Optical Burst Switching Protocols in Next Generation Core Networks

Benon K Muwonge, *Student Member, IEEE,* and Anthony H Chan, *Sr Member, IEEE*

***Abstract*—Optical Burst Switching is the transition technology between the current electronic switching, and future all-optical switching. This paper presents proposals on three main protocols in Optical Burst Switching (OBS): Assembling, scheduling, and reservation protocols. The proposed Assembling protocol, unlike previous ones, assembles traffic with a Forwarding Equivalence Classification (FEC) method, and is strongly dictated by the value of the Hurst parameter, H, a measure of burstiness. The proposed Core Aggregation with Void Filling (CA-VF) scheduling protocol adds a number of fields to the Burst Header Packet (BHP) which enables "burst aggregation" in the core network. CA-VF uses the proposed Optical Congestion Notification (OCN) an optical equivalent of Backward/Forwarding Congestion Notification (BCN/FCN). The Just Enough Time (JET) protocol is used here with slight modifications to allow core "burst aggregation".**

***Index Terms*—Optical Burst Switching, Payload Burst, Hurst parameter, H, Burst Header Packet, BHP.**

I. INTRODUCTION

THERE is a rapid convergence of data and telecommunication services worldwide. This has had implications on how data is transported. The existing electronic switching technologies, SONET/SDH and ATM, which are widely used today, are fast reaching their limits. This is happening because users are demanding more bandwidth, and new applications are requiring faster transmission and more QoS than SONET/SDH or ATM can provide. Coupled with the rapid growth of Internet users in recent years, there has been a vast renewed interest in optical switching. It is widely accepted that optical switching is the alternative to cope with the unprecedented growth in data services worldwide.

OBS is a compromise between electronic and all-optical packet switching, which has not yet matured. In an OBS network, there exist two types of routers; Label Edge Routers (LER), at the edge of the network and Label Switch Routers (LSR) in the core network. Packets arriving into an OBS network are assembled at the LER into bursts which contain the payload data. The payload burst is then buffered electronically, while a Burst Header Packet (BHP) is transmitted ahead. The BHP reserves the required wavelength resources at each Label Switch Router (LSR) in the core and is processed electronically at every LSR. The aim is to avoid opt-elec-opt (O-E-O) conversion of the payload burst. The payload burst is therefore transmitted only in the optical domain. During transmission the wavelength is only held for the duration of the transmission of the payload burst. Fig.1 shows the basic operation of OBS.

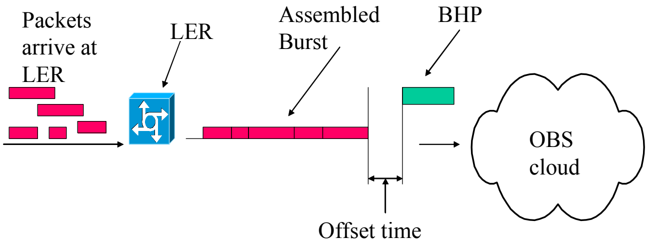

Fig.1. Basic operation of OBS. The BHP has an offset time between itself and the assembled payload burst, which caters for the time taken for electrical processing, and O-E-O conversions.

OBS will be the transition technology from electronic switching, and a future all-optical network, which may still be a number of decades away. OBS is also the most promising technology using currents technologies, to offer a transition to IP over Dense Wavelength Division Multiplexing (DWDM).

The issues in OBS are mainly in resource contention resolution, burst loss at both the ingress and mostly at the core nodes. QoS support for different applications is also another important issue in OBS. As such, there have been many proposals for algorithms that assemble, reserve or schedule payload bursts to address these issues.

Yu et al proposed several assembling algorithms in [1] that theoretically assemble efficiently. Of the proposed algorithms, the min-max assembly algorithm performs best, by assembling bursts of different lengths with a maximum time delay restriction. The algorithm though, does not take into account the different delay tolerances of the different classes of traffic. JET a reservation protocol, is proposed by Yoo and Qiao [2]. Several other reservation schemes are proposed by Xiong et al [3] and by Turner [4], but JET still outperforms them in many respects. The Horizon and Latest Available Unused Connection with Void Filling (LAUC-VF) scheduling

---

Manuscript received May 3 2005. This work was supported in part by the Broadband Centre of Excellence with funding from Telkom, Siemens and the National Research Foundation, South Africa.

B. K. Muwonge is with the University of Cape Town currently doing a masters research in optical networks. 20 Edwin Hart, UCT, Rhodes Avenue, Mowbray, Cape Town, 7700 (e-mail: benon@crg.ee.uct.ac.za)

H. Anthony Chan is professor at the Department of Electrical Engineering, University of Cape Town, Private Bag, Rondenbosh 7701, South Africa (e-mail: h.a.chan@ieee.org).

algorithms are most popular. The LAUC-VF has a higher throughput but substantially slower than Horizon. The proposed scheduling algorithm in this paper uses void filling as the LAUC-VF, and in addition, allows for traffic aggregation.

This paper presents an OBS network model, which can be used for IP over DWDM. The model is optimized for OBS, by proposing new assembling, scheduling and reservation algorithms.

It is assumed throughout this paper that full wavelength conversion is available at each core node, and no FDLs at each of the core nodes.

The rest of this paper is as follows: In section II, scale assembly protocol that uses FEC criteria for grouping traffic is presented. It takes into account the burstiness of the incoming traffic. In section III, A core aggregating scheduling protocol that aggregates core payload bursts is proposed. In section IV an optical congestion notification scheme is proposed. In section V the packet header fields for the proposed scheduling protocol is outlined. In section VI, a model of how the OBS technology may be used is suggested. Lastly, in sections VII and VIII, future work and conclusion of the analysis and proposals is presented.

II. ASSEMBLING OF PACKETS

A. The max-min assembly algorithm

The max-min assembly algorithm sets both a minimum and maximum lengths of the bursts. This is with the assumption that the traffic being assembled is constant-bit-rate traffic. This assumption results in a number of limitations in its performance:

1) Internet traffic is bursty and self-similar, as found in [5]. In cases when the bandwidth usage is low, traffic would be unnecessarily delayed, due to the minimum length restriction. This would affect low delay applications, such as real time video traffic.

2) By fixing one maximum burst length, the assembly does not take into account the different applications that tolerate different delays. Considering different delays for different traffics, would enable different maximum delays for different applications.

Solutions to these short comings would have to consider more factors that affect bursty traffic.

B. Factors affecting burst Assembly

An efficient assembly algorithm is one that assembles packets appropriately, i.e. using FEC, with no packet loss. By definition of self-similarity, aggregated packets must not overlap. To attain efficient assembly of IP packets, the assembled packets must obey this rule. In assembling the packets at the LER, several factors must be considered:

1) Maximum delay tolerance, $\mathbf{t_D}$*, of a given application:*

There is delay incurred by traffic from upper layers, before assembly, $\mathbf{t_d}$. This delay varies, depending on the protocols used. At the point of assembly, it is important for the assembler to determine the remaining delay tolerance, $\mathbf{T_d = t_D - t_d}$, of the application traffic. To determine $\mathbf{t_d}$, at the point of assembly, the TTL field in the IP packet can be used for this purpose. $\mathbf{t_d}$ is determined by knowing the starting value of the TTL, then subtracting from it the current TTL value at the point of assembly.

2) Destination of traffic:

We define two types of destinations: $\mathbf{d_1}$, the egress node, and $\mathbf{d_2}$, the final destination of the assembled traffic. By knowing $\mathbf{T_d}$, and the destination $\mathbf{d_2}$ of the traffic being assembled, traffic may be assembled, by considering either the egress node destination $\mathbf{d_1}$, or type of traffic.

3) *Length of payload burst.*

The length of the payload burst $\mathbf{l_{pb}}$, depends on the nature of traffic being assembled, its rate of arrival $\mathbf{r(t)}$, and the value of $\mathbf{T_d}$. By assembling traffic to a desired length determined by $\mathbf{T_d}$, and nature of traffic, the lengths of $\mathbf{l_{pb}}$, will vary. The different lengths of $\mathbf{l_{pb}}$, will be indicative of the load at the LER.

4) Hurst, H, parameter of traffic, at point of assembly:

The Hurst parameter in Internet traffic ranges between, $0.5 \leq H \leq 1$, which exhibits Long Range Delay, (LRD). The higher H is, the shorter the $\mathbf{l_{pb}}$, and the lower H is, the longer $\mathbf{l_{pb}}$, can be.

When these factors are considered together, QoS of the traffic being transmitted and end-to-end burst transmission can be guaranteed in the core network.

C. A proposed scale assembly method

Fig 2 shows the proposed method of assembly of IP packets at the ingress node.

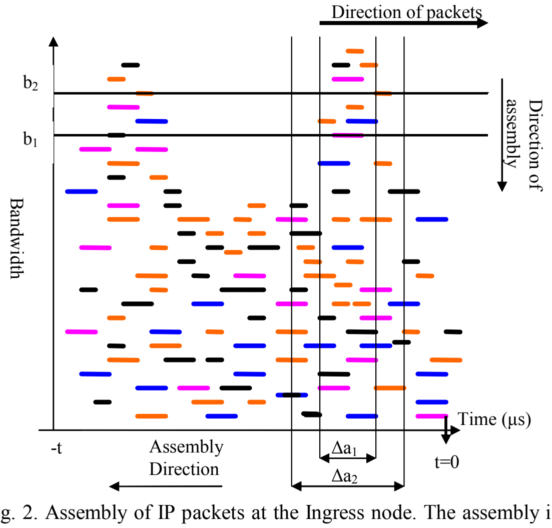

Fig. 2. Assembly of IP packets at the Ingress node. The assembly is done to by the definition of self-similarity. This type of assembly results in FEC.

At **t=0**, packets arrive at the ingress node for assembly. At **t=0**, the assembly algorithm starts to assemble the incoming traffic.

$\mathbf{\Delta a_1}$ and $\mathbf{\Delta a_2}$ are the maximum delay tolerances $\mathbf{T_d}$, of different applications. The values $\mathbf{\Delta a_1}$, $\mathbf{\Delta a_2}$, are up to $\mathbf{\Delta a_n}$, where n is the number of known different applications. This implies that the assembler has intelligence of the expected applications. This allows for differentiated application services. $\mathbf{b_1}$ and $\mathbf{b_2}$, represent the bandwidth at which each packet is assembled. Assembly goes from **t=0**, to an arbitrary **–t**.

The $\Delta\mathbf{a}$ scales all begin at **t=0**, and overlap for different applications. Within each $\Delta\mathbf{a}$ scale, other factors such destinations, $\mathbf{d_1}$ and $\mathbf{d_2}$ are considered.

At every time interval $\Delta\mathbf{a}$, the value of **H** is calculated. By knowing **H**, the expected length $\mathbf{l_{pb}}$, for a given application and FEC can be calculated. This combination of factors would result in a much more efficient algorithm.

D. *Outline of scale assembly protocol*

Starting at t=0, algorithms of different scales $\Delta\mathbf{a}$ run simultaneously, each seeking out packets to assemble.

```
1  Set scale $\Delta a_n$ (n = 1, 2, 3 ---n)
2  Calculate the value of H in each given scale.
3  If packets contain real data, then
     If in scale $\Delta a_n$, packets have
     $\Delta a_n \le T_d$ and $\Delta a_n$ is small then
     Assemble packets into burst and route.
4  If in scale $\Delta a_n$ packets are
     $\Delta a_n \ge T_d$ & d1 and $d_2$ are the same
     Assemble packets into burst if and only if $l_{pb} > l_{pbmin}$.
     Where, $l_{pbmin}$ is the minimum set length for non-real traffic.
5  else If in scale $\Delta a_n$ packets are
     $\Delta a_n > T_d$ and $d_1$ and $d_1$ is the same, then
     Assemble packets into burst, if and only if $l_{pb} > l_{pbmin}$
6  start from step 1.
```

III. CORE AGGREGATION SCHEDULING ALGORITHM

OBS is a one way protocol. Efficient scheduling of the assembled bursts is needed, to avoid burst drop and bandwidth contention in the core network. The LAUC-VF algorithm, mentioned before, and others proposed after it, still encounter core resource contention. We propose a Core Aggregation algorithm with Void Filling, CA-VF that promises better results.

The CA-VF algorithm uses the same scheduling as LAUC-VF, but adds an aggregation feature. By aggregating payload bursts, core resources are more efficiently utilized.

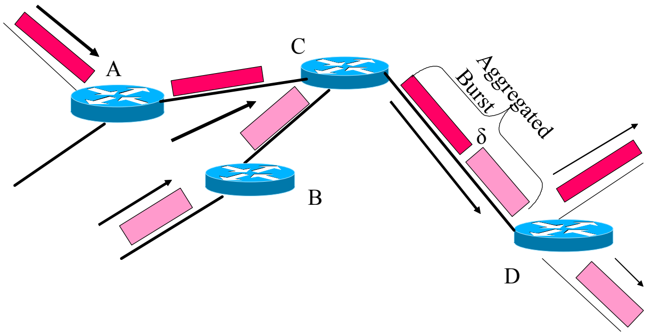

Fig. 3. Payload bursts arriving at LSRs A and B are routed to C. At C, the Bursts are aggregated depending on their time difference, δ, destination, d_1, or d_2, and/or the resources available at node C. The bursts may separate at D depending on their destinations.

Fig 3 shows the concept of core aggregation.

Consider payload bursts at A and B, above, of different lengths. BHPs of each payload burst reach C and give information of their respective bursts. Therefore, before the payload bursts arrive at C, their lengths, $\mathbf{l_{pb}}$ destinations, $\mathbf{d_1}$, $\mathbf{d_2}$, and time differences $\boldsymbol{\delta}$ are known. By knowing $\mathbf{l_{pb}}$, $\mathbf{d_1}$, $\mathbf{d_2}$, and $\boldsymbol{\delta}$, payload bursts may be aggregated when, the conditions below hold:

1. When the routes being taken by the two bursts are the same for one or more nodes.
2. When $\boldsymbol{\delta}$ is very short, such that, no significant time lapses while the wavelength is held for the next burst/s to be aggregated.

At node D, payload bursts may separate depending on their final destinations, in which case, two BHPs will be generated at D for each of the payload bursts.

By aggregating payload bursts, the number of BHPs, reduce. With less BHPs, there are less resources being requested, thereby lessening the resource contention in the core network.

Even with the availability of core resources, payload bursts may be dropped when they arrive at a node before resources have been allocated. This would normally be a result of diminished offset times due to long processing times. BHP processing times are typically low, about 1ns, but tend to increase with increase in number. Therefore, reducing the number of BHPs helps reduce the probability of burst drop.

IV. CORE NETWORK CONGESTION CONTROL

One of the advantages of OBS is in the great simplification of the core network. OBS is intended to have minimal management and therefore cheap to maintain. The immaturity of all optical processing, demands some electrical intelligence at LSRs.

We propose a BCN/FCN equivalent for optical networks to reduce complex signaling schemes used in SONET. We call this Optical-Congestion-Notification, OCN. Fig 4 shows a section of a meshed core network, and how OCN functions.

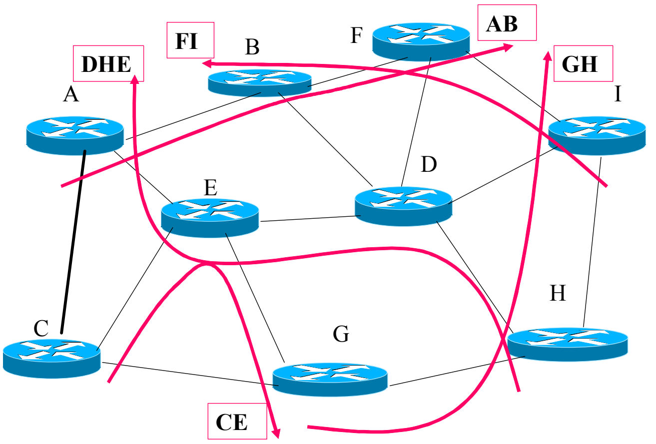

Fig. 4. The paths indicated, show routes taken by BHP packets. Passing each LSR, a BHP packet records the status of each LSR at that time. On reaching the next node, this information is read. The LSRs then use this information to route traffic accordingly.

The paths indicated, represent the different routes that may be taken by BHPs while reserving resources. Consider one of the above routes through nodes H, D, E and A. When the header packet reaches H, information of the available wavelengths at the time of processing is inserted into the header. On reaching D, this information is read , and information about D is also added to the header packet. The same process happens at E. At A, information about the route HDE is in the header packet.

The obvious setback of this is the accumulating information in the BHPs in a dense core network. To avoid this, header packets can be set to only carry LSR state information for a limited number of hops. This does not reduce the efficiency of OCN, since all other header packets in the core do congestion notification.

There are two extreme cases that arise from OCN.

1) When there are many header packets to a node.

When there are many header packets to a node, then more information about the neighboring nodes is known. This then better helps in making routing decisions.

2) When there are very few header packets to a node.

In this case, this then signals to the node that the neighboring nodes are at low capacity. Bursts are then transmitted to the low capacity nodes.

The criteria for a node choosing or rejecting a route to a node, is based on available wavelengths and the delay that results in choosing a given route. This results in traffic in the core network being balanced. As a result, the efficiency of OBS is increased.

V. BHP FIELDS

To enable core aggregation, and OCN extra fields in header packets must be added to those proposed in [3] must be added to header packets. Fig 5 shows the fields in a BHP packet without aggregation.

l_{pb}	**T_a**	**T_d**	**t_x**	**d_1**	**d_2**

Fig. 5. The BHP shows fields appended at the ingress node, on entering the core network.

The descriptions of the individual fields are given below.

l_{pb} – Length of the payload burst. The burst lengths are determined at the ingress nodes. The BHP in this design does not have wait for a burst to be assembled, for **l_{pb}** to be known. The length is known (approximated) using the H value, and method of FEC used. The time at which the BHP is transmitted is then the only modification in the JET protocol.

T_a – Time of arrival of the payload burst. Bandwidth is only reserved at the (expected) point of arrival.

T_d – The maximum delay tolerance of the payload burst. The value of T_d is real time and decreases with propagation delay. This value is used at both the ingress nodes and egress nodes. It helps in determining how long a burst should be held before transmission.

t_x – The QoS offset as described in [6]. The difference in this application, is that **t_x** is not fixed, but varies according to a given application, or FEC classification.

d_1,d_2 – Destinations to egress and final destination respectively. The need for the two separate fields rather than one is due FEC. During FEC payload bursts may be classified with respect to their egress nodes.

With core aggregation the information in the two individual header packets must be put be put in one BHP. Fig 6 shows a BHP packet that reserves resources for an aggregated burst.

l_{pb1}	**l_{pb2}**	**T_{a1}**	**T_{a2}**	**T_{d1}**	**T_{d2}**	**t_{x1}**	**t_{x2}**	**D_c**

Fig. 6. The BHP shows fields appended at the in the core network by an LSR, the node at which burst aggregation occurs.

The descriptions of individual fields are given below.

l_{pb1} l_{pb2} – The lengths of the aggregated bursts, 1 and 2.

T_{a1} T_{a2} – The times of arrival (expected) of the two aggregated bursts. The value of **δ** is then obtained from the difference of the two.

T_{d1} T_{d2} – The remaining tolerable delays for each of the bursts.

t_{x1} t_{x2} – The QoS offsets for each of the bursts.

D_c – The core node for separation of aggregated burst.

The aggregated bursts in Fig3**,** separate at node D. The number of hopes that aggregated bursts may transverse is determined by the egress destinations, and the availability of resources. Not shown in Fig 6, are the final burst destinations **d_1 d_2**. These remain the same as in Fig 5.

With the fields defined above, then core aggregation using LAUC-VF basic reservation scheme can be done as outlined below.

```
1   On arrival of header packets BHP1 and BHP2, the fields $T_a$ $l_{pb}$ $T_d$ and $t_x$ are processed.
2   If $T_{a1} - T_{a2} << l_{pbmin}$
      And destinations $d_1$ and $d_2$ transverse more than one node in the same direction, then
3   Aggregate bursts with destination D.
4   else If $T_{a1} - T_{a2} > l_{pbmin}$
    Do not aggregate.
5   Go to step 1
```

VI. NEXT GENERATION CORE NETWORK WITH OBS

The main goal of the Next Generation Network, NGN, is to reduce cost, and at the same time keep up with the users' demands.

Today, telecommunication companies are spending a lot of resources and money on interworking of different applications. OBS has the capacity to receive packets of all types of media, and transmit them at high data rates. This is because, OBS uses wavelengths to transmit data, and all applications can be done with one suitable modulation scheme. In addition the transmission rates of OBS are much higher than those of ATM, and SONET.

With the proposed scale assembly and scheduling algorithms, the protocols between the IP and physical layers can be bypassed, thus reducing cost. The existing protocols between the IP and physical layer can however, still be used with the OBS protocol in the core network. This further makes OBS both a transitional technology and future technology for the core network.

By using OBS this in effect implements IP over DWDM, a two layer networking model. The two layer model will be a transition from the current five layer model.

The Fiber-to-the-Home, FTTH, concept, where homes have optical fiber right at home, is also realizable. This is due to the much reduced costs, and reliable QoS. Fig 7 illustrates how OBS can be used in the NGN.

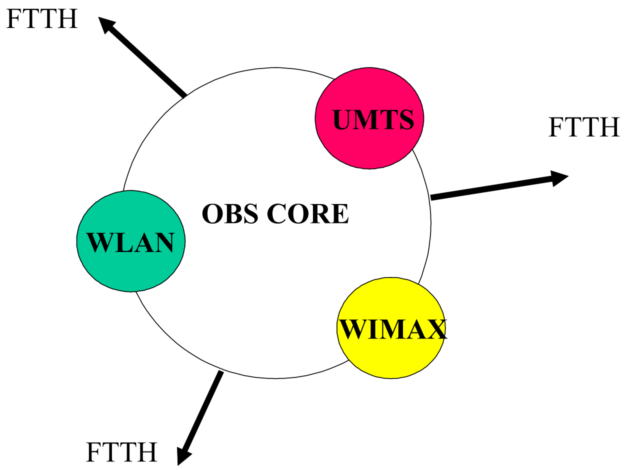

Fig. 7. The OBS core network is able to interwork with other existing technologies and at the same time be directly used by the end user.

Fig 7 shows how OBS may be used in the transition from an electrical, to an all-optical network. The ability of OBS to support end users directly also makes it possible to be used for FTTH.

VII. FUTURE WORK

Future work in the given proposals will validate the models and algorithms by simulations, using an appropriate simulating optical network tool.

VIII. CONCLUSION

OBS is a flexible technology that enables the transmission of data in the core network at an applications' required data rate, and QoS. In this paper, we have proposed two new assembling and scheduling algorithms, the Scale assembly and Core Aggregation protocols. To avoid signaling in the core network, we have proposed an Optical Congestion Notification scheme.

REFERENCES

[1] X. Yu, Y. Chen, and C. Qiao,. "Study of traffic statistics of assembled burst traffic in optical burst switched networks," *.Proceeding of Opticomm,* pp 149-159, 2002.

[2] M. Yoo and C. Qiao, "Just-Enough-Time (JET): A High Speed Protocol for Bursty Traffic in Optical Networks," *IEEE/LEOS Technologies for a Global Information Infrastructure,* pp.26-27, Aug 1997.

[3] Y. Xiong, M. Vandenhoute, and H. Cankaya, "Control architecture in optical burst-switched WDM networks," *IEEE Journal on Selected Areas in Communications*, 18:1838-1851, October 2000.

[4] J. Turner, "Terabit burst switching. Journal of High Speed Networks,,"

[5] W. Willinger, et. al., "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," *Proceedings of SIGCOMM '95*, pp 100--113, Boston, MA, 1995.

[6] M. Yoo and C. Qiao, "A New Optical Burst Switching Protocol for Supporting Quality of Service," *SPIE Proc. of Conf. All-optical Networking,* vol 3531, pp 396-405, 1998.

Benon K. Muwonge is a postgraduate student in the Department of Electrical engineering, University of Cape Town.

H Anthony Chan (M'94–SM'95) received his PhD in physics at University of Maryland, College Park in 1982 and then continued post-doctorate research there in basic science.

After joining the former AT&T Bell Labs in 1986, his work moved to industry-oriented research in areas of interconnection, electronic packaging, reliability, and assembly in manufacturing, and then moved again to network management, network architecture and standards for both wireless and wireline networks. He had designed the Wireless section of the year 2000 state-of-the-art Network Operation Center in AT&T. He was the AT&T delegate in several standards work groups under 3rd generation partnership program (3GPP). During 2001-2003, he was visiting Endowed Pinson Chair Professor in Networking at San Jose State University. In 2004, he joined University of Cape Town as professor in the Department of Electrical Engineering.

Prof. Chan is Administrative Vice President of IEEE CPMT Society and had chaired or served numerous technical committees and conferences. He is distinguished speaker of IEEE CPMT Society and is in the speaker list of IEEE Reliability Society since 1997.