An IP over WDM Protocol with Routing and Transport Capabilities

Shou-Kuo Shao and Jingshown Wu
Room 519
Department of Electrical Engineering and
Graduate Institute of Communication Engineering
National Taiwan University
Taipei, Taiwan 107
R.O.C.
TEL: 886-2-23635251 ext.: 519,
FAX: 886-2-23638247,
Email: yvonneko@cctwin.ee.ntu.edu.tw , skshao@ms.chttl.com.tw

Abstract

A new IP over WDM protocol with routing and transport capabilities is proposed. This protocol utilizes the concept of label switching and wavelength information of Wavelength Division Multiplexing (WDM). This protocol uses labels and length information as the physical transport framing and data link delineation at the same time and can handle routing processes in the core network. Thus, it can speed up the processes in future terabit networks. The transport adaptation and encapsulation scheme of this protocol is simple and transparent to protocol data unit. Most of all, this protocol not only provides transmission mechanism, but also integrates the Multiprotocol Label Switching (MPLS) traffic engineering control concept as the IP transport and switching control plane. In this paper, we present the adaptation and encapsulation scheme of the proposed protocol and analyze its framing and delineation performance.

1 Introduction

While IP Backbone systems are being realized using Synchronous Optical Network / Synchronous Digital Hierarchy (SONET/SDH) and Asynchronous Transfer Mode (ATM) technologies, existing transport networks will not yield the bandwidth-effective transporting and routing schemes for IP traffics. This is because that exiting telecommunication networks are being designed for traditional telecommunications services in mind, and not specifically for IP traffic. The integration of IP and Dense WDM (DWDM) has been the most popular research and development area for next generation Internet [1].

MPLS is rapidly emerging as an Internet Engineering Task Force (IETF) standard [2,3]. MPLS intends to enhance the speed, scalability, and service provisioning capabilities in the Internet. MPLS uses the technique of packet forwarding based on labels, to enable the implementation of a simpler high-performance packet-forwarding engine. MPLS is best viewed as a 2.5th layer protocol, integrating layer 2 and layer 3 functions. With this point of view, optical label switching has been deemed as the most probable technique for the future all optical networks [1].

However, at the present time, due to the gap to success of optical elements, optical labeling is not so realizable. Other techniques using optical sub-carriers to carry the labels may suffer the optical transparent problem in all optical networks. For this sake, we need to reconsider 1) the properties of optical transmission, 2) the properties of IP traffic, and 3) future requirements for IP backbone traffic. What we need are as follows: 1) an IP over WDM transport mechanism that can handle physical framing, data link delineation and part of layer 3 routing functions at the same time. This mechanism also needs to speed up the processing, while keeps transparent to all optical networks as possible. 2) As for IP traffic properties, we need additional traffic engineering models as the control plane for IP networks. In other words, we need a high-speed asynchronous transport mechanism integrating with label switching techniques and the advantages of control plane functions. This mechanism must be able to migrate to future all optical networks smoothly.

In this paper, we propose a new IP over WDM protocol with routing and transport capabilities. This protocol utilizes the concept of label switching and integrates MPLS and WDM technologies. This protocol uses labels and length information as the physical transport framing and data link delineation at the same time, and can handle routing processes in the core network. The proposed protocol performs physical framing, data link delineation and part of layer 3 routing functions in one process. Thus, it can speed

up the processes in future terabit networks. The transport adaptation and encapsulation scheme of this protocol is simple and transparent to protocol data unit. Most of all, this protocol not only provides transmission mechanism, but also integrates the MPLS traffic engineering control concept as the IP transport and switching control plane. This protocol provides simplest and fastest framing architecture for asynchronous and variable length packet compared with [4,5]. It can migrate into future all optical networks without further modification as long as optical elements can handle label switching and checking. By applying the protection and restoration functions at IP layer, this protocol can further reduce the cost and complexity of the systems for future all optical transport networks. However, these topics including 1) network scenarios, 2) combining MPLS traffic engineering model with DWDM network elements, 3) protection and restoration at IP layer [6] and 4) network management, etc., are going to be submitted in subsequent papers and will not be presented in this paper. The performance of the proposed protocol utilizing MPLS concept as the physical layer transmission mechanism will be described in detail in the subsequent sections.

2 Transmission Format of the Proposed Protocol

The concept of the proposed IP over WDM protocol is not merely an IP over DWDM adaptation and encapsulation scheme, but also the concept of utilizing MPLS Label Distribution Protocol (LDP/CR-LDP) as the DWDM network signaling protocol. It also integrates MPLS traffic engineering control for optical network elements. However, each transport protocol does have its own framing and encapsulation method. In this section, we will first introduce the transmission format of the newly proposed IP over WDM protocol.

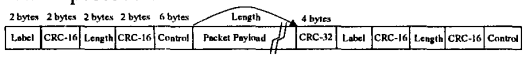

Fig. 1 Physical and Data Link Layer Frame Delineation of the Proposed IP over WDM Protocol

The transmission format of the proposed protocol is shown in **Fig. 1**. The frame encapsulation is done by using <Label> <CRC-16> <Length> <CRC-16> <Control> and followed by the protocol data unit of variable length and a <CRC-32> field as the trailer for payload CRC checking. As shown in Fig. 1, the "Label" field has two bytes and is used as the label switching function. The CRC-16 circuit with generating polynomial: $g(x) = x^{16} + x^{12} + x^5 + 1$, is used to check if the "Label" field is correct or not during searching process of re-framing. These bytes with "Label" field together form the framing indication of physical layer. The "Length" field together with the second CRC-16 bytes is used to indicate the length of the protocol data unit, and to be the data link layer frame delineation. Once the framer arrives at the "Sync" state, the CRC-16 check circuit is triggered as the single-bit error correction circuit (so does at the "Post Sync" state). So, the framer just need one CRC-16 and one logic control circuit which records the status of the framer and controls the operation of CRC-16 circuit to be multiple errors detection or single-bit error correction. This simplifies the hardware of the framer a lot. "Control" field is used to indicate the type of protocol data unit, scrambler synchronization status and maintenance information (not going to be discussed in this paper) like the functions listed in [4,5]. Since these bytes can be used in multi-purpose, we just leave them there as option bytes and may use these bytes in subsequent papers. In this paper, they are less significant than the framing and delineation bytes, so we do not intend to investigate them at the present time and will focus on the first four fields to see if this framing and delineation mechanism is feasible and outstanding than others.

In order to avoid the inefficiency of using preamble techniques as the bit synchronization method, there must have the inter-packet fills to maintain the channel continuity. Those idle inter-packet fills contain only the necessary <Label>, <CRC-16>, <Length> and <CRC-16> fields. Maintenance packets like those to deliver OAM messages will have additional <Control> and <CRC-32> fields. These OAM packets will periodically be inserted into optical channel with higher priority than the normal data packets. The optimal period of the OAM packets and related usage and functions will be given in the subsequent papers.

We propose 16 bits for labels in the proposed IP over WDM protocols. To accommodate all the addressing range, i.e., 2^{20}, and incorporate the technologies of wavelength routing, we adopt 16 different wavelength, $\lambda_0, \lambda_1, \cdots, \lambda_{15}$, as a group of MLPS path addresses. In that way, we can fully utilize the wavelength information while keeping electrical labels as small as possible. In this case, it is possible that we could integrate the MPLS technologies with wavelength routing. These similarities between wavelength and label have been used for combining MPLS traffic engineering control with optical cross-connects (OXCs) [7]. Additional need for a General Switch Management Protocol (GSMP) [8] or GSMP-like protocol to interface the optical router gear [9] may also be considered. This part of adaptation and integration with MPLS, wavelength routing, MPLS traffic engineering control is related with the network scenario of the proposed IP over WDM protocol and will be discussed in subsequent papers.

3 Analysis of Framing and Encapsulation Performance

The beauty of transport function of the proposed IP over WDM protocol, is its simplicity in the framer. To find the frame in "Sync" state at the first time that the system is powered up or at the time of re-framing after loss of frame detected, the framer is in "Hunt" state. The framer uses byte-by-byte search to find the first valid CRC-16 sequence of label field. The framing and delineation state transition diagram of the proposed IP over WDM protocol is shown in **Fig. 2**. The framing state transition diagram consists of four distinct states: "Hunt", "Presync", "Sync" and "Post Sync" states. Fig. 2 is not the same as the state transition diagram shown in [4,5] and ATM circuits [10], in which the state transition diagram has only three distinct states: "Hunt", "Presync" and "Sync" states.

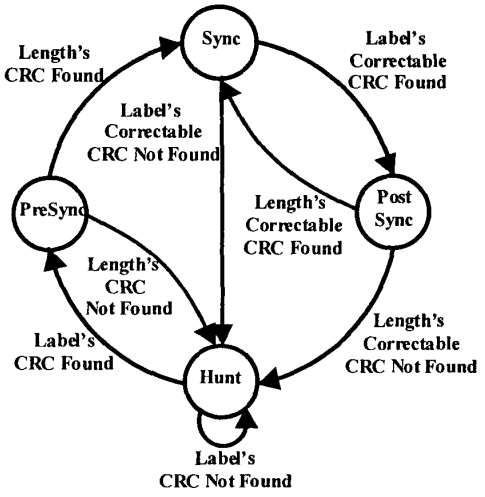

Fig. 2 Framing and Delineation State Transition Diagram

The procedure of framing and delineation is as follows: When the framer is first powered up or in the state of re-framing, the framer tries to find out the first valid Label field in a byte-by-byte shifting and checking manner. If no valid header is found, the framer continues the shifting and check processes, i.e., staying in the "Hunt" state. Once upon finding such a valid CRC-16 sequence, the framer moves to the "Presync" state, and records the status of framing. The control logic of framer, at this state, continues to perform the second CRC-16 checking for the subsequent specific 4 bytes (Length field + second CRC-16 field) to find out the second correct CRC-16 pattern of Length field. If the framer fails in finding the second valid CRC-16 pattern after subsequent 4 bytes, it then returns back to the "Hunt" state. If it finds the second valid CRC-16 sequence, the framer goes into "Sync" state. At that point, the framer is considered in frame. The framer determines from the Length field to find out how many bytes there are until the beginning of next Label field of next frame.

Once the framer gets into the "Sync" state, the control logic circuit of framer enables the CRC-16 error checking circuit into a single-bit error correction circuit. The framer now detects the CRC-16 sequence with no error bit and with one error bit pattern for the Label and Length field of next frame. If it does not find the next single-error correctable CRC sequence, it goes back to the "Hunt" state to perform the searching process again. If the framer detects the first correctable CRC sequence of Label field of next frame, it goes into "Post Sync" state (with CRC-16 as a single-bit error correction circuit). The framer then continues identifying the subsequent 4 bytes with CRC-16 as single-bit error correction circuit to see if they are correctable for the Length field. If not correctable, the framer goes back into "Hunt" state to perform shifting and checking process. If the framer finds out the correctable CRC-16 sequence of the Length field, it goes into "Sync" state again and maintains the CRC-16 circuit in single-bit error correction status. This process continues until lost of frame detected and all the re-framing processes go over again.

There are three statistics that are very important in determining the framing performance of framing algorithms:

1) Probability of loss of frame (PLF)
2) Probability of false frame (PFF)
3) Mean time to frame (MTTF)

We are going to analyze them as follows.

3.1 Probability of Loss of Frame

It is worth to note that at the receiver side, the framer is first delineated using <Label> and <CRC-16> to arrive at "Presync" state and then using <Length> and second <CRC-16> fields to get into "Sync" state. Probability of loss of frame (PLF) is directly a function of optical channel bit error rate (BER). Once the synchronization is achieved, single-bit error correction is activated. Thus, for a loss of frame, more than one error in the <Label> and <CRC-16> fields or less than one error in <Label> and <CRC-16> fields but more than one error in <Length> and second <CRC-16> fields is required. Since all these fields are 2 bytes and the single-bit error correction is enabled, the probability of the frame loss due to the channel corruption in framing and delineation is:

$$PLF = PLF_{Label} + PLF_{Length} \quad \text{.....[1]}$$

where, PLF_{Label} is the probability that loss of frame happened in checking valid CRC-16 sequence of Label field and PLF_{Length} is the probability that loss of frame happened in checking valid CRC-16 sequence

of Length field. These two probabilities are shown as follows:

$$PLF_{Label} = 1-[(1-P_e)^{32} + 32P_e(1-P_e)^{31}] \quad \text{.....[2]}$$

$$PLF_{Length} = PLF_{Label}(1-PLF_{Label}) \quad \text{.....[3]}$$

where, P_e is the optical channel BER. **Fig. 3** shows the probability of loss of frame of the proposed IP over WDM protocol.

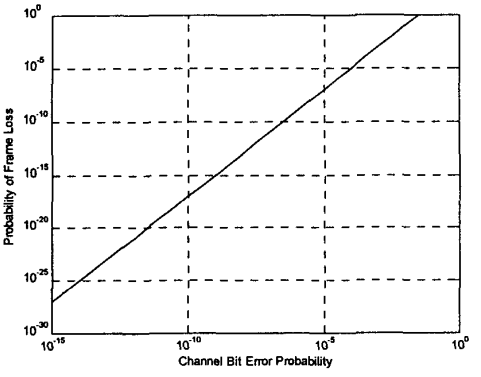

Fig 3 Probability of Frame Loss versus Channel BER

After we obtain the PLF, we can roughly estimate the "Mean Time To Frame Loss" (MTTFL) which can be expressed as follows [5]:

$$MTTFL = \frac{1}{PLF * FrameRate} \quad \text{.....[4]}$$

where, *FrameRate* is equal to *Channel Bit Rate / Frame Length*. However, since the transport is asynchronous and with variable packet length, we can't get the accurate "Mean Time To Frame Loss", unless we know the packet length distribution. We only can roughly estimate this parameter with a given packet length. **Fig. 4** shows the graph of *MTTFL* (in Years) versus packet length (in Bytes) at optical channel bit rate 10Gbps and BER 10^{-12}. We can see *MTTFL* even at very small packet length distribution can still get almost a billion years!

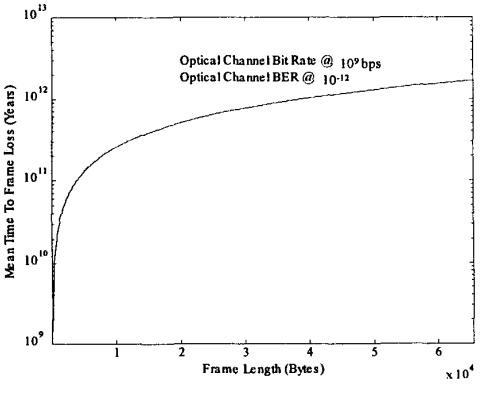

Fig. 4 Mean Time To False Frame versus Packet Length

3.2 Probability of False Frame

When parsing the frame byte-by-byte in search of a valid header, it is possible that the valid CRC-16 pattern is found inside the protocol data unit. This is the case that the CRC-16 header is simulated by the payload data. And the probability of false frame (*PFF*) is equal to the probability of two consecutive CRC-16 matches and is shown as follows:

$$PFF = \left(\frac{1}{2^{16}}\right)^2 = (2^{-32}) = 2.33\text{x}10^{-10} \quad \text{.....[5]}$$

Even if "false framing" happens, the framer does not stay in "Sync" state for the coming frames unless it finds more consecutive two CRC-16 matches at the right places. It could be an exact match or a match with only one bit in error. This means that the probability of falsely staying in "Sync" state is *Prob* (false sync) = $(33\text{x}2^{-16})^2 \approx 25\text{x}10^{-8}$. This makes the probability of rejecting false frame becomes 1 - $25\text{x}10^{-8}$ and is equal to 0.99999975. From this, we can see that the probability of rejecting staying in false "Sync" state of the proposed protocol is nearly equal to 1.

3.3 Mean Time To Frame

Re-frame refers to the transition from "Hunt" state to "Sync" state as shown in Fig. 2. Re-frame time, or mean time to frame (MTTF), is the duration of time required for a frame alignment circuit to re-establish frame alignment from being powered up or from a misalignment condition. A re-frame procedure consists of two separate processes: search and confirmation [11]. A search process sequentially searches the CRC-16 sequence for Label header to identify the beginning of each packet. If there is a match, the position is provisionally accepted as a candidate and a confirmation process is called in to test validity of the candidate position. If the confirmation criterion is met with the candidate position, frame alignment is declared. If not, a search process is reinitiated, followed by another confirmation process and so on.

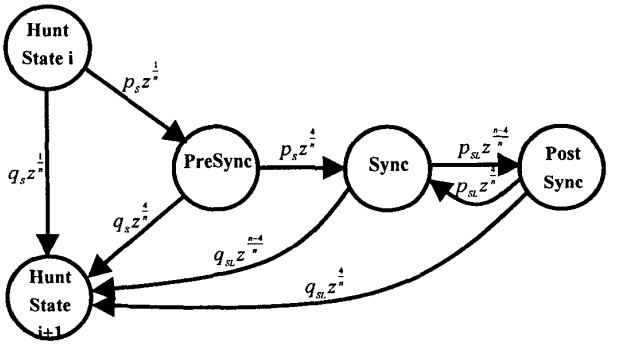

Fig. 5 Falsehood Detection Process in Search Process

3.3.1 MTTF Search Process

To consider the worst case, a search process is initiated just one byte after the correct frame alignment position. Therefore, all the n bytes in each packet, where n is variable for each packet, have to be checked. For the first $(n\text{-}1)$ byte position, a search process takes time to verify them wrong. If the payload simulates the CRC-16 sequence, a confirmation process is called in to detect its falsehood. After the falsehood is detected, the search process is re-initiated and the next byte position is checked. At the n *th* byte position the search process accepts it.

Using the technique of probability of generating function (PGF) in the generalized state transition diagram of Sitter [12] and also in [11], we can derive the state transition diagram of search process as shown in Figure 5. The search process moves from the byte position i to $i+1$ in one byte duration. Since the proposed protocol is a packet based framing and delineation architecture, the more meaningful units in qualifying the MTTF duration is in packets. So, we denote one byte duration as $1/n$ packet, where n is the packet length in bytes. If the CRC-16 pattern of Label header is simulated, the confirmation process takes $4/n$ packet to detect its falsehood. The probability of CRC-16 pattern is simulated by the payload is p_s and is equal to $1/2^{16}$. The probability of not being simulated as the CRC-16 pattern is q_s and is equal to $1-p_s$. If the simulation continues happening, the framer will go into the false framing state and the probability of staying in "Sync" state and "Post Sync" state is p_{SL} which is equal to $33\times p_s$ and the probability of back to "Hunt" state is q_{SL} which is equal to $1-p_{SL}$. From **Fig. 5**, we then can show that the PGF of the state transition from state i to $i+1$ is given by:

$$q_S Z^{\frac{1}{n}} + P_S Z^{\frac{1}{n}}[q_S Z^{\frac{4}{n}} + p_S Z^{\frac{4}{n}} q_{SL} Z^{\frac{n-4}{n}} + p_S Z^{\frac{4}{n}} \frac{p_{SL} Z^{\frac{n-4}{n}}}{1-p_{SL}Z} q_{SL} Z^{\frac{4}{n}}]$$

$$\cong q_s Z^{\frac{1}{n}} + P_s q_s Z^{\frac{5}{n}} \quad \text{.....[6]}$$

The maximum mean time to frame requires the falsehood detection for the first $(n-1)$ byte position and a correct CRC-16 pattern for the last n *th* byte position. Therefore if the correct CRC-16 pattern is not corrupted by optical channel bit errors, the search process takes the signal transfer function $\tau(Z)$ to the right position, where $\tau(Z)$ is given by:

$$\tau(Z) = [q_s Z^{\frac{1}{n}}(1+P_s Z^{\frac{4}{n}})]^{n-1} Z^{\frac{1}{n}}$$

$$= q_s^{n-1} Z (1+P_s Z^n)^{n-1} \quad \text{.....[7]}$$

3.3.2 MTTF Confirmation Process

After the correct position of CRC-16 pattern is found, the confirmation process is called in. **Fig. 6** shows the state transition diagram of the reframe declaration procedure. After the correct position is found, the probability that the framer goes into "Presync" state is P_D. And P_D is equal to the $(1-P_e)^{32}$, where P_e is the optical channel BER. The probability that the CRC-16 pattern is corrupted by the channel noise is q_D, which is equal to $1-P_D$.

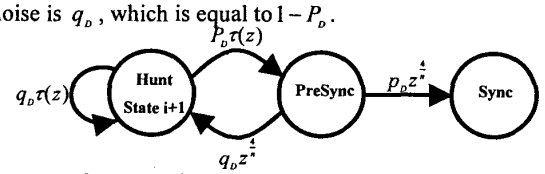

Fig. 6 Reframe Declaration Procedure

The overall transfer function from the "Hunt" state to "Sync" state is given by

$$P_{RF}(Z) = \frac{p_D^2 \tau(Z) Z^{\frac{4}{n}}}{1 - q_D \tau(Z)[1 + p_D Z^{\frac{4}{n}}]} \quad \text{.....[8]}$$

which is the PGF of maximum average MTTF. Therefore, the maximum average MTTF, T_{RF} is given by

$$T_{RF} = P'_{RF}(1)$$

$$= \frac{1}{p_D^2}[\tau'(1) + \frac{4}{n} p_D] \quad \text{.....[9]}$$

Fig. 7 (a) to **(d)** show the maximum MTTF with packet length 65535 bytes, 1500 bytes, 354 bytes and 576 bytes, respectively. These packet lengths are typical lengths found in the proposed protocol and in [1,4]. The packet length of 1500 bytes is the maximum transfer unit (MTU) of Ethernet and the length of 576 bytes is the length that all routers must support.

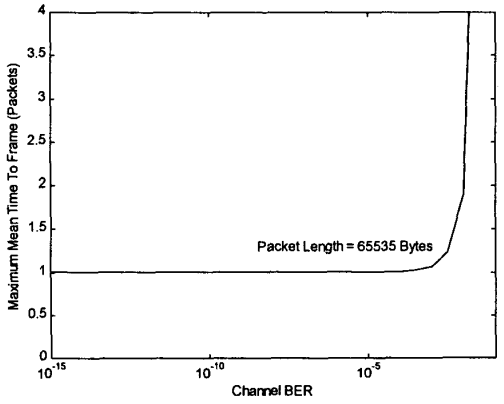

Fig. 7(a) Maximum Mean Time To Frame versus Channel BER with Packet Length 65535 Bytes

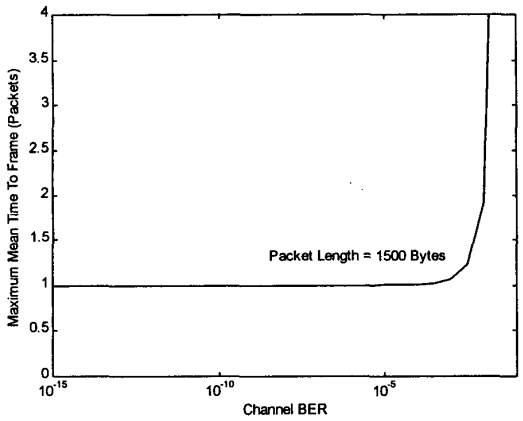

Fig. 7(b) Maximum Mean Time To Frame versus Channel BER with Packet Length 1500 Bytes

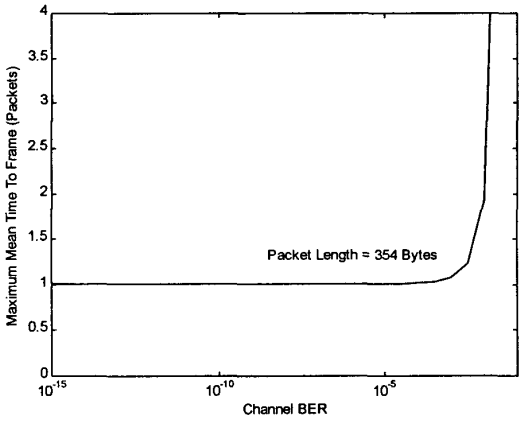

Fig. 7(c) Maximum Mean Time To Frame versus Channel BER with Packet Length 354 Bytes

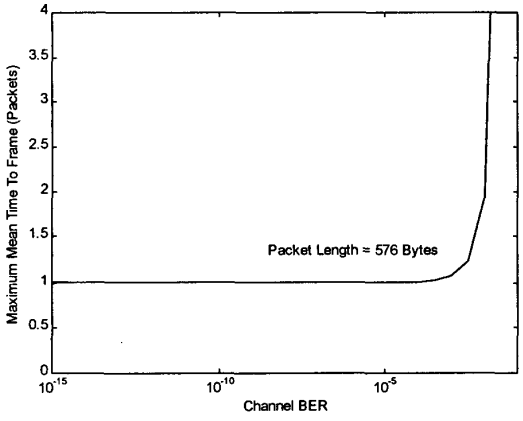

Fig. 7(d) Maximum Mean Time To Frame versus Channel BER with Packet Length 576 Bytes

From Fig. 7 (a) to (d), we can see that when optical channel BER is less than 10^{-5}, the maximum average MTTF is almost equal to one frame. This is quite straightforward, since the proposed protocol is actually in "Sync" state in less than one packet duration. The maximum average MTTF will be at most one packet to get into "Sync" state. This is much faster for the framing of variable packet length than in [4,5]. Each different length packet arriving at "Sync" state is almost the same at channel BER less than 10^{-5}. We can see that although the packet length is a random variable and will affect the accuracy of actual MTTF results, we still can conclude that the MTTF will be almost less than one packet.

4 Conclusion

In this paper, we present a new IP over WDM protocol, which is asynchronous and variable packet length. This protocol is suitable for high-speed WDM of IP transport networks. The framer architecture is simple and gets satisfactory performance. The proposed protocol is transparent to the protocol data unit and can support multi-protocols. It is also transparent to all optical networks with minimum delay. This protocol can migrate to future all optical networks without any further modification as long as optical elements can handle label switching and error checking. The proposed protocol integrates the concept of MPLS and wavelength routing. This protocol can perform physical framing, data link delineation and part of layer 3 routing functions in one process. It is very suitable for IP over WDM transmission.

5 Acknowledgement

The authors want to thank to the support of the National Council of Science under Grant NSC 89-2215-E-002-014.

6 Literature

[1] EURESCOM Project P918-GI, "Integration of IP over Optical Networks: Networking and Management," Deiverable 1, Oct. 1999.

[2] R. Callon, P. Doolan, N. Feldman, A. Fredette, G. Swallow and A. Viswanathan, "A Framework for Multiprotocol Label Switching," Internet Draft, Work in Progress, Sep. 1999.

[3] E. C. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture," Internet Draft, Work in Progress, Aug. 1999.

[4] B. T. Doshi, S. M. Dravida, E. J. Hernandez-Valencia, W. A. Matragi, M. A. Qureshi, J. Anderson and J. S. Manchester, "A Simple Data Link Protocol for high-Speed Packet Networks", *Bell Labs Technical J.*, pp85-104, Jan.-Mar. 1999.

[5] Bijan Raahemi, "Comparison of Frame Delineation Performance Between 10GE WAN PHY and 8B/10B," Nortel Networks' submission to IEEE802.3 HSSG, ver. 1.0, Nov. 1999.

[6] R. D. Doverspike, S. Phillips, and J. R. Westbrook, "Future Transport Network Architecture," *IEEE Commun. Magazine*, vol. 37, issue 8, pp.96-101, Aug. 1999.

[7] D. O. Awduche, Y. Rekhter, J. Drake, R. Coltun, "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects," *Internet Draft*, Work in Progress, Nov. 1999.

[8] Tom Worst, "General Switch Management Protocol," *Internet Draft*, Work in Progress, Oct. 1999

[9] John Crowcroft, "IP over Photons: How not to Waste the Waist of the Hourglass," *IEEE Conference IWQoS '99, Seventh International Workshop on Quality of Service,* pp.9-11, April 1999.

[10] ITU-T Recommendation I.432, "B-ISDN user-network interface – Physical layer specification: General characteristics", Aug. 1996.

[11] DooWhan Choi, "Frame Alignment in a Digital Carrier System – A Tutorial," *IEEE Commun. Magazine*, vol. 28, issue 2, pp.47-54, Feb. 1990.

[12] R. W. Sitter, "Systems Analysis of Discrete Markov Processes," *IRE Trans. on Circuit Theory*, Vol.CT-3, pp.257-266, Dec 1956.