufficient detail in the article. The authors point out to a forthcoming longer article on the same topic in the Sept issue of IEEE Communications Magazine.

The third article is by Mohit Chamania, Admela Jukan, Oscar Gonzales de Dios, Javier Jimenez Chico on offloading excess IP traffic with optical bypass. The article is important to providers from an opex savings perspective and gives a new direction to optical networking research, especially making them more relevant to IP/MPLS networks. The article paves a way to a more pragmatic view of IP over WDM optical networks.

As you are aware the editorial desk does highlight events, call for papers and new standards work – all of which would be relevant information for our readers. So please do send in any such information that you would like to see online for the next issue of Prism!

We also do hope readers would send in their thoughts on how to make Prism better – we would be happy to publish their messages even if all of these cannot be adopted at the same time!

We invite prospective authors send articles of up to 4 pages (single column, 10 point font, with all one-inch margins) to submissions@ontc-prism.org. The deadline for reception of articles is Sept 30, 2010 for the next issue of Prism.

On behalf of the TAB we are thankful to the IEEE Communication Society as well as to the ONTC officers Byrav, Suresh, Admela and Dominic who have supported us in making this newsletter happen.

It is our hope that the newsletter would bring the community together and identifying areas of growth and common interest.

Ashwin Gumaste, IIT Bombay.

Message Board

Standardization:

IEEE SIEPON		http://grouper.ieee.org/groups/1904/1/
IEEE Interim Meeting (802)	Sept 13-16, York, UK	http://www.ieee802.org/1/meetings/index.html#sep10gen
IETF	Nov 7-11, Beiging, China	http://www.ietf.org/meeting/79/index.html
IEEE Plenary	Nov 7-11, Dallas, USA	http://ieee802.facetoface-events.com/future
ITU SG15	Oct 18-22	http://www.itu.int/ITU-T/studygroups/com15/index.asp

Academic Conferences

IEEE/OSA OFC 2011	March 6-10. Los Angeles	http://www.ofcnfoec.org/
IEEE Globecom 2010	CFP March 31 Conf Dec 6-10. Miami	http://www.ieee-globecom.org/
IEEE LCN 2010	CFP April 12. Conf Oct 11-14 Denver, CO.	http://www.ieeelcn.org/
IEEE ICC 2011	CFP Sept. 7: June 5-9, 2011, Kyoto, Japan.	http://www.comsoc.org/confs/icc/2011/index.php
IEEE ANTS 2010	CFP July 15: Conf: Dec 15-17, Bombay, India	www.ieee-ants.org
Infocom 2011	Conf April 10-15, Shanghai, China.	http://www.ieee-infocom.org/2011/

The next ONTC meeting would also be held in Miami, during IEEE Globecom.

We would be happy to include more conferences in the above list, if readers email editor@ontc-prism.org a CFP of the conference. The conference must be at least technically supported by ONTC or ComSoc to be included in the list above and follow the ONTC endorsement policy.

Key journals reporting results in the optical networking area:

IEEE/OSA Journal of Optical Communication and networks (JOCN)
http://www.opticsinfobase.org/jocn/journal/jon/author.cfm
IEEE/OSA Journal of Lightwave Technology: http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=50
IEEE/ACM Transactions on Networking: http://www.ton.seas.upenn.edu/
IEEE Communications Magazine: http://mc.manuscriptcentral.com/commag-ieee
IEEE Network: http://dl.comsoc.org/ni/

MPLS-TP

Nurit Sprecher, Nokia Siemens Networks

Yaacov Weingarten, Nokia Siemens Networks

Introduction:

The evolution of the Service Provider network is being driven by different factors. A major factor is the accelerating growth and penetration of new packet based services (such as L2/L3 VPNs, VoIP, IPTV, RAN backhauling, etc.). This factor provides the SPs with massive growth opportunities to support high BW intensive services that require a guaranteed SLA. In addition, the need to increase their revenue, while remaining competitive, forces service providers to look for the lowest network Total Cost of Ownership (TCO). A single, simple, end-to-end manageable, packet based network architecture supporting multiple solutions (such as mobile backhauling, enterprise and residential services, etc.) would give service providers the flexibility to rapidly deploy new revenue generating services in an economic way.

For many years, SONET/SDH has provided service providers with a high benchmark of reliability and operational simplicity. When migrating to packet based transport networks, some challenges need to be resolved. The solution must:

- be cost effective such that the investments in both CAPEX and OPEX will be minimal.
- ensure quality and per service end-to-end SLA, and
- support any service (including legacy), at any scale, with global reach (across cities, countries and continents).

MPLS-TP, the transport profile of MPLS, which is currently being defined in the IETF jointly by the IETF and the ITU-T, can enable the migration to a packet-based converged network that will easily scale to support Carrier Ethernet Solutions in a simple and cost-effective way.

MPLS is a mature, proven, and ubiquitous technology which already plays an important role in packet networks and services. It includes both a basic architecture (dealing with forwarding, framing, and encapsulation) and a comprehensive set of tools. The tool-set supports the ability to –

(1) setup and maintain of reliable and quality transport paths and
(2) enable the delivery of the services.

The IETF/ITU-T Joint Working Team (JWT) was able to show that a great deal of the IETF protocol, design and architectural work could be reused to meet transport network requirements. The work on MPLS-TP is aimed to integrate the MPLS transport profile into the IETF MPLS and PWE3 architecture, combining the packet experience of MPLS with the operational experience of SONET/SDH. This work is being carried out under the following guiding principles:

(1) reuse (as much as practically possible) existing MPLS standards, and
(2) ensure complete interoperability and architectural soundness with IP/MPLS.

The work on MPLS-TP extends the existing MPLS toolkit with additional architectural elements and mechanisms that are needed to enable optimal Operational Expenditure for transport requirements. MPLS-TP is a profile of the extended MPLS toolkit that defines the necessary and sufficient subset of tools which ensure scalable and quality transport while providing optimal TCO.

It should be noted that by defining a toolkit, these new tools may also be used in an existing IP/MPLS environment to enhance the general behavior of the network and the services. A toolkit allows SPs to expand and customize their service offering efficiently with minimal effort.

MPLS-TP Scope:

The scope of MPLS-TP is the packet transport layers. The goal is to enable the delivery of L2, MPLS(-TP) and IP services over different transmission media with MPLS-TP providing the platform for intelligent and efficient scaling and traffic engineering.

MPLS-TP supports the separation of the service delivery architecture, in the *service layer*, from the underlying transport architecture, in the *packet transport layer*, in order to provide scalability, efficiency, and security.

The main role of the packet transport layer is to provide connectivity. This connectivity needs to provide efficient, reliable, long-standing, aggregated (P2P and P2MP) *transport paths* between service termination or switching points. The transport paths are transparent to the service creation and modification, thus avoiding unnecessary protocol complexity. Bandwidth management, resiliency, QoS, OAM, and provisioning are all performed effectively and in a scalable way per the aggregation level. The number of states maintained in the network is relatively low; being proportional to the number of supported transport paths, independent of the number of services. Packet forwarding is based on SP-controlled mechanisms and addresses.

The service layer, on the other hand, is service and application aware. It is a high touch networking, which deals with the dynamic subscriber management and addresses all of the protocol complexity related to BW management, QoS, resiliency, OAM and provisioning that is configured per service. The number of states maintained at this level, is proportionate to the number of services supported by the network.

The services are handled at the edges of the packet transport network (service termination/switching points) and are encapsulated and transmitted transparently across the transport network.

MPLS-TP Data-Plane

MPLS-TP is based on existing pseudowire and LSP constructs and does not modify the MPLS forwarding architecture. It provides support for unidirectional LSPs, co-routed bidirectional P2P LSPs, and unidirectional P2MP LSPs. Corouted bidirectional LSPs, consist of two unidirectional LSPs that both traverse the same bidirectional transport path, and the intermediate nodes are aware of the pairing relationship between these LSPs.

MPLS-TP supports –

- delivery of highly-reliable, differentiated transport services.
- Traffic Engineering for efficient utilization of network resources.
- mechanisms to address QoS (end-to-end BW guaranteed) and performance requirements (such as throughput, controlled jitter and delay, packet loss, etc.) while utilizing the network resources efficiently and reliably.
- deterministic operation that ensures secured networks and services
- MPLS-TP provides efficient transport for all services: Ethernet, TDM, ATM, FR, MPLS(-TP), IP, etc. and
- supports all transport service types (p2p, p2mp, mp2mp).

MPLS-TP does place certain restrictions that are a result of the nature of transport networks. These include –

- LSP merging is disabled to allow a label to uniquely identify a transport path (as OAM operations require a packet context).
- PHP is disabled by default.

In-band communication Channels:

MPLS-TP defines different in-band communication channels to comply with the following fundamental requirements:

- MPLS-TP must support the full functionality (i.e. OAM, protection switching, operations, and configuration) without relying on the existence of a control plane, management plane, and without IP functionality (encapsulation, routing, and forwarding).
- OAM packets and the user traffic must share the same transport paths. In addition, it must be possible to differentiate between the different packets

Different communication channels are therefore defined to support the following functions at different nested levels: OAM functions, Coordination of protection state, Signaling (SCC) and Management (MCC).

To fulfill these requirements, MPLS-TP extends the mechanism that already exists in MPLS - the PW Associated Channel Header (ACH) that was defined for PW-VCCV [RFC5085]. The extension, the MPLS Generic Associated Channel (G-ACh), allows MPLS-TP to provide common in-band control channels also at the LSP and link levels.

G-ACh provides support for multiple channels between the end points of the transport path for different control channels. In order to differentiate the packets intended for the G-ACh, MPLS-TP designated a reserved label (13), GAL (G-ACh Label), that will appear in all messages. GAL always appears directly below the monitored transport path label and at the BoS. In addition, the packets include a G-ACh Header that includes a “Channel Type” field which indicates the type of control channel message that is being carried. This header appears directly after the GAL in all control packets

OAM

MPLS-TP provides a comprehensive set of OAM tools for fault management and performance monitoring, supporting the network and the services at different nested levels (i.e. at the end-to-end level, a segment of a path, and link level). The OAM tools may be used to monitor the network infrastructure, to enhance the general behavior, and performance level of the network. The tools may also be used to monitor the service level offered to the end customer, allowing verification of the SLA parameters, and enabling rapid response in the event of a failure or

service degradation. The OAM tools help reduce OPEX, minimizing the overhead of trouble shooting, and enhancing customer satisfaction which, in turn, helps to enable the delivery of high-margin premium services.
All OAM messages in MPLS-TP run in-band (using the G-ACh) and share their fate with data packets.

Architecture: MEPs and MIPs

The OAM architecture is described in the OAM Framework document. The main elements of the architecture are the Maintenance Entity Group (MEG), i.e. the portion of the transport path that is being monitored or maintained, the end-points of the MEG, known as the MEPs, and the intermediate points that can be monitored, the MIPs. The OAM messages can be transmitted between the MEPs of a MEG or from a MEP to one of the MIPs (dependent upon the nature of the OAM message). MIPs, however, do not initiate any OAM message, but may send responses.

A side-effect of this last rule (i.e. that MIPs do not initiate OAM messages) is the reason for the specification of an additional OAM construct – the Path Segment Tunnel (PST). This construct allows the system to perform OAM functionality on a segment of a transport path. By configuring the PST at a segment, it allows configuring the end-points of the segment to be the MEPs of the PST (originally they are MIPs of the end-to-end transport path) and thereby allows them to generate the needed OAM messages.

OAM Tools

MPLS-TP supports the following OAM functions:

- Continuity Check and Connectivity Verification: used to detect loss of continuity and misconnections between endpoints. To support pro-active Continuity Check and Connectivity Verification, BFD is being extended to support the requisite behavior. To support on-demand connectivity verification, LSP-Ping is being extended to comply with the requirements.
- RDI: Used by an endpoint to notify its peer that a defect has been detected on a bidirectional connection between them. BFD is being extended to include the RDI function within the Continuity Check and Connectivity Verification messages.
- Route Trace: Used to determine the route of a transport path across the MPLS-TP network. LSP-Ping is extended to comply with the requirements. A new message is being defined for this functionality.
- Data-plane loopback: this is an out-of-service message that allows testing of reachability to different points on the transport path. The intermediate points may support more than one testing point for this functionality. A new OAM message is being defined for this functionality.
- Packet Delay Measurement: used to measure the one-way or two-way delay of packet transmission between a pair of endpoints. A new message is being defined and the functionality will be based on the functionality of Y.1731. Note that a certain degree of synchronization of the time clocks of the two-ends of the transport path is needed.
- Packet Loss Measurement: Used to measure the packet-loss ratio between a pair of endpoints. A new message is being defined and the functionality will be based on the functionality of Y.1731.
- Diagnostic Tests: Used between endpoints to verify bandwidth throughput, bit errors, etc. The functionality is still under study. A new message will be defined.
- Client Fault Indication: Used to propagate a Client Fault indication to the far-end peer when alarm suppression in the client layer is not supported. A new message is defined in the PWE3 WG and will be transmitted via ACH.
- Alarm Reporting: ensures scalable operations! A server-layer (e.g. link) notifies the endpoints of its clients of a failure condition in order to allow alarms that may have been generated, as a result of this failure condition, to be suppressed. For data-plane implementation, a new message is defined. The function is supported at the control-plane and management-plane as well.
- Lock Reporting: similar to the Alarm Reporting function, this is used to suppress alarms at the MPLS-TP layer when a server-layer has been administratively locked by the operator. A new OAM message is being defined based on the functionality of Y.1731
- Lock Instruct: used by an endpoint to inform its peer endpoint that the operator has issued a lock on the MEG. A new OAM message is being defined based on the functionality of Y.1731.

Survivability

MPLS-TP supports enhanced protection switching and restoration mechanisms at different nested levels, to recover from “failed” or “degraded” transport entities and to support the most demanding services (sub-50 ms protection switching).
MPLS-TP supports different elements of control that may trigger a recovery action, these include in-band OAM fault management or performance monitoring indicating a defect or degradation condition in the network, or a

indication of a physical network failure (e.g. loss of light), or manual control by the operator, or control-plane signaling (when control-plane is active).

MPLS-TP supports multiple levels of recovery. The level of recovery directly affects the service level provided to the end user in the event of a network failure (in terms of the amount of lost data and the recovery time), and there is a correlation between the level of recovery provided and the cost to the network. The different levels supported are-

- Dedicated recovery where dedicated resources are allocated to protect a working entity.
- Shared protection where the resources for the recovery entities of several services are shared. These may be shared as 1:n or m:n, etc.
- Restoration and repair where the resources are allocated only when a defect or degradation is detected.

All existing GMPLS and MPLS recovery mechanisms are applicable to MPLS-TP. It should also be possible to provision and manage the related protection entities and functions defined in MPLS and GMPLS using the management plane.

MPLS-TP supports both general mechanisms, e.g. alternate paths or bypass tunnels, and also optimized mechanisms for specific topologies, such as mesh or ring networks.

Linear protection provides rapid and simple protection switching in a mesh network. It provides a clear indication of the protection status and supports an in-band protocol to coordinate the protection state between the edges of the protection domain when needed (e.g. bi-directional protection switching).

Some Service Providers have expressed great interest in the operation of MPLS-TP in ring topologies; they require a high degree of survivability functionality in these topologies. A study, presented to the IETF, indicated that existing MPLS-TP mechanisms (including linear protection) can be operated to optimally comply with the criteria specified in the MPLS-TP requirement document [RFC5654].

Provisioning and Management

MPLS-TP supports fast and automated provisioning and management of transport paths and services that traverse either a single domain or multiple domains, preserving the look and feel to which service providers have become accustomed.

Both static OSS based and dynamic control-plane based provisioning and management are supported. Full operation of MPLS-TP networks is possible independent of the way the network is configured and managed.

The control-plane of choice for MPLS-TP is GMPLS, the unified control plane used for core tunneling technologies (such as optical switches, TDM, Ethernet, MPLS, etc.). GMPLS supports fast, efficient, dynamic, and reliable service provisioning in multi-layer and multi-technology environments.

GMPLS supports mechanisms which –

- Provide Traffic Engineering, constraint-based routing and explicit path control.
- Address QoS and performance requirements (such as throughput, delay, packet loss, etc.), while utilizing network resources efficiently and reliably.
- Provide a comprehensive set of mechanisms for protection and fast restoration, as well as graceful operations.
- Partitioning of the managed network into separate peer or hierarchical control domains. It supports logical or physical separation of the control and management planes from the data plane.

Note that GMPLS supports in-service transfer of the ownership between control and management planes. That is, if a network is provisioned using an OSS-based solution and then at a later stage, a control-plane is deployed, it is possible to transfer, in-service, the ownership of the already provisioned transport path to the control-plane.

Standardization Status

Work on the standardization is continuing in the IETF, with periodic review at ITU-T. The timetable is very aggressive, especially for IETF timetables, but is progressing close to schedule. Currently –

- Requirements documents, i.e. TP-Requirements, Network Management Requirements, and OAM Requirements, are waiting for final approval as RFC, or, in the case of TP-Requirements, published as an RFC.
- Framework documents, i.e. TP-Framework, Network Management Framework, OAM Framework, Survivability Framework, Control-plane Framework, are in final review by the WG and ITU-T prior to entering the RFC process. Additional architecture documents, i.e. data-plane and point-to-multipoint, are planned for Q310 approval.
- Associated Channel documents – basic structure and the associated TLV structure – are both published RFCs.

- OAM tools documents are currently beginning their journey through the process, with three documents accepted as WG drafts and numerous individual drafts. The scheduled timeframe for their approval to enter the RFC queue is Q111.
- Survivability mechanisms, i.e. Linear Protection, Ring Protection, and Mesh Protection, are also starting their journey with a WG draft for linear protection and several individual drafts proposed for ring protection. The timeframe for these aims for early approval of Linear Protection in Q111and full toolset approval in Q411.

In addition to the specification effort, the ITU-T has started modeling MPLS-TP to allow it to be integrated into current transport equipment and networks. The work in the ITU-T references the work done in the IETF.

Conclusion

After years of debate on L2 vs. L3, on the technology of choice (VLAN-XC, PBB-TE, T-MPLS, L2-MPLS), it is possible to state that Carrier Ethernet is happening and that there is a very positive ecosystem surrounding MPLS-TP, the technology of choice. Many tier-1 service providers, suppliers and chip vendors are taking an active role in the standardization effort of MPLS-TP. Service Providers have begun asking for MPLS-TP and many RFIs already include a request for MPLS-TP implementation.

Shortest Path Bridging: A Novel Control Plane for Ethernet

David Allan, Peter Ashwood-Smith, Nigel Bragg, János Farkas, Don Fedyk, Mick Seaman, Paul Unbehagen

Introduction

The traditional Ethernet Bridging and data link layer has served well for several decades now and forms the foundation for many protocols including the IP/MPLS suite. Furthermore, as a service layer, Ethernet provides end to end connectivity complemented by robust and fully featured OAM capabilities. What modern Ethernet deployments have been missing is a more robust control plane that is adaptable to topology changes, easy to use, aware of the full topology, and capable of using multiple equal paths. The IEEE 802.1aq Shortest Path Bridging protocol provides all of these capabilities for both traditional 802.1ad (Q-in-Q) and 802.1ah (MAC-in-MAC) encapsulation methods.

Operationally, traditional Ethernet's model has enabled a "plug-and-play" environment, which is highly valued by consumers and operators alike. This desire to maintain simplicity also led to relatively simple control planes in the form of the Spanning Tree Protocol (STP), and the later Rapid STP and Multiple STP. These protocols preserved the Ethernet data link properties by maintaining a spanning tree as the active topology with near zero configuration.

One major shortcoming of spanning trees is that they do not utilize all links in certain topologies. The resulting connectivity is less efficient for traffic that is not from or to the root of the spanning tree. Furthermore, the protocols themselves are distance vector based and use transactional exchange in lieu of a database, which may extend the convergence time of an Ethernet network. To date, this has constrained the scale and utility of pure Ethernet networks. To deploy and maintain STP, MSTP or RSTP domains at significant scale requires an in-depth knowledge of the protocols and painstaking management, which moves away from the "plug-and-play" operational model. The significant increase in data plane scaling offered by 802.1ah Provider Backbone Bridging (PBB) [4] has pretty much mandated the development and specification of a new control protocol to address these issues.

The interaction of any new Ethernet control protocol with the data plane must maintain all of the key architectural properties of Ethernet specified by IEEE 802.1, and minimize changes to the technology base so that the huge body of existing work and implementations can be leveraged. A specific key requirement was to use minimum cost paths within an arbitrary mesh and not to be limited to a single shortest path between any two points.

The 802.1aq Shortest Path Bridging (SPB) [1] control plane managing the B-MAC layer of the PBB encapsulation is ideally suited to building layer 2 VPNs, and truly extends the scalability of Ethernet several orders of magnitude. This is important for any service that must offer Virtual LAN segments (the MEF E-LAN [7] service); in addition to its obvious applicability in campus backbones, another use is the direct support of any-to-any (IP subnet) connectivity in LTE wireless backhaul deployments. It was also required to support the VLAN and Provider Bridge (PB) data path albeit at smaller scales, so that existing and inexpensive ASICs could also provide shortest path functionality.

A link state protocol is the state of the art in distributed routing. SPB uses the ISO standard Intermediate System to Intermediate System (IS-IS) routing protocol [3]. This protocol only requires modest extensions to support SPB and is referred to as ISIS-SPB. IS-IS has a long track record of large scale deployment with robust behavior, and it was felt that 1000 node or larger pure Ethernet networks were not out of the question. Early small pre-standard live deployments (see [6]) and much larger emulations bear this out.

SPB is specified as an amendment to IEEE 802.1Q [2] standard and meets the backwards compatibility requirements inherent to such amendments, and hence SPB is able to interoperate with MSTP, RSTP and STP. Furthermore, SPB is equipped with the OAM specified by 802.1ag Connectivity Fault Management (CFM) [5].

Shortest Path Bridging Principles

Shortest Path Bridging (SPB) [1] provides frame forwarding on Shortest Path Trees. The mechanism for tree construction is distributed computation, achieved by running ISIS-SPB on all bridges in a network domain, known as an SPT Region. An SPT Region is seen as a single bridge from outside the region. ISIS-SPB uses the standard IS-IS procedures to construct and update the link state database in each SPT Bridge. SPB extensions to IS-IS are minimal and include only the parameters (TLVs) that are strictly necessary for the link state control of a bridged network.

ISIS-SPB sets up and maintains at least one SPT for each bridge to every other bridge in an SPT Region. Each SPT Bridge roots at least one SPT and an SPT Bridge only sends frames on one of its own SPTs. Thus, SPB implements source rooted trees, which is favorable for efficient multicast forwarding and applications using it, for example broadcast IPTV.

The SPTs have to meet two congruency criteria. Forward and reverse paths must be the same between any two bridge pairs, and unicast and multicast paths have to be congruent. This is key for the support of MAC address learning, OAM fate sharing, and preserving packet ordering guarantees. Therefore, SPT Bridges have to implement an order independent (deterministic) tie-breaking extension to the algorithms used for shortest path computation. Figure 1 (below) shows this result visually for two trees (A and B). Varying the tie-breaking rules enables the generation of multiple Equal Cost Trees (ECTs) in order to implement load balancing, which is discussed later.

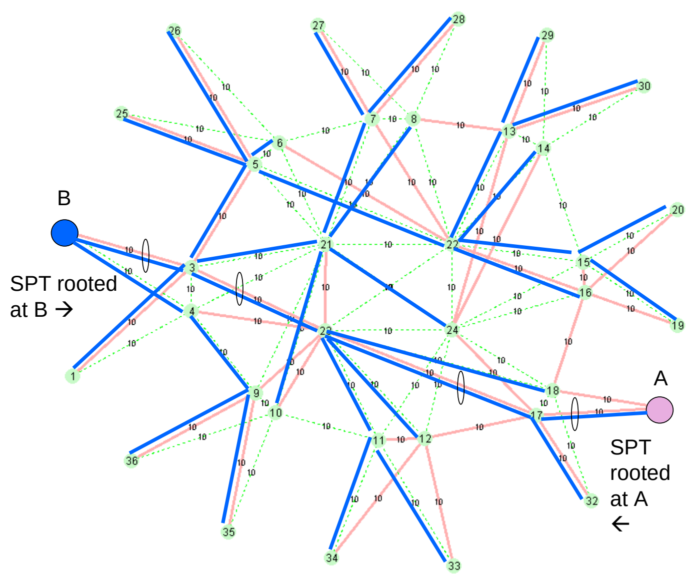

Figure 1 The symmetry of two Shortest Path Trees A and B: the path between A and B is the same in both trees

Loop avoidance[1] is an essential criterion of an Ethernet network in order to avoid the multiplication of looped broadcast and multicast frames. 802.1aq incorporates both loop prevention and loop mitigation techniques. ISIS-SPB implements a loop prevention mechanism for multicast forwarding. Neighbor bridges exchange digests of the topology database to check whether they have the same view of the physical topology, and by inference agreement on the distance to all SPT roots. An SPT Bridge removes multicast state when the distance relative to the root of that state has crossed a threshold beyond which looping is theoretically possible. The SPT Bridge only re-installs such affected state once its peer digests re-match, and hence agreement on the distance to all roots via that peer has been re-established. Trees on which the local distance to their root is unaffected by a topology change see no interruption. Loop avoidance is re-enforced with ingress filtering. Ingress filtering mitigates loops by auditing the port of arrival

[1] From time to time the use of TTL for loop mitigation has been suggested. IP uses TTL for loop mitigation, and for other diagnostics like trace route. Of all the techniques for loop mitigation, TTL is one of the simplest but not one of the best. The IEEE considered TTL and other techniques when specifying SPB but in the end extended the Loop prevention model of Spanning tree, by modifying the operation of the FDB when controlled by ISIS-SPB, and by adding Ingress checking. It is important to understand that while IS-IS behaves exactly as per specification, population of Multicast forwarding is gated by neighbor synchronization. Ingress checking is a stringent loop mitigation mechanism suitable for unicast and multicast which checks that a frame is received on a valid SPT from the source, by ensuring that the port of arrival is on the shortest path route to the source in the reverse direction.

of a frame, to ensure that it arrives on the port from the source SPT, and this allows relaxation of loop prevention for unicast traffic.

SPB takes advantage of the auto discovery mechanisms built into IS-IS, extending this to include service membership discovery. For this, the VIDs or PBB Service Identifiers (I-SIDs) configured on SPT Bridges are carried in ISIS-SPB TLVs. Thus, ISIS-SPB can configure frame forwarding specific to the registered services without extra discovery mechanisms. This innovation eliminates the need for signaling, and the delays associated with sequential hop-by-hop messaging,

SPB has two operating modes, distinguished by the applied SPT identification. Shortest Path Bridging VID (SPBV) uses a VID per bridge to identify each SPT, and this collective set of SPVIDs supports a VLAN. Within that VLAN, conventional "reverse path learning" is used to optimize unicast forwarding. A distinguished VID, the Base VID, is used to identify the VLAN in management operations in SPBV, and in Shortest Path Bridging MAC (SPBM) mode too.

In contrast a Backbone MAC address is used as the identifier of an SPT in SPBM, which uses the 802.1ah encapsulation. The resultant scalability makes it appropriate for larger fully managed networks. SPBM is used in a PBB network where all the B-MAC addresses are managed, and so MAC address learning from data frames is turned off. Otherwise, both modes operate along the same SPB principles and network can support either mode or both simultaneously since they are separated by Base VID. The operation of SPB will be described more in detail in [7] and other publications in addition to the standard.

Use of Equal Cost Paths

Shortest path forwarding results in better utilization of a mesh network. Even greater utilization can be obtained by the spreading of offered load across multiple diverse equal cost shortest paths. 802.1aq achieves this via manipulation of the criteria for selecting between equal cost paths and leveraging the ability of the Ethernet data plane to instantiate different path permutations in different VIDs. This is particularly important for Campus and Data Centre applications, where requirements for capacity and robustness are achieved by spreading load over a broad, massively connected, hierarchy of switches, exemplified by the "Fat Tree" structure.

When presented with equal cost paths, the SPB congruency properties can only be achieved when all bridges make the same path choices for each path permutation, requiring independence of the computation order, and of the network position of the computation. Each bridge has a Bridge ID, which ensures uniqueness. A PATHID is specified as the lexicographically sorted list of the Bridge IDs which the path traverses. Thus, all nodes implementing the same logic choose the same path from the multiple options (for example the one having the lowest PATHID). By using a set of globally defined transformations of the Bridge ID prior to sorting, different paths are selected. Since Priority is numerically the most significant part of the Bridge ID, the masking of bridge priority results in a different sort order when different priorities are used. 802.1aq specifies 16 Bridge ID priority transforms and makes possible the application of further tie breaking methods to choose from the nearly infinite set of other possible permutations.

The initial 16 equal cost trees have unique properties. Firstly, the path congruence property means that 802.1aq actually supports equal cost routing for multicast and broadcast traffic. Secondly, since assignment of traffic to a path is done at the ingress to the 802.1aq network, the operator has the ability to forgo random assignment and place the traffic based on estimates of utilization, which can be considered a very lightweight form of traffic engineering. For example, all traffic, both unicast and multicast, for one subnet in a Data Center can be constrained to one set of shortest paths, while traffic for a different subnet can be constrained to a second.

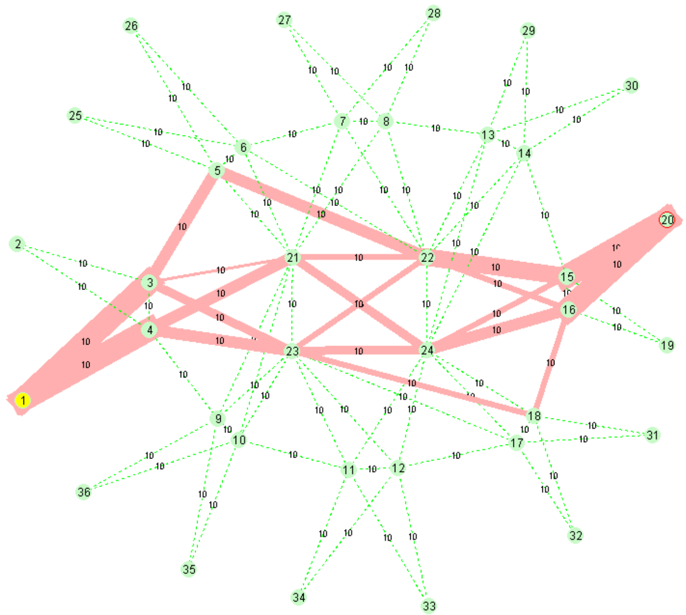

Figure 2 802.1aq Network Emulation view of all 16 Shortest Paths between a pair of nodes. Operator can select any of them per service.

In order to foster continued exploration in this area of 802.1aq traffic placement a framework has been created to allow future extensions to the Equal Cost modes. Research is already demonstrating techniques for even greater mesh utilization and controllability.

Figure 2 above, gives a concrete visual representation of 802.1aq's multi path routing features. Shown are all 16 individual shortest paths superimposed on the network between a given pair of nodes (1 and 20). The distribution compares favorably with the hop by hop hashed-based approaches of IP ECMP, however, unlike ECMP, 802.1aq permits per service assignment by the network operator. In addition to tuning load by adjusting service to route assignment, the operator can also 'tweak' the ECT behaviors giving a high degree of traffic control without resorting to full Traffic Engineering (TE) protocols.

In Conclusion

Over the past several years Ethernet has accreted significantly enhanced functionality, measurability and scalability, which have simply awaited the right control plane to be unleashed. The 802.1aq project is defining that control plane, which will bring unprecedented scale, resiliency and efficiency to Ethernet networking. In essence 802.1aq is the application of sophisticated routing algorithms to commodity technology. The result will set new benchmarks in price, performance and operational simplicity.

References

[1] IEEE Std. 802.1aq D3.0, "IEEE Draft Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks – Amendment 9: Shortest Path Bridging," June 2010.

[2] IEEE Std. 802.1Q, "IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks," 2005.

[3] ISO/IEC 10589, Information Technology – Telecommunications and Information Exchange Between Systems – Intermediate System to Intermediate System Intra-Domain Routing Information Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless- Mode Network Service (ISO 8473)," 2nd ed., 2002.

[4] IEEE Std. 802.1ah, "IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks – Amendment 6: Provider Backbone Bridging," November 2007.

[5] IEEE Std. 802.1ag, "IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks – Amendment 5: Connectivity Fault Management," December 2007.

[6] D. Allan, P. Ashwood-Smith, N. Bragg and D. Fedyk, "Provider Link State Bridging," IEEE Communications Magazine, September 2008.

[7] D. Allan, P. Ashwood-Smith, N. Bragg , J. Farkas, D. Fedyk, M.Ouellete, M. Seaman, P. Unbehagen " Shortest Path Bridging: Efficient Control of Larger Ethernet Networks," IEEE Communications Magazine, October 2010.

[8] Metro Ethernet Forum, MEF D00057_006, Ethernet Services Definitions - Phase 2, Approved Draft 6, February 2008

DAVID ALLAN (david.i.allan@ericsson.com) is Sr. program manager for standards and strategy at Ericsson, and a former distinguished member of technical staff at Nortel. He has been active in data telecommunications standards for the past 15 years including WG chair roles in the Broadband Forum and IETF. He has been active for over 25 years as an architect, design engineer, and developer of real-time systems in diverse areas of technology ranging from process control and avionics to financial transaction processing. His current role at Ericsson is focused on carrier infrastructure based on MPLS and Ethernet. He has a B.Eng. (1978) from Carleton University in Ottawa.

PETER ASHWOOD-SMITH (peter.ashwoodsmith@huawei.com) is a Huawei Fellow with B.S and M.S. degrees in computer science from the University of Toronto. He has worked on the design, standardization, implementation, deployment, and support of many modern routing protocols (e.g., early label switched networks including ATM, MPLS, and GMPLS). He has approximately 30 networking related patents. His interests and research are now tending in the direction of computation-based/non-label-switched protocols.

NIGEL BRAGG (nbragg@ciena.com) has degrees from Trinity College, Cambridge, and Southampton University. He has spent 20 years in the telecommunications industry, including 13 years with Nortel. For the last seven years he has focused on carrier packet transport, and is a co-inventor of PBT and PLSB. He has recently joined Ciena.

JÁNOS FARKAS (janos.farkas@ericsson.com) is a research engineer at Ericsson Research since 1999. He is an active contributor to the IEEE 802.1aq Shortest Path Bridging project. He has focused on carrier transport networks based on Ethernet in the past few years. His former research activities include IP QoS solutions for mobile access networks and network traffic management. He has 18 patent applications and 12 research papers in the area of telecommunications networks. He has a Master's degree in electrical engineering from the Budapest University of Technology and Economics

DON FEDYK (Donald.Fedyk@alcatel-lucent.com) Don is currently the Editor of the Shortest Path Bridging project (IEEE 802.1aq). Don is an inventor on approximately 20 communications related patents and holds B.S. and M.S. degrees in Electrical Engineering from the University of Waterloo, Ontario, Canada.

MICK SEAMAN (mickseaman@sbcglobal.net) is an independent consultant working on protocols and operational methods for Ethernet deployment in public and mission critical networks. Mick has led the design, implementation, and standardization of routing and bridging switches since the mid 1980's – at DEC, 3Com (VP & CTO Enterprise Systems Business Unit, 1996-99), and Telseon (Founder, EVP & CTO) – and holds more than 20 patents in the field of switching system and protocol design. Mick chaired the IEEE 802.1 Interworking (1986–2007) and Security (2004–2010) Task Groups and was responsible for the IEEE 802.1D, 802.1Q, 802.1X, and 802.1AE standards. Mick has a Master's Degree in Physics from Cambridge University.

PAUL UNBEHAGEN (paul.unbehagen@alcatel-lucent.com) is an active contributor to the IEEE 802.1 WG and the IETF. He has participated in several IETF WGs to include IS-IS, BGP, L2VPN, and IPVPNs and is currently the editor of the IP/SPB IETF draft. He has numerous patents in communication protocols such as IS-IS and BGP. Paul has also had a diverse background working in a vendor but also on live networks at a Tier 1 Service Provider, an ASP, U.S. Military as well as a few startups. Paul thus has 15 years of deployment, operational, network design and architectural experience in live networks ranging from Enterprise to Carrier. Paul has a B.S in CIS and a B.S. in BA from NC Wesleyan College.

Offloading Excess IP Traffic with Optical Bypass – A Simple Capacity Upgrade, or More?

Mohit Chamania, Admela Jukan, TU Braunschweig, Germany
Oscar Gonzales de Dios, Javier Jimenez Chico, Telefonica, Spain

The coming years are bound to a huge traffic growth expected in the metro and the backbone networks due to the widespread deployment of fixed and mobile broadband access networks coupled with growth in new services, such as HD and ultra HD video. Given this scenario, the network operators are considerably concerned about the scalability of the IP based network architectures. The cost of the network depends on the line rates used in the infrastructure, and a continuous data increase (and consequently cost) coupled with the traditional flat-rate pricing models for end-users would soon affect operator margins, which could eventually impact the end-users in the form of an Internet access prices increase. In order to drive down network costs, a dynamically reconfigurable optical network presents an excellent alternative due to the fact that the cost of optical switching is independent of the traffic rate of the transported signals. However, as optical technologies currently lack the flexibility provided by IP routers in handling traffic, it is clear that the next generation of high speed networks will not be all-optical and will, in all likelihood, will follow a “IP-over-optical” networking paradigm.

While various IP-over-optical architectures have been extensively studied in research, the real practice lags far behind. Instead of fancy multi-layer optimization solutions for IP networks, broadly addressed in the research community under the umbrella of Multi-layer Traffic and Network Engineering, network providers prefer extreme over-provisioning paradigms with up to 70% headroom in IP links. In a typical carrier network, whenever the IP traffic exceeds a nominal threshold, the capacity of IP links is upgraded so that the link utilization at peak network loads is limited to about 30%-40% of the link capacity or less. Given that traffic increases steadily with time, frequent network planning exercises (approx. every six months) upgrade the port sizes in the IP network to make it future-proof, and the granular nature of available capacities of these ports (1G , 10G , 40G) ensure that IP links have significant headroom. These practices are targeted to facilitate smooth and stable operation in networks, which is the core objective for operation in commercial networks. However, it is clear that the volume of traffic in the future will render traditional over-provisioning practices ineffective, and new paradigms are required to support IP traffic offloading of peak traffic.

When designing multi-layer traffic engineering and network engineering methods for efficient operation in carrier networks, what most research proposals do not consider is the existence of a legacy IP/MPLS infrastructure which is the fundamental enabler for almost all of the operator’s services including fixed internet, triple play, mobile internet and L2/L3 corporate services. Frequent changes in the IP/MPLS topology and operation are not desired as they not only impact the schedule and frequency of network planning exercises in the legacy network, but also trigger significant reconfigurations in legacy management systems. These are the important reasons why IP link capacity is exclusively upgraded by simply added the capacity to the existing link. Other reasons are perhaps rather operational, and go back to the traditional organizational separation of “IP” and “transport networks” telco’s business units.

So what is the right way to go when scaling up the capacity of the IP network in the future? In other words, what is the evolutionary “IP-over-optical” approach for excess IP traffic offloading, given the requirements on stability of legacy network systems?. To maintain the stability of the legacy system, we propose *optical bypasses* to be used in the IP layer which are hidden from the legacy IP/MPLS routing services. By optical, we mean any underlying transport technology capable of setting up circuits between a pair of IP routers based on optical WDM transmission, such as carrier grade Ethernet or OTN. Bypasses are different from “new IP links” which would be created in a virtual topology design and network planning phase, because they do not affect the legacy IP routing although they are dynamically setup to offload the IP traffic. In this architecture, dynamic circuits are temporarily used to bypass traffic across congested links or routers in the network.

But, how this can be realized? Where do we setup a bypass in the network? How do we simply “add a new port” on a router? We do not have answers to all the questions but we can speculate.

First and foremost, these optical circuits must ensure that the IP network routing remains stable, which means although they are setup and used by the IP flows, they are not advertised in the IP layer, and effectively “hidden” from the IP routing service. The packets redirected to these bypasses must not experience any change in routing, and should remain fully unaware of the existence of the bypass. Therefore, the bypass has to be established between a pair of routers along the original route. For instance if the original route of a packet flow is Frankfurt-Barcelona-Madrid, the bypass can be setup between Frankfurt and Madrid, so when the packets arrive in Madrid they remain unaware of their past hops, as IP packets are routed based on destinations. Hence, if the optical circuits are to keep the IP routing stable, they have to be constrained to bypass traffic between IP routers along the original routing path from a router upstream of the congested site to a router downstream.

Second, by not advertising optical bypasses as new links in the IP network, we ensure that IP/MPLS routing re-convergence is not required, and resulting large scale routing modifications are avoided. In this way, we also ensure that no other segments in the network observe increase in traffic thus guaranteeing temporal network stability. In order to re-route flows onto optical bypasses, only minor forwarding modifications are required at the source router, which can either be facilitated by tagging appropriate traffic flows at the IP layer using VLANS, or by using SDN based switches such as OpenFlow to divert the traffic onto an optical bypass. However, how the flows are identified to be diverted onto bypasses and also whether the setup of a circuit used for IP traffic outside from the IP routing service will receive a wide acceptance in the IP network community, remains an open question.

We believe that the usage of optical bypass for IP traffic offloading when applied in conjunction with technologies where circuit capacity on the port can be changed on demand, the overall performance is expected to be close to optimal, especially when the traffic demand is predictable. Example of such a technology is carrier-grade Ethernet, where a high capacity interface of 100Gb/s can be dynamically partitioned into circuit and packet switched ports. With the proposed approach, existing IP/MPLS infrastructure is not substituted by other switching nodes (e.g. OTN, MPLS-TP, etc), but it is complemented by them. This approach ensures a stable operator service provisioning and the scalability of IP network. In this way, potential IP scalability problems generated by full meshed IP topologies with hundreds or thousands of adjacencies are gracefully avoided. In particular, by balancing the network load between IP/MPLS and a dynamically switched optical network, the proposed architecture does not eliminate current IP/MPLS routers while significantly reducing the required investments for new transit network resources.

From the network architecture perspective, the optical bypass technique, when applied in the access and metro networks, exploits the hierarchical architecture of the legacy network, where multiple operator`s POPs (Point of Presence) are connected to a transit IP router, and similarly multiple transit routers are then in-turn connected to the core IP router. Even with the significant rise in P2P traffic, a large percentage of traffic from the POP is destined to and beyond the core router. Eliminating transit nodes in such a scenario is also not efficient as it leads to inefficient traffic grooming onto a circuit. Here, when large traffic volumes are observed at POPs, an optical bypass can be established from the access POP to the core router, thus relieving load on the transit router. Note here that traffic destined from the said access POP to other POPs belonging to the same parent transit router are not bypassed thus ensuring load on the core router-transit router link does not increase. In summary, the application of this technique can significantly reduce the IP processing capability required at the IP transit nodes, but it solves the potential scalability and stability concerns at the IP control plane that other offloading approaches often neglect.

Mohit Chamania is currently pursuing his PhD at the Technische Universität Carolo-Wilhelmina zu Braunschweig, Germany. His research interests include inter-domain routing, Ethernet and Optical technologies.

Admela Jukan is a Chair Professor of Electrical and Computer Engineering at the Technische Universität Carolo-Wilhelmina zu Braunschweig. Her research interest is in optical networking, broadband wireless and wireline access networks, and network control and management platform design.

Óscar González de Dios joined Telefónica I+D in 2000. In 2005 he joined the Advanced Network Planning department in Telefónica I+D, where he has been working in the analysis and performance evaluation of optical networks. He has participated several R&D European projects, like IST NOBEL (I and II), e- Photon/One+, BONE and AGAVE. He has coordinated the CELTIC project BANITS 2. He has also worked in several internal projects for the Telefonica group regarding optical network planning and optical network analysis.

Javier Jiménez joined Telefónica I+D in 2003, being involved since then in the Network Management, Innovation Strategy and Core Networks and Protocols departments. He joined in 2007 the New Network technologies division, where he has been involved in Future Internet projects such as IST 4WARD or IST GEYSERS, and CELTIC 100GET.es project on 100Gbps Optical Carrier Ethernet, as well internal projects related to the integration and coordination of IP and Optical transport technologies.

About PRISM (*www.ontc-prism.org*)

Aim: To disseminate relevant content pertaining to optical networking and related growth areas across industry and academia. To promote the growth of optical networking activity by creation of a unified knowledge base. To create a communication bridge between industry and academia in terms of research frontiers and complementary strategies for future growth.

Scope: The optical networking community stands at a point where its potential is not fully realized. The bandwidth offered by the fiber at price points that currently prevail is a fantastic business case for Internet services for providers the world over. Optical networking has transcended itself from a point-to-point communication service to a WDM based multi-point granular networking hierarchy. This journey was made possible through successful and important innovations in the optics and networking domain, bringing together a rich technology set for deployment in telecommunication networks. It would be fair to say that without optical networking, the scope of the Internet would not reach its global scale that it has presently reached. In the future, optical networking has the potential to impact the telecom world through new innovations in architecture, protocol and devices that would lead to new service offerings impacting human lives. Amongst these futuristic offerings are cloud computing, energy efficient systems, data-centers, 100 Gigabit Ethernet, WDM PON, multi-point communication systems, sub-wavelength grooming and transparent ROADM-based services. It is clear, and especially pronounced in Asia and parts of Europe that optical networking will play a very important role in the design of future networks. Whether it is the GENI project in the US or the Akari in Japan – optical networking finds a clear way into technological offerings for the future of the telecommunication industry. From a historical perspective, optical networking has offered significantly to the telecom industry – we distinctly note that after the telecommunication bubble burst, it was the area of metropolitan networks that led to the re-bounce of telecommunications the world over. It is always important to highlight such historical perspectives from industry leaders and pioneers to bring the optical community closer. We continue to exploit the latest advances in this area of telecommunications – delving on the research and development of optical networking solutions.

The scope of the newsletter is as follows:

- A forum that brings the optical networking community together, through leadership articles in technology and research.
- Bring to the fore issues that both industry and academia are working on, with the focus of being able to minimize this gap through interaction via the newsletter forum.
- Highlight important events related to the area of optical networking, in particular focus on consortiums, projects, awards, seminal breakthroughs, standards and industry related information.
- Research: Focus on research issues pertaining to optical networking. Showcase key growth areas (like data centers, metro ROADMs, 100GE, etc.).
- Consortiums and Projects: Focus on consortiums and projects relevant to optical networking, in which the primary entities are research focused (non-profit groups like universities etc.)
- Developing Economies: Focus on emerging economies and the networks there.
- Standardization activity: The newsletter will periodically discuss standard related activities especially when new drafts are circulated or a standard in form of an MSA is accepted. A standard pioneer will be invited to write about the standard. Our focus will be on the IEEE 802 working group, the ITU groups and FSAN groups in terms of coverage.
- Industry information: latest *technical* happenings will be reported from the industry. These will be critically based on demonstrations at international tradeshows such as OFC, ECOC and World Broadband Forum. Care will be taken not to report any company specific information and ensure vendor neutrality in the newsletter.
- Service provider focus: Since a key consumption point of our industry are service providers, it is most important to focus a section of the newsletter on them. We will in every newsletter focus on the latest happenings in the provider space – whether it is adoption of new technologies or new deployments or even network designs, we will cover these through neutral writings. In particular, we will ensure that no names are taken in the coverage, making it generic – for example, "a select provider in the North America has decided to deploy ROADM technology using WSS cross-connects [source]..".
- Periodically create a roadmap of technologies in different domains pertaining to optical networking. The roadmap would be a team effort by multiple experts in association with the editor.
- Optical Networking is Fun (ONiF): a section devoted to humor in optical networking – puzzles, crosswords and "did you know" for after-hours research.

Submit your article as a .pdf file to submissions@ontc-prism.org. Note that you must have a covering note that descibes the nature of the article from one of the above scope keywords. **The scope keywords are: consortiums, projects, growth areas, emerging economies, Standardization activity, Industry information, Service provider focus, roadmap of technologies and Optical Networking is Fun.**

Note to prospective authors: ONTC Prism follows strict policies mandated by the IEEE Code of Ethics. We will strongly enforce plagiarism and self-plagiarism as a review criteria. For more information visit:http://www.ieee.org/web/publications/rights/ID_Plagiarism.html.

Editor:

Ashwin Gumaste,
James R. Isaac Chair,
Department of Computer Science and Engineering
Indian Institute of Technology Bombay,
Contact Information:
Room 208, Kanwal Rekhi Building, IIT Bombay, Powai, Mumbai, 400076
Email: ashwing@ieee.org, Web: www.ashwin.name.