

## An Efficient and QoS Guaranteed Network Architecture for Real-Time Premium Traffic over the Internet

<sup>1</sup>Manodha Gamage, <sup>2</sup>Mitsuo Hayasaka, <sup>3</sup>Tetsuya Miki

<sup>1,2,3</sup>The Department of Electronics, Faculty of Engineering, The National University of Electro-Communications, 1-5-1, Chofugaoka, Chofu-Shi, Tokyo, 182-8585, Japan.

### **Abstract:**

*The QoS provided by today's best effort Internet is not good enough for real-time premium traffic (PT). It is believed that QoS guarantees could be better provided by the connection oriented networks such as MPLS. However, these connection oriented networks are inherently more prone to network failures. Failures of connection oriented networks can be broadly classified into two types namely: link/path failures and degraded failures. Degraded failures that account for about 50% of total failures are mostly detected by control plane timers. The control plane and the data plane of connection oriented networks such as MPLS are logically separated and therefore a failure in the control plane should not always disconnect the communications in the data plane. The Virtual Path Hopping (VPH) concept in this study distinguishes these two types of failures and avoids the disconnections of communications in the data plane due to degraded failures. The computer simulations performed and the results are presented to support the concept. The Dynamic Virtual Path Allocation (DVPA) algorithm to improve the availability of the connection oriented backbone networks without compromising the network resource utilization efficiency is also discussed here. DVPA is to achieve very fast re-routing for link/path failures. Therefore the connection oriented architecture with VPH and DVPA could provide QoS guarantees and high availability for real-time applications over the Internet.*

**Keywords:** *Virtual Path Hopping, Dynamic Virtual Path Allocation, Multi Protocol Label Switching, premium traffic, QoS, connection oriented networks, real-time, degraded failures, link/path failures*

### **1. Introduction**

With the explosive growth and operational cost reductions of the Internet, the real-time multimedia applications over the Internet are increasingly becoming popular. The real-time Premium Traffic (PT) over the Internet prefers guaranteed QoS with minimum delay and jitter, guaranteed bandwidth, high availability, fast recovery from failures etc. The connection oriented packet networks such as Multi Protocol Label Switching (MPLS) [1] are better able to provide the required QoS in many of these dimensions compared to the connectionless, best effort IP networks. On the other hand connection oriented networks are potentially more vulnerable to failures and the focus of this study is to overcome the problems due to the network failures. According to [2], the failures in connection oriented networks such as MPLS can be categorized into degraded failures and link/path failures. Degraded failures are mainly due to the links of lower layers are not in suitable quality to guarantee the data transmission. They are caused by the control plane failures, restarts of the control plane nodes, tear downs of control plane sessions due to loss of adjacencies, congestion etc. They should not immediately disconnect the communications in the data plane because the data plane and the control plane in MPLS are separated logically. If degraded failures continue for more than a predefined period of time the data plane communications would be disconnected. The Virtual Path Hopping (VPH) concept identifies these degraded failures before the data plane communication session is terminated and the traffic carried Virtual Path (VP) is changed by way of a VP hop. The results of the computer simulations show that this is a promising proactive technique to minimize

failure occurrence in the data plane, especially for real-time PT.

There are many strategies to recover from the failures of the networks and provide high availability, but they are always with huge tradeoffs with the utilization of the network resources. Clearly for the emerging real-time premium applications, the service interruption times should be very small if cannot be completely eliminated and with the current strategies, a high availability in the backbone network could be achieved only with a network resource utilization of well below 50%. Therefore it is a timely requirement to find a solution to improve this efficiency while achieving a very high availability. Connection oriented network architecture with Dynamic Virtual Path Allocation (DVPA) algorithm and VPH that could achieve this in the future Internet is discussed here. In this architecture the VPH eliminates the degraded type failures and the DVPA algorithm achieves a very fast recovery from link/path failures without compromising the network resource utilization efficiency. Rest of this paper is organized as follows. In the next section, the problem description and the existing solutions are analyzed briefly. In section 3, the proposed VPH concept is discussed in detail. A brief description of the DVPA algorithm is done in the section 4. A numerical analysis of DVPA algorithm is done in section 5. The performance of the VPH concept is evaluated by way of computer simulations and results are presented in Section 6. Finally the paper is concluded in Section 7.

## 2. Problem Analysis and Existing Solutions

MPLS provides guaranteed QoS in terms of delay, jitter and bandwidth, because of its connection oriented nature and it supports constraint based routing. Therefore the following are the two key problems in achieving the preferred QoS in the connection oriented networks for real-time PT.

- a) *Minimize the failures* in the network and if there are any failures, minimize the interruptions to data traffic due to them.
- b) *Distribution of traffic and Utilization* of the network resources efficiently.

Degraded failures [2] initially start as temporary interruptions to the applications in data plane. If this continues beyond a pre-determined time period the relevant application would be terminated. In other words degraded failures should not immediately terminate the data plane communications as the data and control planes of MPLS are logically separated. Most of the degraded failures are usually detected by control plane timers; RSVP Hello State Timer in RSVP-TE [7] or the Keep Alive Timer in LDP. The values of these control plane timers are usually decided at the time of formation of the control plane session and are in the range of 30-40s, but they can be as large as 60-90s. Studies done on actual ISP networks have shown almost 50% of total network failures are very short lived degraded failures and they are suspected to be due to control plane failures such as misunderstandings of loss of adjacencies, teardowns of control plane sessions etc [3]. The terminations of data communications in the data plane due to these degraded failures that are detected by the control plane timers can be avoided if they can be rectified before the control plane timers expire.

On the other hand, link/path failures should be overcome by re-routing. Two such conventional solutions for network failures in real-time premium applications are 1+1 protection, 1:1 protection (extendible to m:n protection). As we know, 1+1 has very fast recovery times because the backup path always carries the same traffic as the active path but very inefficient with respect to the usage of bandwidth and other resources. On the other hand 1:1 improves the bandwidth efficiency at the expense of the recovery time because the backup path always available for pre-emptible low priority traffic when the active path has no failure. Backup bandwidth (BBW) sharing is becoming increasingly popular due to the improved bandwidth efficiencies as a single backup path can be shared by many link-disjoint active paths [5, 6], but they have their share of limitations too. Almost all of these proposals consider single fault situations with the assumption that any failure could be

repaired before the next failure occurs. Studies have shown that the algorithms that provide 100% restorability for single failures could provide only 60%-70% restorability for dual failure situations and this can be even as low as 20% [18]. The expansion of networks and increased durations of applications prefer future networks to possess 100% restorability even for dual faults. Usually, the network status varies rapidly over time and therefore very high BBW sharing attribute at a given time may not be true at another time and therefore the bandwidth utilization may not be as high as expected. Furthermore, none of the above mentioned solutions that are very widely used at present distinguish the two types of failures; degraded type that can be avoided before terminating the data plane communications and link/path type failures that occur in the data plane.

The graceful restart and fault tolerance [14, 15] of the control plane protocols are such attempts to avoid the control plane failures terminating the data plane communications, but all the routers do not support them. Even if all the routers could preserve the forwarding states, there would be scalability problems. Furthermore they are not much suitable for real-time interactive communication as they might lose some data at a fail over. The VPH concept proposed here is an alternative proactive technique to avoid terminations of data communications in the data plane due to degraded failures detected by the control plane timers. The next problem to be solved is to perform very fast re-routing for link/path failures in the data plane of PT while maintaining a higher utilization efficiency of network resources. The DVPA algorithm proposed here solves that problem.

### 3. Virtual Path Hopping Concept

Main objective of the proposed VPH is to eliminate terminations of data communications in the data plane due to degraded failures that are detected by control plane timers (T) of connection oriented network such as MPLS that uses in-band signaling and thereby achieve a very high availability. VPH is proposed especially for PT. VPH refers to “changing a degraded active VP (AVP) in the data plane to a new VP that can guarantee QoS before it fails due to the expiration of control plane timers (T).” This is done by creating a new control plane session to activate an unused VP in a pre-determined VP-pool between an ingress/egress pair and transferring the traffic in the degraded AVP, whose control plane timer is about to expire. On the other hand the term re-routing refers to ‘the change of data transmission from active path to a backup path, after a failure in the data plane’. Any restoration by re- routing after a failure would have an outage of service where as VPH has no service outages as it uses make before break concept [7] when hopping the VPs. Though the explanations here are for IP/MPLS with RSVP-TE in the control plane we are confident that this concept can be equally applicable for any connection oriented network with minor changes according to the protocols. In IP/MPLS the VP is called a LSP.

The ingress nodes of IP/MPLS or any other connection oriented IP networks will have to play a major role in the implementation of VPH concept. When the traffic arrives at ingress, ideally a link- disjoint VP-pool, which contains the candidate VPs that participate in VPH is decided between the ingress and egress as shown in figure 1. This VP-pool should contain at least three link-disjoint VPs for the effectiveness of this concept and with the increment of number of VPs in the VP-pool, the effectiveness will increase. The number of VPs in the VP-pool depends on several factors such as the load of the network, Service Level Agreements (SLA) of clients, failure analysis and failure distributions of the network etc. It is not necessary to restrict VPH only for link-disjoint situations and block the traffic, whenever it is not possible to find a minimum of 3 fully link-disjoint VPs for a VP- pool. Therefore we suggest that it is preferred to find the VP-pool with fully link-disjoint VPs for the best performance of VPH, but if this condition cannot be met, it is acceptable to create the VP-pool with VPs that are not fully link-disjoint provided the QoS requirements of the traffic accepts it. This entirely depends on the requirements of the applications. However two VPs of a VP-pool should not share the same tunnel at all. VPH would improve the performance in both situations but the 100% reliability in all occasions might not be achieved, when all VPs are not fully link-disjoint. It is important to have this option as there may be certain applications that are better off with starting

the communication with a VP-pool with not fully link-disjoint VPs rather than blocking them. There are many algorithms proposed in the literature [8-12] to find link-disjoint paths between a pair of ingress and egress. It is beyond the scope of this paper to discuss them in great detail. Also instead of using Shortest Path First (SPF) only, to determine the VP-pool, it is necessary to use QoS routing along with SPF as proposed in [8] and [10]. This will necessitate the routers to advertise the available bandwidth, delay jitter, delay, packet loss rate, QoS requirements of traffic etc.

In this discussion, VP-pool of n VPs is considered. All VPs in the VP-pool should be ranked (from rank#1 to #n such that the best VP is rank #1) considering factors such as QoS they can provide, costs in terms of bandwidths, reliability and the length of the VP. VPH doesn't reserve resources of all the VPs in the VP-pool as it is very inefficient. At the time of forming the VP-pool the ingress only does the path computations and it does not do any label bindings. The resource reservation and establishment of label binding for every VP is done by exchanging PATH and RESV messages of RSVP-TE, just before that they are to be used by the intended traffic. Therefore any best effort traffic (BET) that arrives after forming the VP-pool can use any resources of the unused VPs in that VP-pool. This BET is preempted by PT, similar to conventional 1:1 protection, if there is a degraded failure in PT carried AVP. In other words VPH reserves resources and uses only one VP at any given time. At the beginning of the communication session, ingress will always start the communication via the rank #1 VP and then it would hop periodically to rank #2, rank #3.....rank #n, back to rank #1 and so on in a cyclic manner assuming those VPs are neither used by other traffic nor they are failed. If a VP is not free for VPH then the ingress will have to update its VP-pool with a new VP. The period of VPH is a vital parameter in this concept and for optimum results it should be less than the minimum threshold value of timers used to detect the control plane failures as shown in the simulation results of one of our previous conference papers [16].

Once the ingress explicitly decides rank #1 VP, it would first form the control plane RSVP session of that VP and reserve network resources to guarantee the QoS of the data to be communicated. Usually the control plane timer threshold values mentioned in section 2 are decided by negotiating with neighboring peers. Therefore these threshold values of the control plane peers of the same session could vary from each other. The period of the VPH should be decided by the ingress after deciding these threshold values and it should be less than the minimum threshold value of all the peers of the session. For example, as shown in the figure 1, if the timer values of the 3 links of VP#1 are 40s, 30s and 50s respectively, then the period of VPH should be less than 30s (the minimum of the 3).

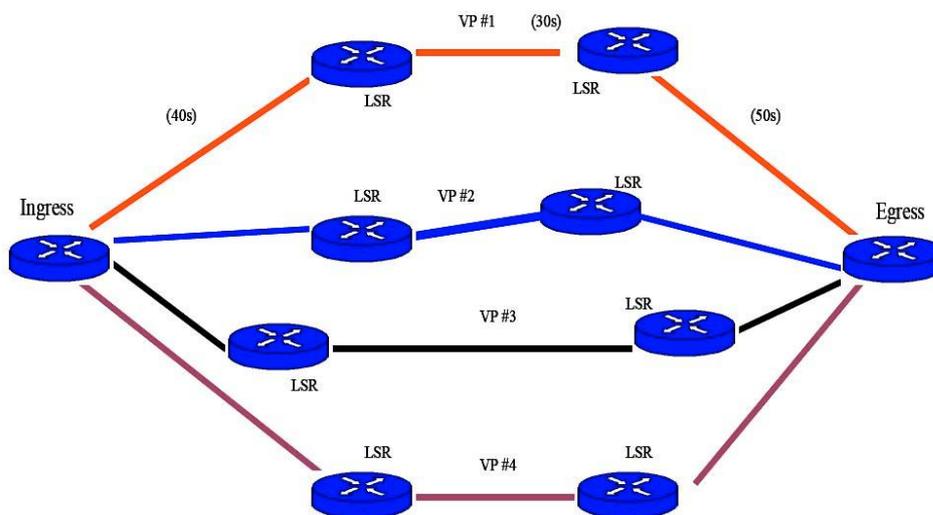


Figure 1: The concept of Virtual Path Hopping with a VP-pool of 4 VPs

It should be less than 30s such that VPH can be performed before the 30s control plane timer expires. The timer threshold value information can be informed to the ingress with the RESV messages of RSVP-TE. In this situation the VPH concept will make sure it will change the active VP, before the control plane timer expires and RSVP session is torn down causing a disconnection in the corresponding data plane. Obviously the new VP is formed and maintained by a new control plane RSVP session. Since RSVP-TE always exchange PATH and RESV messages between an ingress-egress pair to form a new control plane session and corresponding VP, it is clear that round trip time (RTT) + processing time at each node is the transit time of a VP hop. Therefore this transit time depends mainly on the distance between the ingress and egress. If necessary it is possible to set priority as Key to all the PATH and RESV messages involved in VPH to reduce the transition time further. 'Actively reserved bandwidth architecture' [19] can be used for fast allocation of resources and minimize the call rejections due to lack of resources. It is necessary to maintain a VP-pool Table at the ingress, in order to implement the VPH smoothly. Such a table essentially should include information such as LSP ID, rank of LSP and status of LSP (active or inactive) apart from the information of a conventional routing table of MPLS.

All this explanation of VPH is for unidirectional LSP/VP. It is essential to have bi-directional VPs for interactive real-time communication and this can be done in 3 ways according to the existing protocols [20].

- a) It is possible to have two unidirectional VPs created separately (with two separate control channels) for the forward and backward directions. In this situation VPH concept can be applied to forward and backward directional VPs separately by having two separate VP-pools. In other words a degraded failure in the forward VP doesn't warrant a VPH in the corresponding backward directional VP and vice versa. The network administration tools could tie the two directional VPs and manage them as a single entity.
- b) Use the 'Upstream Label' object in the PATH message and follow the procedures explained in GMPLS [21] to create bi-directional LSP/VPs by exchanging only two messages similar to unidirectional VPs.
- c) A more secure solution to follow the same procedure as in b) above and use RESV\_CONF message in addition to PATH and RESV messages in order to confirm the successful formation of bi-directional VPs.

Since all a), b), and c) are provided by the existing protocols, the implementation of either method in VPH is technically feasible and the most suitable method depending on the network and traffic type could be selected.

Whenever a VPH is performed according to the above algorithm, there will be some overhead traffic added to the network. This added overhead can be minimized by minimizing the number of VP hops. Therefore instead of periodic VPH, a more efficient VPH, where VP hops are triggered by another timer called VPH\_Timer is evaluated in this study. This is referred to as non-periodic VPH. The VPH\_Timer is reset to zero every time a VPH is done or a PDU is received. When RSVP-TE is used in the control plane, RSVP Hello State Timer (T) of control plane peers is decided by negotiating with the neighboring peers. Usually  $T=k*R$ ; where R is the refresh period and k is an integer (by default  $k=3$  or  $4$  and  $R=10s$  in most routers) [7, 13]. In the Same way VPH\_Timer values also can be decided by each control plane peer that participates in VPH and they should be less than T. Whenever the VPH\_Timer expires, T should be reset to zero and a Refresh\_Timer should be started in the same way as of graceful restart in RSVP and LDP [14]. This "refresh" state is used to buy time to inform the ingress about the expiration of the VPH\_Timer and to carry out the VP hop smoothly without any data losses in the data plane. Therefore data plane communications should not be terminated during the "refresh" state and this refresh state is used to implement make before break concept to make sure no data losses occur when hopping the VPs. The Refresh\_Timer value should be decided based on the transit time of VPH that mainly depends on the RTT between ingress and egress. It is possible to use the "Lost" state of Hello messages to inform the expiration of VPH\_Timer to the ingress as it is used in graceful restart of RSVP-TE implemented by some widely used routers [15]. If VPH is not performed, the data plane communications are terminated whenever

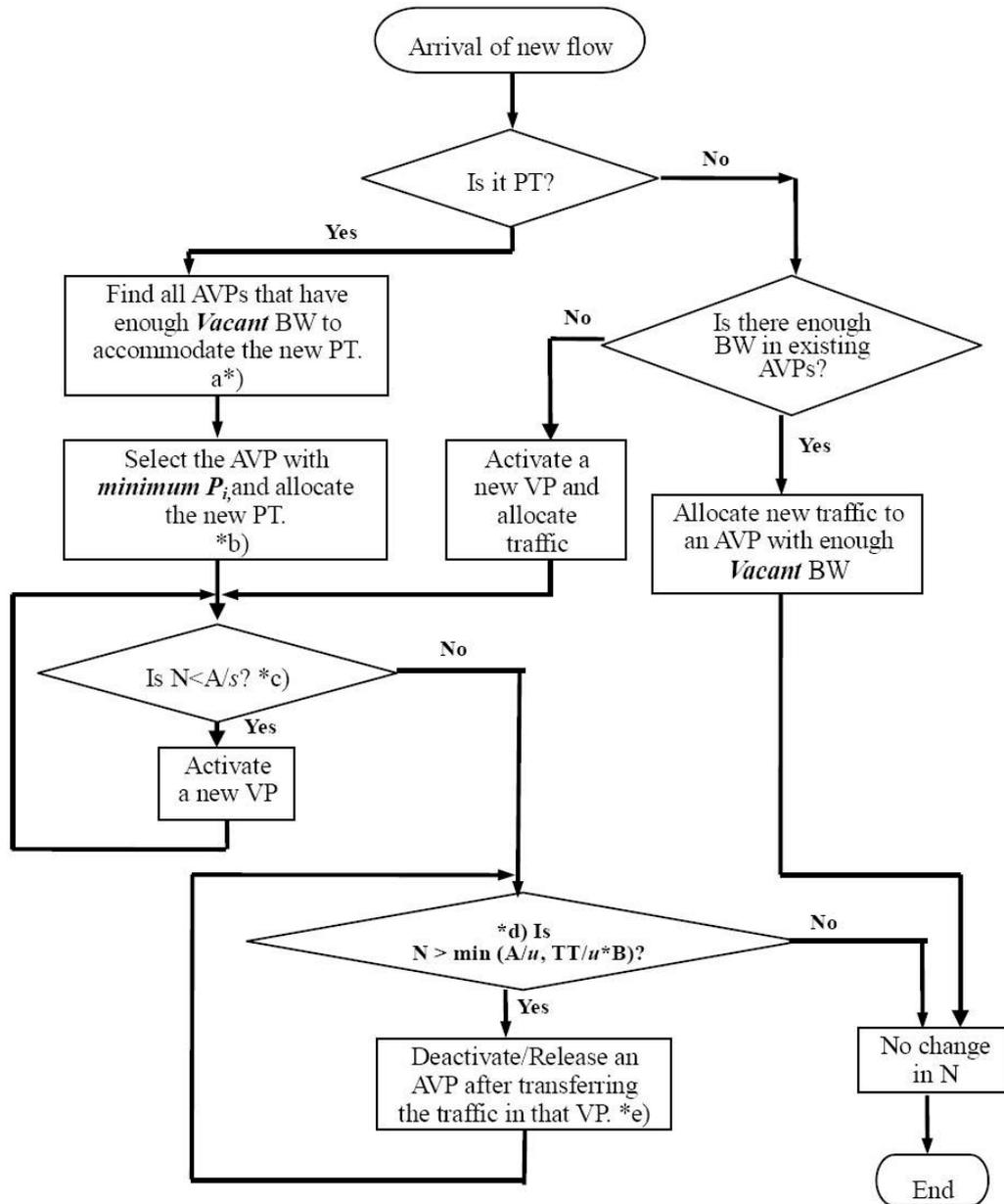
the timer (T) to detect the degraded failures is expired. Then there should be a re-routing to a BP to recover the communications. In the event of a re-routing, there will be a data loss and VPH will eliminate the re-routings in data plane due to degraded failures that are identified by the control plane timers. This will enable to achieve a very high availability and reliability for PT.

VPH also has some other advantages. Since the main objective of the VPH is to overcome degraded failures of the connection oriented networks that uses in-band signaling, it is beyond the scope of this paper to evaluate all the additional advantages in detail. But some advantages such as traffic distribution, avoiding trap topologies are briefly mentioned below. As there is a VP-pool between each pair of ingress and egress, it is possible to achieve 100% restoration even for dual fault situations without disturbing the TE of network. VPH also has the ability to dynamically distribute the traffic well, through the network. This will distribute the network load evenly throughout the network reducing the stress on links and nodes due to high congestion. This in turn will reduce the probability of failure in a node or in a link [22]. VPH can also avoid 'trap topologies', which causes problems in other existing BBW sharing techniques. Trap topology is the situation where it is impossible to find a link-disjoint BP for a certain APs at a given point of time. If this happens for a communication session, according to the current setup, it will last through out the session and this is a very critical situation, especially for PT. In VPH, before a communication session begins it will decide a VP-pool mostly consists of link-disjoint VPs avoiding 'trap topologies'.

#### 4. Dynamic Virtual Path Allocation (DVPA) Algorithm

Link/path failures should be overcome by fast re-routings, transparent to the users. Only a preplanned backup path (BP) can achieve this in real time. The recovery time consists of two main parts; time taken to detect failures and time taken to activate the BP. The proposed DVPA algorithm virtually makes the time taken to activate the BP to be 'zero' because BP for PT is always active. The most important point is the higher utilization efficiency (UE) of resources that can be achieved even with an active BP in this algorithm because DVPA dynamically varies the number of active VPs (AVPs) to make sure that BET occupies the ABP when there are no failures. In the proposed algorithm, the traffic received at the ingress is divided into two types, namely PT and BET. In DVPA also, the VP- pool is used as in the VPH concept, but instead of one AVP in VPH concept, it is suggested to have few (at least 2) AVPs in DVPA. Preplanned protection is only for PT, where the re-routing should be done within a very short time (typically less than 50ms) and the ingress will calculate PT ratio ( $P_i$ ), which is defined as the ratio of PT of  $i$ th AVP to Total BW of  $i$ th AVP. DVPA increases the number of AVPs with the increase of traffic and reduces the number of AVPs if the traffic decreases in order to maintain a higher the UE.

The flow chart in figure 2 indicates how the VP allocation is done for newly arrived traffic flows in the proposed algorithm. According to this algorithm when the PT arrives at the ingress, it should be allocated to an AVP with minimum  $P_i$ . This helps to distribute PT among AVPs as much as possible and keeps the maximum  $P_i$  ( $\max(P_i)$ ) to a minimum value. This minimizes the spare capacity (SC) necessary to protect PT from failures as explained below. SC to protect PT of an AVP is defined as the difference between the total capacity and BW used by PT of an AVP. In this algorithm the SC to protect PT should be always large enough to recover a PT of  $\max(P_i)*B$ ; where B is the bandwidth of  $i$ th AVP, (i.e. maximum PT in an AVP of the VP-pool) in a single failure protection situation and it should be  $2*\max(P_i)*B$  in a dual failure protection situation as explained in the next section.



**BET:** Best Effort Traffic

**PT:** Premium Traffic

**TT:** Total Traffic; where  $(TT=BET+PT)$

**Vacant BW** of an AVP = Total BW available in a AVP – Total BW used by PT&BET of that AVP

**N:** Number of Active Virtual Paths (AVP) between a pair of ingress and egress

\*a) If no AVP with enough VB can be found, it is necessary to activate another VP in the VP-pool and if no more VPs are available to activate, this traffic has to be dropped due to lack of BW availability.

\*b) For better distribution of PT, so that a lower value of  $\max(P_i)$  will lead to less spare capacity requirement.

\*c) When single failure at  $j^{th}$  AVP:  $A = \max(P_i) + (\sum_{i=1}^N P_i) - P_j + 1$  and  $s$ - Service factor, where  $0 < s < 1$  and  $1 \leq i, j \leq N$

\*d) B- Bandwidth of the VP and  $u$ -Utilization factor, where  $u > 1$

\*e) Release the VP with minimum traffic

Figure 2: Algorithm to allocate AVPs in DVPA

Obviously the selected AVP should have enough vacant bandwidth (VBW) to accommodate newly arrived traffic flow. The VBW of an AVP is defined as the difference between the total BW available and total BW used by PT and BET of an AVP. On the other hand if the newly arrived traffic is BET, it is simply allocated to an AVP with enough VBW. If there is no enough VBW in

AVPs for newly arrived PT or BET another VP in the VP-pool is activated. If no such VP to be activated is available, the newly arrived traffic is dropped due to the lack of BW. In this algorithm, if an AVP (say AVP\_1) fails, the affected PT is recovered by re-routing them to another existing AVP (say AVP\_2). If the SC to protect PT of AVP\_2 is not enough to accommodate the PT of failed AVP\_1, then the affected PT is distributed among many AVPs (AVP\_2, AVP\_3 etc.) with enough SC to protect PT. Therefore in some situations BET of AVP\_2 and/or AVP\_3 also might have to be preempted to accommodate PT of failed AVP\_1. All preempted BET can be recovered by activating another VP in the VP-pool, if there is no enough capacity to accommodate them among the already activated VPs as BET is not time critical. Here too the bi-directional VP requirement can be handled as explained in the section 3. Also the dynamic variation of number of VPs in the DVPA is explained in the next section.

## 5. Numerical Analysis of DVPA Algorithm

In the analysis performed here for DVPA algorithm, it is assumed that VP-pool consists of  $n$  link-disjoint VPs each with a bandwidth of  $B$ , and there are  $N$  AVPs at any given time ( $2 \leq N \leq n$ ). The initial analysis is done for a single failure scenario and then the analysis is extended for dual failure situations that might be more appropriate for future networks.

PT ratio of  $i^{\text{th}}$  AVP;

$$P_i = \frac{PT\_in\_i^{\text{th}}\_AVP}{Total\_BW\_of\_i^{\text{th}}\_AVP}, \text{ where } (1 \leq i \leq N)$$

Total PT in  $N$  AVPs:

$$B * \left( \sum_{i=1}^N P_i \right)$$

Total available BW after failure of one AVP:

$$B * (N - 1)$$

The available BW for recovery of PT (SC to protect PT), if  $j^{\text{th}}$  AVP fails ( $1 \leq j \leq N$ ) is given by;

$$B \left( N - 1 - \left( \left( \sum_{i=1}^N P_i \right) - P_j \right) \right)$$

Here the bandwidth used by the BET is not considered because the objective of this method is to provide guaranteed QoS and availability for PT. Maximum PT ratio to be recovered is  $\max(P_i)$  and DVPA algorithm always should have enough spare capacity to recover it.

For 100% restorability of PT (Single failure);

$$B \left( N - 1 - \left( \left( \sum_{i=1}^N P_i \right) - P_j \right) \right) > B * \max(P_i)$$

$$\therefore N > \max(P_i) + \left( \sum_{i=1}^N P_i \right) - P_j + 1$$

In figure 2;

$$A = \max(P_i) + \left( \sum_{i=1}^N P_i \right) - P_j + 1$$

A service factor  $s$  is considered to allow some extra BW to make sure that the AVPs are not overloaded. Always  $N > A$  should be maintained and if  $N < A/s$ , where  $0 < s < 1$ , another VP should be activated in the VP-pool in order to increase  $N$  by one. The PT ratio for every AVP is calculated by the ingress and these values are used to allocate PT to the AVP with minimum  $P_i$  as explained in the section 4. Whenever a new allocation of PT is done the PT ratios are updated. The utilization factor,

$u$  is defined such that  $0 < u < s < 1$  and if  $N > \max\{A/u, TT/(B \cdot u)\}$ , then the utilization of the AVPs are low. Therefore try to reduce  $N$  by one as explained before.  $u$  is used to avoid frequent fluctuations in number of AVPs that can occur if a single threshold value is used to increase and decrease  $N$ . The values of  $s$  and  $u$  can be decided by the network administrator according to the needs of the network. Similar analysis can be extended for a dual failure situation, where  $j^{\text{th}}$  and  $k^{\text{th}}$  VPs fail and then it is possible to obtain;

$$A = 2 \max(P_i) + \left(\sum_{i=1}^N P_i\right) - (P_j + P_k) + 2$$

This means that there should be enough spare capacity to recover  $2 \max(P_i)$  of affected PT.

## 6. Simulation and Analysis of Results

The performance of VPH and DVPA were evaluated by computer simulations. The simulation programs were done using Visual C++ and the number of links and their connectivity were randomly decided for each network topology. The traffic flow arrivals were simulated according to Poisson distribution. Many simulations were performed with different average traffic flow arrival values such as 3s, 5s, 10s, 30s, 60s, and 300s. The durations of communication sessions were decided based on an exponential distribution with different averages of 300s, 600s, 900s, 1800s, and 3600s. The bandwidths of the sessions were randomly decided with averages of 1Mbps, 5Mbps, 10Mbps, and 20Mbps. After many simulations, it was found that the improvements due to the VPH and DVPA were not much sensitive to these average values. The results shown here are for averages of 10s, 1800s and 10Mbps for traffic flow arrivals, session durations and bandwidths of sessions respectively. The number of sessions in a flow arrival was randomly decided to be between 1 and 10. According to many simulations, the parameters of the distribution of repair times of failures had negligible effects to the performance of VPH and therefore it was assumed to be a constant for VPH.

### 5.1. VPH

Different network topologies with nodes 10, 20, 40, 50, 60 and 90 with different number of links were simulated for many failure combinations in nodes and links similar to in Table 1. The results indicated almost complete elimination of terminations of the data communications in the data plane due to degraded failures, irrespective of the network topology. Therefore the results of the topology with 90 nodes and 270 bidirectional links, simulated for failure combinations in Table 1 are presented here. For the same failure combination, over 30 connectivity orientations of topologies were simulated. Simulations were performed for VPH and without VPH scenarios. Whenever a network failure occurs in the data plane a fast re-routing is performed to restore it and a free VP from the VP-pool was always used as a backup path in all simulations. Therefore number of re-routings done will be a count for network failures and this is used as a performance evaluation measure of VPH concept. VPH\_Timer values were randomly decided for each peer to be a multiple of 10s in the range of 30s-80s. Figure 3 indicates the Frequency of occurrence vs. Number of re-routings per month graphs. The dotted lines of these graphs indicate the most frequently occurred re-routings per month and all these distributions are concentrated around the dotted lines. According these results, VPH always reduced the re-routings by about 50%, irrespective of the failure probability of nodes and links indicating almost 50% reduction of failures in the data plane. This accounts for almost 50% of degraded failures that are identified by the control plane timers and can be eliminated by VPH. This elimination of terminations of data communications in the data plane due to degraded failures in connection oriented networks improves the availability drastically.

## 5.2. DVPA Algorithm

A VP-pool consists of 6 link-disjoint VPs was simulated to evaluate the performance of the proposed DVPA algorithm for single faults. The bandwidth of each VP was considered to be 100Mbps. All simulations were carried out for durations of 24 hours. The utilization factor,  $u$  and service factor,  $s$  explained in the numerical analysis were set to 0.7 and 0.9 respectively. The DVPA algorithm was also evaluated for dual fault situations and it was necessary to have a VP-pool of 7 link-disjoint VPs, in order to accommodate the same amount of traffic as in single fault scenario. Networks with 6 fixed AVP and 7 fixed AVP between an ingress/egress pair were also simulated in order to compare the above single and dual fault scenarios respectively. The fixed numbers of AVPs were decided as 6 and 7 for the two scenarios because they were the minimum bandwidth (600 and 700Mbps) required to accommodate all traffic during the peak time. Figure 4 shows the Utilization Efficiency (UE) Vs Load graphs for the proposed DVPA and for fixed number of AVPs for both single (S) and dual (D) fault scenarios. Simulations were done for 10%PT, 20%PT, 25%PT, 30%PT, 35%PT, 40%PT, 45%PT, and 50%PT. Figure 4(a) and (b) are for 25% and 50% of PT respectively. In all these simulations there were no dropped traffic flows even at the peak time. The Utilization Efficiency (UE) here is defined as the ratio of used bandwidth (PT and BET) between the ingress and egress to the total available active bandwidth between them. According to the graphs in figure 4(a), when the proposed algorithm was implemented for single fault scenario, the Utilization Efficiency (UE) for 25% PT is over 80% for loads over 0.3. Even at low loads (0.1-0.2) it is around 50%. In contrast to this the Utilization Efficiency (UE) for fixed AVP (for single fault case) is very low for the loads less than 0.5 and it is less than 80% even for loads of 0.8.

Table 1: Different failure combination scenarios simulated

Combination	Failures/link/month	Failures/node/month
I	0.01	0.01
II	0.01	0.1
III	0.05	0.1
IV	0.05	0.01

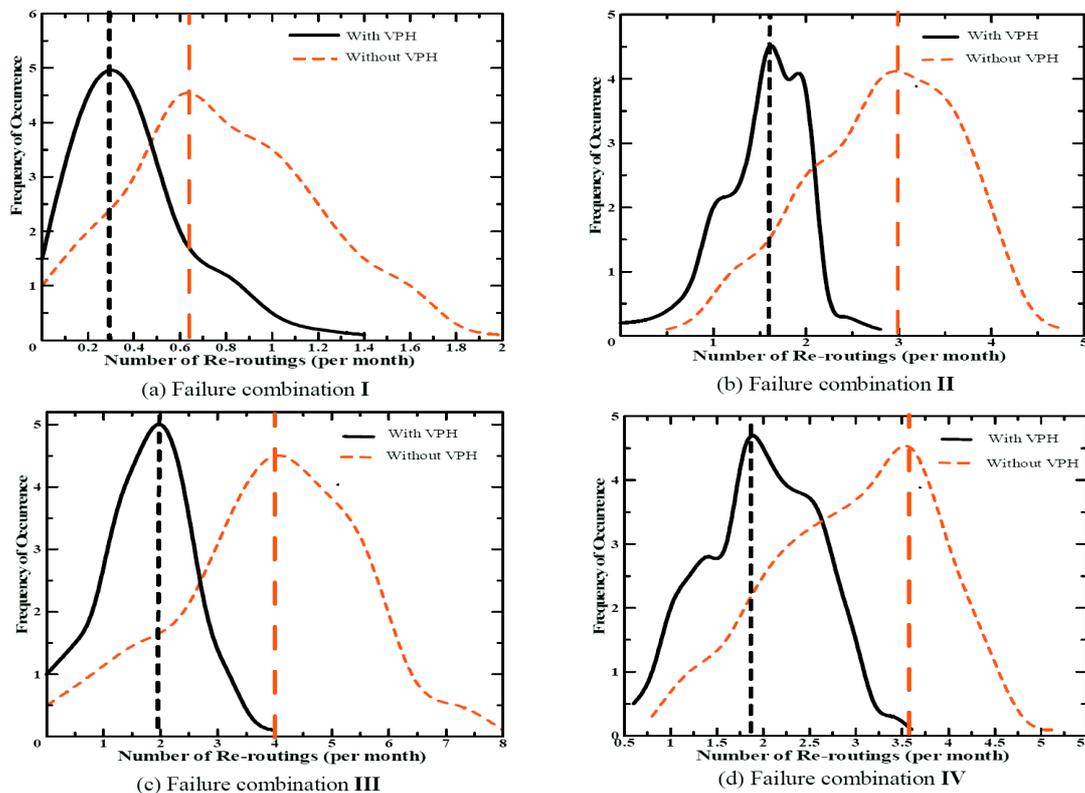


Figure 3. Frequency of occurrence Vs Number of re-routings (90 nodes)

As shown in the Figure 4(b) the Utilization Efficiency (UE) is slightly reduced when PT is increased up to 50%, but still DVPA improves UE drastically compared to fixed VP case. As expected when average PT was over 50%, first it started to drop few BET calls and when further increased, it started to drop few PT too. This was due to lack of bandwidth and it can be overcome by increasing the bandwidth of each VP or by increasing the number of VPs in the VP-pool. Increasing the VPs in the VP-pool does not affect the Utilization Efficiency (UE) in the proposed algorithm as it dynamically varies the number of AVPs. However increasing number of VPs between any pair of ingress and egress would make the conventional fixed VP method more and more inefficient during low traffic times. In all these simulations there were no interruptions for applications, as the recovery was very fast when ever there was a failure.

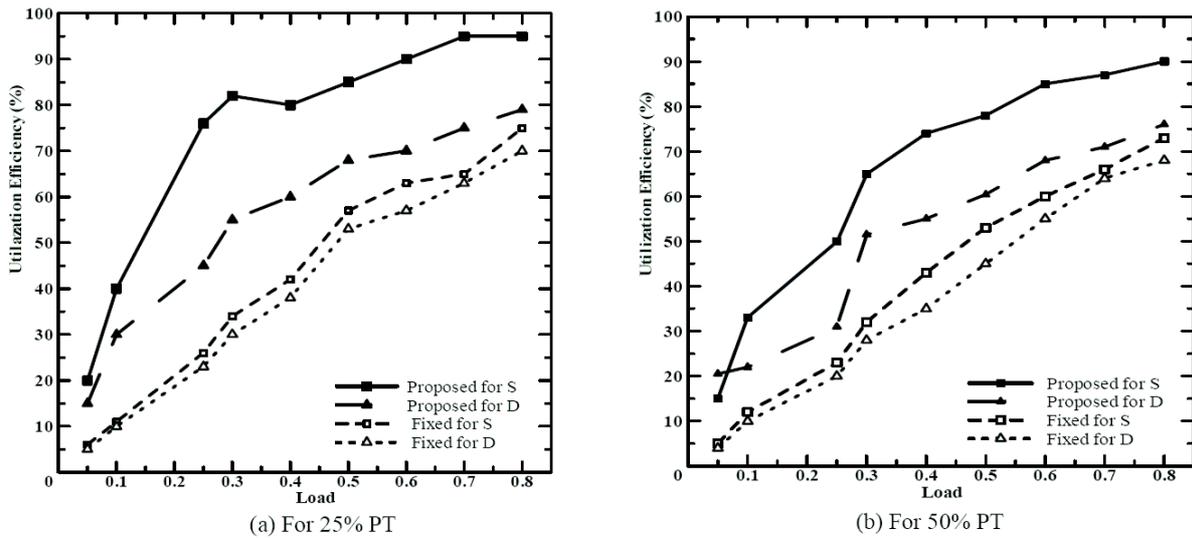


Figure 4. Utilization Efficiency Vs Load

Figure 5 describes how the proposed algorithm distributes the PT among all AVPs for both Off-peak, figure 5(a) and peak, figure 5(b) conditions. Figures 5(a) and 5(b) are for when the load is 0.4 and 0.8~0.9 respectively. These figures are the results of the simulations performed for single fault situation when there is an average of 50% PT. Figure 5(a) indicates that only 3 VPs are activated (VP#1, #2 and #3) at low loads and figure 5(b) shows, when the load is very high the algorithm activates all VPs in the VP-pool. This clearly shows how the distribution of PT in DVPA algorithm can improve the Utilization Efficiency (UE) of the network resources.

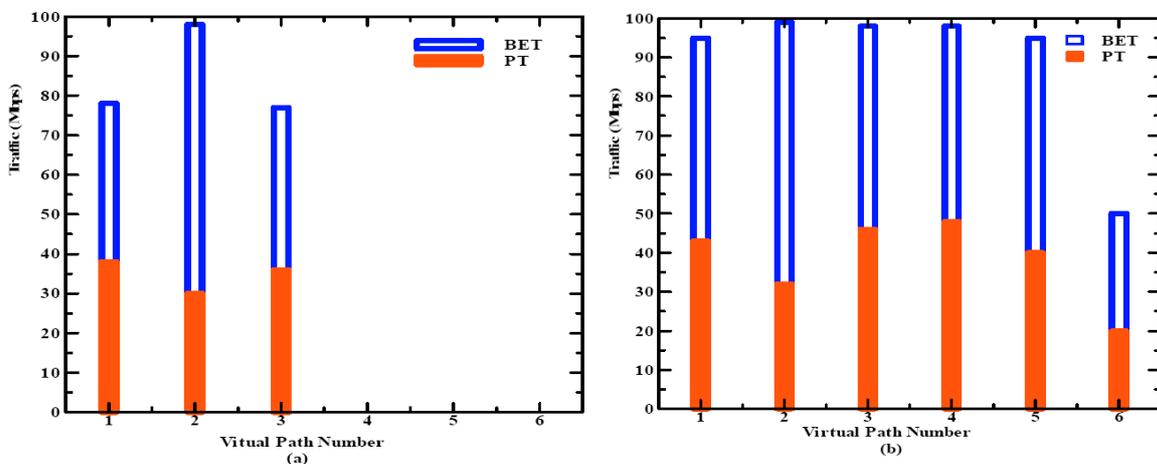


Figure 5. Traffic with respect to VP Number (a) Off-Peak and (b) Peak

## 7. Conclusions and Further Work

The rapid growth of real time multimedia applications over the Internet demands guaranteed QoS and near 100% availability. Therefore connection oriented networks that can meet QoS demands with respect to delay, jitter, and guaranteed bandwidth are to dominate the future Internet. Connection oriented networks are more vulnerable to network failures and it is a timely requirement to find a solution to achieve 100% availability not only for single failure situations but also for dual failure situations in these networks. DVPA algorithm with VPH concept provides this demanded availability without deteriorating the Utilization Efficiency (UE) of network resources as the results of this study show. The terminations in the data communications due to the degraded type failures identified by the control plane timers can be completely eliminated by VPH. The link/path failures of data plane for PT can be recovered with very low recovery times while keeping the Utilization Efficiency (UE) of resources over 70% to 80% for almost all very common load values using the proposed DVPA algorithm. The proposed algorithm here virtually takes no time to activate the backup path in the event of a link/path failure as it uses an AVP as a backup path. Therefore it is possible to conclude that implementation of DVPA algorithm in conjunction with VPH concept in connection oriented networks such as MPLS can achieve the QoS requirements of the future real-time premium applications over the Internet.

In the very near future we expect to carryout more simulations, especially for the DVPA algorithm and obtain more generalized results. Furthermore it is expected to investigate the possibilities of testing VPH concept in conjunction with DVPA algorithm in a real network using real-time traffic. It is also worthwhile to measure the blocking probability of session arrivals due to unavailability of link- disjoint VP-pool and to find a suitable solution to minimize it, as future work of this study.

## References

1. E. Rosen et al., "Multi Protocol Label Switching Architecture", IETF RFC 3031, Jan 2001.
2. V. Sharma et al., "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", IETF RFC 3469, Jan 2003
3. Gianluca Iannaccone et al., "Analysis of link failures in a IP backbone" Proceedings of Internet Measurement Workshop 2002, <http://www.icir.org/vern/imw-2002/imw2002-papers/202.pdf>
4. D. Awduche et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels", IETF RFC 3209, Dec. 2001.
5. Yijun Xiong, Lorne G. Mason, "Restoration strategies and spare capacity requirements in self-healing ATM networks", IEEE/ACM Transactions on Networking, no. 1, Feb 1999 pp. 98-110.
6. Li Li, Milind M. Buddhikot, Chandra Chekuri, Katherine Guo, "Routing Bandwidth Guaranteed Paths with Local Restoration in Label Switched Networks", Proceedings of the 10th IEEE International Conference on Network Protocols 2002.
7. P. Pan, G. Swallow, and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels" <http://www.ietf.org/internet-drafts/draft-ietf-mpls-rsvp-lsp-fastreroute-07.txt>
8. S. D. Nikolopoulos, A. Pitsillides and D. Tipper, "Addressing Network Survivability Issues by Finding the K- best Paths through a Trellis Graph", Proceedings of IEEE INFOCOM 1997.
9. B. Szviatovszki, A. Szentesi and A. Juttner, "On the Effectiveness of Restoration Path Computation Methods", <http://www.cs.elte.hu/~alpar/publications/proc/RestoPath.pdf>
10. Y. Guo, F. Kuipers, and P. V. Mieghem, "Link-Disjoint Paths for Reliable QoS Routing", <http://www.nas.its.tudelft.nl/people/Piet/papers/dimcra.pdf>
11. Y. Bejerano, Y. Breitbart, A. Orda, R. Rastogi, and A. Sprintson, "Algorithms for Computing QoS paths with Restoration", <http://www.paradise.caltech.edu/~spalex/pub/conferences/infocom03.pdf>
12. G. LIU, Y. Yang, and X. Lin, "Performance Evaluation of K Shortest Path Algorithms in MPLS Traffic Engineering", IEICE Trans. Commun., vol.E87-B, no.4, pp.1007-1010, Apr. 2004.
13. R. Braden L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReServation Protocol (RSVP)", IETF RFC 2205, Sept. 1997.
14. M. Leelanivas, Y. Rekhter, and R. Aggarwal, "Graceful Restart Mechanism for LDP", RFC 3478, February 2003
15. Cisco product features, "MPLS Traffic Engineering-RSVP Graceful Restart", [http://www.cisco.com/en/US/products/sw/iosswrel/ps1829/products\\_feature\\_guide09186a008027db17.html](http://www.cisco.com/en/US/products/sw/iosswrel/ps1829/products_feature_guide09186a008027db17.html)
16. [16]M. Gamage, H. Mitsuo and T. Miki, "Virtual Path Hopping to overcome Network Failures due to Control Plane Failures in Connection Oriented Networks", Proc. of joint conference of the 10th APCC and the 5th MDMC, China. Aug. 2004.
17. Manodha Gamage et al. "Implementation of Virtual Path Hopping (VPH) as a Solution for Control Plane Failures in Connection Oriented Networks and an Analysis of Traffic Distribution of VPH", to be appeared in Proceedings of 3rd International Workshop on QoS in Multi-service IP Networks (QoS-IP 2005), Feb. 2005, Catania, Italy.
18. M. Clouqueur and W. D. Grover, "Availability Analysis of Span-Restorable Mesh Networks", IEEE Jurnal on Selected Areas in Communications, vol. 20, no. 4, May 2002.
19. G. Kuo, and C. T. Lai, "A new architecture for transmission of MPEG-4 video on MPLS networks", ICC 2001 - IEEE International Conference on Communications, no. 1, June 2001 pp. 1556-1560.
20. N. Jerram, A. Farrel, "MPLS in Optical Networks", White Paper of Data Connection Ltd, <http://www.dataconnection.com>.
21. E. Mannie, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", IETF RFC 3945, Oct. 2004
22. G. Weichenberg, V.W.S. Chan, and M. Médard, "A Reliable Architecture for Networks under Stress," Fourth International Workshop on the Design of Reliable Communication Networks (DRCN), Canada, Oct. 2003

---

Article received: 2005-06-23