



A Resilient Path Management for BGP/MPLS VPN

Jong T. Park

School of Electrical Eng. And Computer Science
Kyungpook National University
park@ee.knu.ac.kr

1

APNOMS03

Abstract

BGP/MPLS VPN is recently receiving much attention from industries and standard bodies. It enables ISP to provide IP-service with QoS guarantee to the customers over shared MPLS backbone. In this paper, we present a resilient path management mechanism for BGP/MPLS VPN. An existence condition for finding fast backup path which satisfies the resilience constraint is derived for a full mesh BGP/MPLS VPN. It is based on Hamiltonian cycle which is a minimal set of traversing links for a given graph. A simple path finding algorithm with $O(n)$ computational complexity is presented. Finally, we develop a decomposition theorem which enables the approach to be extended to the hierarchical MPLS backbone which is well scalable in a full mesh structure, using MPLS label stacking technique.



Introduction

- Convergence of IP with optical : Resilience becomes important to both customers and service providers
- Lower layer failures may generate hundreds of upper layer failures at MPLS hierarchy
- Need of automatic service provisioning mechanism which enables a minimal disruption of service, meeting the customer resilience requirement for BGP/MPLS VPN
- Propose a *resilient path management* mechanism which can *dynamically* configure the paths (LSPs in MPLS network) satisfying the TE resilience requirement from the customers.
- Specifically, we present (1) condition for existence of fast solution based on Hamiltonian cycle, (2) a simple resilient path management algorithm, and (3) decomposition theorem applicable to both intra and inter-domain full mesh BGP/MPLS VPN.

2

APNOMS03

- Recent advent of IP over optical convergence : Resilience becomes a critical issue and active work in standard bodies.
- Failures at lower layers may generate *hundreds of link or node failures* at higher layers of a MPLS network.
- BGP/MPLS VPN is configured in a *full mesh* or hub-spoke, and instrumented by establishing and maintaining LSPs of MPLS.
- It is necessary to provide a contracted service to the customers with minimal or no disruption of service in case of unexpected multiple failure occurrences.
- We propose a *resilient path management* mechanism which can *dynamically* configure the paths (LSPs in MPLS network) satisfying the TE resilience requirement from the customers.
- Specifically, we present (1) condition for existence of fast solution based on Hamiltonian cycle, (2) a simple resilient path management algorithm, and (3) decomposition theorem applicable to both intra and inter-domain full mesh BGP/MPLS VPN.



BGP/MPLS VPN Overview

- BGP/MPLS Virtual Private Network (VPN) enables service provider to provide IP-based VPN service to customers
 - Layer 3 VPN Solution standardized by IETF PPVPN WG (RFC2547bis)
 - MPLS is used to transmit VPN traffic and BGP is used to distribute the routing information across MPLS backbone
- Features of BGP/MPLS VPN
 - Supporting customer transparency to VPN service provisioning
 - Use of full MPLS TE capabilities supporting multiple QoS classes
 - Preserving of customers' IP address schemes
 - High scalability and security
- Proprietary MPLS VPN solutions such as Cisco's BGP/MPLS VPN, Nortel's MPLS-based Virtual Node, and Lucent's Virtual Node.
- Other VPN Solutions: L2TP, PPTP, IPsec, Virtual Leased Line, etc

Recently, the provisioning of virtual private networks (VPN) over the global Internet is gaining much popularity. VPN provides interconnections of customer sites over a shared network infrastructure. Traditionally, VPNs have been mostly provided by the leased lines, but the development of new technology such as ATM and Multi-Protocol Label Switching (MPLS) enables the service providers to look for the better cost-effective solutions in terms of scalability, security and quality of service. The provisioning of VPN over BGP/MPLS among different Autonomous Systems has been being standardized by IETF RFC 2547bis, and several vendors are already providing proprietary solutions such as Cisco's BGP/MPLS VPN, Nortel's MPLS-based Virtual Node, and Lucent's Virtual Node.

In BGP/MPLS VPN, the VPN routing information is distributed by MP-BGP over the service provider's backbone network, and MPLS is used as an underlying network infrastructure device to forward the VPN traffic among the participating VPN sites. In BGP/MPLS VPN, the connection-less IP traffic of a VPN customer site is transparently transmitted through a provider's connection-oriented data paths, more specifically label switched paths (LSPs), of a MPLS backbone. The efficient design of these paths for BGP/MPLS VPN is an active research area. RSVP-TE or CR-LDP is recommended for setting up these paths for BGP/MPLS VPN traffic. Currently, in many cases, these paths are manually designed by network administrators.

As the size of VPN and the number of its constituent components become very large due to the rapid increase of customer sites, the importance of the automatic provisioning of the BGP/MPLS VPN will be more important to not only customer's satisfaction but also service provider's revenue generation. Furthermore, when components of MPLS backbone would fail due to some hardware and/or software errors, the VPN service may create a tremendous amount of damage to both customers and service providers. The detection of failed components, identification of the causes for errors, and rapid recovery of the failed components are indispensable procedures for the reliable VPN service, being constrained with service level agreement (SLA). The resilience of the VPN connection can be achieved by provisioning of so called Make-Before-Break backup paths in BGP/MPLS backbone, and rapid restoration of the failed components to sustain the guaranteed level of quality of service.



Related Work

	Protection	Rerouting (RR)
Recovery Time	Fast	Slow
Survivability	Low	High
Resource Usage	High	Low
Path Selection	Static	Dynamic
Related Standard Work	IETF RFC 3469 1:1, 1:N, M:N Protection	IETF RFC 3469 Make-Before-Break

- ✓ Recovery time in RR = Path Selection + Signaling + Resource Allocation
- ✓ Path Selection : Finding Alternate Paths in case of Multiple Failures, currently active research area in MPLS

5

APNOMS03

There has been many research work on MPLS recovery mechanism by IETF. The MPLS recovery mechanism is roughly classified into two techniques: protection switching and restoration/rerouting. In the protection switching, often called fast reroute, the alternative LSP paths are pre-provisioned to minimize the disruption of the service in case of network failure.

Protection switching is further classified into several ways; 1+1 and 1:1, 1:N, M:N, and split path protection. Each method can be applied to a link of LSP, a segment of LSP, called span, or an entire LSP path. In 1+1 approach, traffic is transmitted simultaneously on two links and at the receiving node, the best traffic is chosen to choose the best source. The 1:1 approach allows the traffic to be transmitted on the alternative backup path in case of the failure of the primary path. Protection LSP can be established either at end node or at intermediate node.

In MPLS restoration mechanism, the backup path is established after failures in a primary path occur, and then the data traffic is switched to the backup path. The required resource may be dynamically reserved, so that it usually take longer time than protection mechanism. The restoration can also be performed on a link, a node, path segments, or entire end-to-end path. Restoration mechanism is sometimes called a rerouting mechanism, and usually adopts the *make-before-break* principle. A backup path may be decided completely on-demand or a set of backup paths are decided in advance with pre-computation, and an optimal one is selected after failures occur. For each backup path, the required resource may be reserved at each node along the path using signaling but not be committed, for example, not to be cross-connected in optical cross-connect. In other words, the resources of the backup path are reserved only in the control plane lever, and the actual cross-connection can be performed after failures occur.



Problem Formulation

For a given BGP/MPLS VPN network consisting of a set of nodes, links and resilience constraints, establish and maintain the primary and backup paths such that the disruption of service is minimized for multiple component failures while satisfying the resilience constraints.

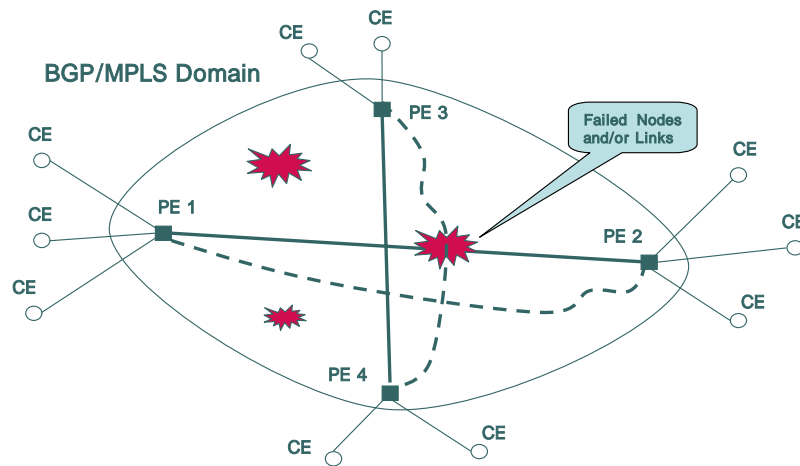
6

APNOMS03

In this paper, we propose a resilience-based approach to BGP/MPLS VPN management in which a resilience of a path is used as a criterion for the path selection of MPLS VPN LSPs. More specifically, a set of auxiliary backup paths are pre-provisioned with a primary path for VPN data traffic which can be used for rapid restoration and/or load distribution. For example, when a failure occurs in the MPLS backbone, the failed primary path can be rapidly switched back to the backup path so that the loss of packet can be minimized. The resilience of the VPN paths can be achieved by the construction of a backup path for every VPN connection in which the primary and backup paths may be using different nodes and links.

We find out conditions for the existence of backup path for a given resilience constraints, and present fast backup path design algorithms which could run in the order of $O(n)$ where n is the number of nodes. In most cases, the number of failures would be small, so that the algorithm would generate the feasible candidates for the backup paths very efficiently.

BGP/MPLS VPN with Failed Nodes and/or Links



7

APNOMS03

The figure shows the conceptual model for Intra-Domain BGP/MPLS backbone where there exist two set of resilient VPN paths: one from PE 1 to PE 2 and the other from PE 3 and PE 4. As shown in the figure, VPN service for each pair consists of two paths, i.e., a primary path and a backup path. Some failure of link and/or node is detected in the primary path from PE 1 to PE 2, and these failed components may affect the backup path from PE 3 to PE 4. This area containing the failed components which may be set of links and/or nodes are indicated by failed spot in the figure. We are more interested in finding out the properties of network component failures with regard to the path protection and restoration. Therefore, it is assumed that any node is not disconnected from the network due to the failures, i.e., each node having at least degree of 2 after components failures. We can also apply the technique to the rerouting approach without any pre-established backup paths.



Definition of Path Resilience

- ◆ Path resilience informally implies the recovering capability of a path without disruption of service in case of multiple failures.
- ◆ Definition : A path resilience in MPLS network is defined as a real-valued function such that

$$\text{path resilience} = \sum_{\text{ProtectionSet}} \frac{1}{m} \cdot \frac{\text{Number of Protected Components}}{\text{Total Number of Components}}$$

where m is the multiplicity factor of a primary path and *ProtectionSet* denotes the set of all the backup paths to protect the primary path

In MPLS, a set of attributes is defined to control LSPs, and among these, the resilience attribute is used to determine the behavior of LSPs when failures occur. A basic resilience attribute determines whether the failed LSP is to be rerouted when segments of its path fail. Extended resilience attributes are used to specify specific recovery mechanisms and policies to govern the relative preference of each specified backup path. The user service level with regard to reliability may be mapped to these resilience attributes of LSPs. We present a simple model to represent this resilience attribute for MPLS recovery mechanism below. A path resilience in MPLS network is defined as a real-valued function as expressed above. The multiplicity factor m of a path defines the number of total primary paths which are sharing a backup path, or a segment. It is used to represent the sharing of backup path in MPLS recovery mechanism. Number of Components implies the total number of components in a path, without including the end-nodes of GMPLS network since they are always shared among primary and backup paths. The Total Number of Protected Components implies the total number of components in the path which are protected by backup paths. The path resilience is defined to be zero if there is no backup path, or all the components of the backup path are shared with the primary path. For the brevity of expression, we simply call resilience instead of path resilience. The resilience model enables service providers or network operators to automatically support a range of different service levels to optimize their service revenue with respect to available network resources.



Condition for the Existence of Path with Resilience 1

For a BGP/MPLS backbone with N nodes in full mesh structure, where $N \geq 3$, there exists a path with resilience 1 between any pairs of PE even though any $(N-3)$ links or nodes or together between PEs fail where the failure of a node implies the removal of the node and all the links emanating from it.

9

APNOMS03

Definition : A graph G is said to be k -component Hamiltonian if the removal of the any j components,

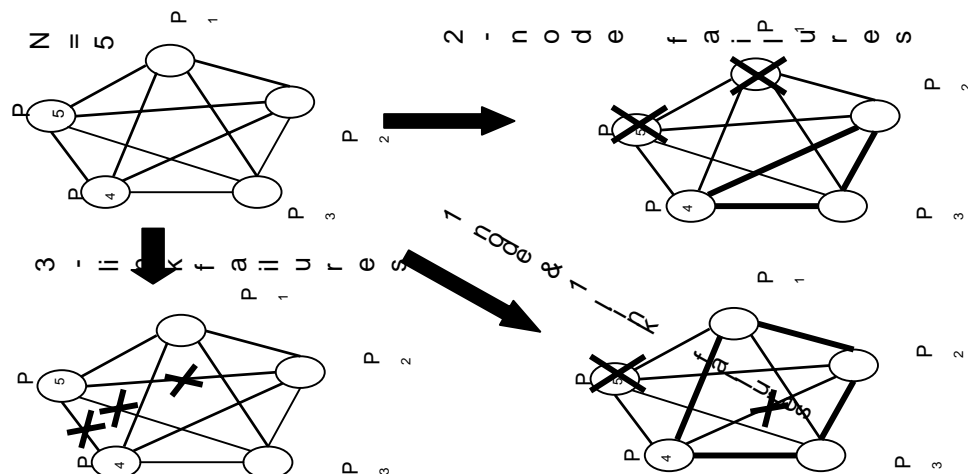
$0 \leq j \leq k$, from G leaves a Hamiltonian graph.

- Here, a component is either node or link. If a component is a node, a removal of the component implies the removal of the node and all the links emanating from the node.

Theorem 1: A complete graph with p nodes, where $p \geq 3$, is a minimum $(p-3)$ -component Hamiltonian graph.

Proof) We simply illustrate the concept graphically.

(A formal proof is rather lengthy).





Resilient Path Management for BGP/MPLS VPN

Procedure Dynamic_Path_Mangement (Failure_Notification);

Step (1) If failure notification is not related to paths from PE,
Then return ("Irrelevant Failure Notification");

Step (2) If a primary path is damaged due to the
Failure_Notification and a backup path is available
satisfying the resilient constraint,
Then switch the VPN data traffic to the backup path;

Else construct the backup path satisfying the resilient
constraint and reroute the traffic to backup path ;

Step (3) As the components in the primary path are repaired,
revert to the primary bath;

10

APNOMS03

The construction of backup path is accomplished by the Algorithm:

Algorithm Construct_Backup_Path (Failure_Notification)

Begin

If (Resilience_ = 1) and (Number of Failure_Notification (N - 3))

Then { Pointer := 0; BackupPath[Pointer] := ingress_PE;

Do {Select a node R such that R does not belong to both
Path and BackupPath, and there exists a link
<BackupPath[Pointer], R>;

BackupPath[Pointer + 1] := R ;

Pointer := Pointer + 1;

UNTIL (R = egress_PE or All nodes which are not
belonging to Path is covered);}

}

End

Note that the backup path construction algorithm run in the order of $O(n)$ where n is the number of nodes

Decomposition Theorem

For a BGP/MPLS network where the collection of nodes is decomposed into sets of AS domains, assume that each AS domain is structured as a full mesh, and all the ASs are fully connected to each other via a PG. Then, there is a path with resilience δ between any pair of PEs such that

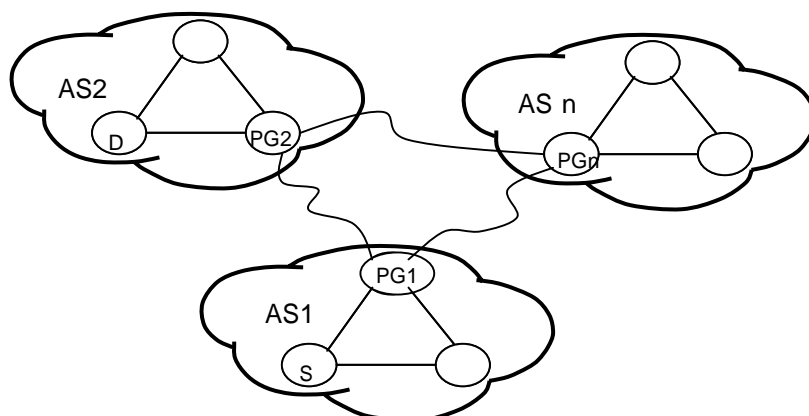
$$= \begin{cases} 1 & \text{PGS are used as sources and destinations} \\ (\delta - 2) / \delta & \text{Otherwise} \end{cases}$$

where δ is equal to the number of the components in the path. A gateway (PG) for a given AS takes care of external connections to all the other AS domains.

11

APNOMS03

Proof: Let the BGP/MPLS network to be modeled as a graph G. We show that there is always a backup path for a given primary path from any node S in G to any other node D in G such that there exist two gateway nodes PG1 and PG2 which are shared in the primary and backup paths. We proof by contradiction. First, suppose that there is no backup path from a node S in G to a node D in G. Suppose that S and D belong to the identical AS domain. Then, we can always construct a separate path from S to D by Theorem 1 since AS is fully connected. So, it leads to the contradiction. Secondly, suppose that S and D belong to different AS domains such that S belongs to AS1 and D belongs to AS2 as shown in the figure. Since a backup path can always be constructed from PG1 to PG2 and a different sub-backup path can be constructed from S to PG1 and from PG2 to D, there always exists a separate backup path from S to D in which PG1 and PG2 are the only shared nodes. This also leads to the contradiction. Since PG1 and PG2 are always shared in the primary and backup paths, the resilience δ is equal to $(\delta - 2) / \delta$ where δ is equal to the number of the components in the path. For the case with S and D are identical to PE1 and PE2, respectively, resilience become 1. This completes the proof.





Conclusions

- BGP/MPLS VPN is a promising solution to service provider, which supports *private* IP-based connectivity to customers over shared *public MPLS* infrastructure.
- Resilient path management is becoming more important in future (optical) data network to provide non-disrupted guaranteed service on multiple component failures due to HW/SW errors, security attack, disastrous events, etc.
- We present (1) condition for existence of fast solution based on Hamiltonian cycle, (2) a simple resilient path management algorithm, and (3) decomposition theorem in a full mesh BGP/MPLS VPN.
- These results can be used to *dynamically* configure both primary and backup paths together satisfying a TE resilience requirement in MPLS backbone.
- The decomposition theorem allows both intra & inter BGP/MPLS networks to be managed efficiently.

In this paper, we have presented a fast resilient path management architecture and algorithm for BGP/MPLS VPN which is a full mesh structure. A model representing a resilience constraint is presented and an existence condition for finding fast backup path for the resilience constraint is derived for a full mesh BGP/MPLS VPN. It is based on Hamiltonian cycle. A simple path finding algorithm with $O(n)$ computational complexity is presented. Finally, we develop a decomposition theorem which enables the approach to be extended to the hierarchical MPLS backbone which is well scalable in a full mesh structure, using MPLS label stacking technique.

References are listed below:

[IETF MPLS] IETF MPLS Working Group, <http://www.ietf.org/html.charters/mpls-charter.html>

[IETF PPVPN] IETF PPVPN Working Group, <http://www.ietf.org/html.charters/ppvpn-charter.html>

[RFC2547bis] Eric C. Rosen, Yakov Rekhter et al., "BGP/MPLS IP VPNs," draft-ietf-ppvpn-rfc2547bis-04.txt, May 2003

[RFC2702] D. Awduche, J. Malcolm, et. al., "Requirements for Traffic Engineering Over MPLS" RFC 2702, Sept. 1999.

[RFC3031] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, January 2001.

[RFC2858] T. Bates, Y. Rekhter, et al., "Multiprotocol Extensions for BGP-4," RFC2858, June 2000

[RFC3469] V. Sharma, et al., "Framework for Multi-Protocol Label Switching (MPLS) - based Recovery," RFC3469, February 2003.

[Hara72] F. Harary, Graph Theory, Addison-Wesley Publishing Company, Reading Massachusetts, 1972.