Burst-mode transmitter / receiver sharing for VM migration of cloud edge

Kyota Hattori\(^{1a}\), Masahiro Nakagawa\(^2\), Masaru Katayama\(^2\), and Jun-ichi Kani\(^3\)

\(^1\)NTT Technology Planning Department, NTT Corporation, 1-5-1 Otemachi, Chiyoda-ku, Tokyo, 100-0004 Japan
\(^2\)NTT Network Service Systems Laboratories, NTT Corporation, 3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan
\(^3\)NTT Access Network Service Systems Laboratories, NTT Corporation, 1-1, Hikarinooka, Yokosuka-shi, Kanagawa, 239-0847 Japan

\(^a\) kyota.hattori.vj@hco.ntt.co.jp

Abstract:

Network virtualization technology such as cloud edges (CEs) will contribute to the cost-effective construction of carrier networks while maintaining saving of hardware and bandwidth resources. In this letter, we propose a future metro network architecture based on Next-Generation Passive Optical Network Stage 2 technologies to support CEs. This makes it possible to reduce the number of transmitters and receivers for the accommodation of both user traffic and traffic between CEs cost-effectively by sharing burst-mode transmitters and burst-mode receivers between CEs. We execute the preliminary evaluation of network capital expenditure with the proposed method by numerical simulation.

Keywords: Optical passive metro network, NG-PON2, sharing burst-mode transmitters and burst-mode receivers

Classification: Network system

References

[1] ETSI, “GS NFV001 V1.1.1,” Oct. (2013).
[2] A. Misawa, K. Mochizuki, H. Tsuchiya, M. Nakagawa, K. Hattori, M. Katayama, and J. Kani, “Virtual Edge Architecture with Optical Bandwidth Resource Control,” IEICE Trans. Commun., Vol. E99.B, No. 8, pp. 1805-1812, Aug. 2016.
[3] K. Hattori, T. Homemoto, M. Nakagawa, N. Kimishima, M. Katayama, and A. Misawa, “Optical Layer 2 Switch Network with Bufferless Optical TDM and Dynamic Bandwidth Allocation,” IEICE Trans. Electron., Vol. E99-C, No. 2, pp. 189-202, Feb. 2016.
[4] ITU-T Rec. G989.2, “40-Gigabit-capable passive optical networks 2 (NG-PON2): PMD layer specification,” (2014).
[5] A. V. Tran, C. J. Chae, and R. S. Tucker, “Bandwidth-Efficient PON System for Broad-Band Access and Local Customer Internetworking,” IEEE Photon. Technol. Lett., Mar. 2006.

[6] K. Hattori, M. Nakagawa, T. Matsuda, M. Katayama, and K. Koda, “Passive Optical Metro Network based on NG-PON2 with Sharing Burst-mode Receiver between Continuous-mode and Burst-mode Transmitters to Support Cloud Edges,” Proc. ECOC, W.4.P1.SC6.64, Sept. 2016.

[7] J. Chen and L. Wosinska, “Analysis of Protection Schemes in PON Compatible with Smooth Migration from TDM-PON to Hybrid WDM/TDM PON,” OSA J. Opt. Netw., Vol. 6, Issue 5, pp 514-526, May 2007.

[8] EU FP7 OASE D4.2.2, June 2013.

[9] P. Öhlen, B. Skubic, A. Rostami, M. Fiorani, P. Monti, Z. Ghebretensae, J. Martensson, K. Wang, and L. Wosinska, “Data plane and control architectures for 5G transport networks,” J. Lightw. Technol., Vol. 34, no. 6, pp. 1501-1508, Mar. 2016.

1 Introduction

Network function virtualization [1] enables the application of virtualization technologies to carrier networks (NWs) such as virtual switches to reduce capital expenditure (CAPEX) through effective use of hardware and bandwidth resources.

(a) Architecture of proposed future metro NW

(b) Summary of the proposed method (Logical NW)

Fig. 1. Proposed future metro NW.
This is especially progressing for further development of expensive edge equipment such as cloud edges (CEs) [2]. A key technology facilitating this will be virtual machine (VM) migration according to resource usage of CEs. Therefore, future metro NWs will have to accommodate the traffic, whose directions and volume dynamically change, with supporting VM migration of CEs. To achieve this, we previously proposed the Optical Layer-2 Switch NW (OL2SW-NW) [3], which allows bandwidths in metro NWs to be shared with optical time division multiplexing (TDM) paths according to traffic volume. The links of OL2SW-NW consist of optical TDM SWs. Therefore, when the OL2SW-NW is applied to areas of low traffic demand, the cost of OL2SW-NW will become relatively high due to the optical TDM SWs.

In this letter, we propose a future metro NW based on Next-Generation Passive Optical Network stage 2 (NG-PON2) technologies [4], whose optical passive devices are used in the links, to reduce NW CAPEX even in areas of low traffic demand. The proposed NW enables a burst-mode (BM) transmitter (BMT) for VM migrations to share between BM receivers (BMRs) of channel terminations (CTs) in different optical line terminals (OLTs).

2 Method to share a BMT to support cloud edges

2.1 NW architecture

Fig. 1(a) shows the architecture of proposed future metro NW as an extended implementation of NG-PON2. This is a bufferless wavelength division multiplexing and TDM single ring NW which provides variable bandwidth to logical paths not only between access NWs and CEs but also between the CEs. This NW can perform Dynamic Bandwidth Allocation (DBA) for both traffic from CEs ($T_C$) and traffic from access NWs ($T_A$). As a method of DBA for metro NWs, we introduced two cycles to increase bandwidth utilizations [3]. To reduce NW CAPEX, the links of proposed NW consist of optical passive devices except for amplifiers. Here, we label the nodes: the ones connected to the access NW and the CE are called “A-node” and “C-node”. Each node has ether-burst-converters (EBCs) that convert between Ethernet frames and BM signals according to allocated timeslots (TSs). To apply NG-PON2 technologies, the C-node’s EBC is based on an OLT CT because C-nodes have to aggregate traffic. The A-node’s EBC is based on an optical network unit (ONU). The proposed NW has a master node called “M-node”. The M-node allocates TSs for logical paths and controls logical connections between access NWs and VMs by managing externally connected media access control (MAC) addresses of VMs. C-nodes have to communicate both A-nodes and other C-nodes. Therefore, when applying the original NG-PON2, each C-node has to prepare additional continuous-mode (CM) transmitters (CMTs) and receivers (CMRs) according to the number of C-nodes ($N_C$) to receive CM signals from other C-nodes and deploy L2SWs to avoid collisions of data from A-nodes and other C-nodes (We define this as a conventional method) [5]. This may increase NW CAPEX. For this issue, we proposed the method of sharing a BMR between CMTs and BMTs [6]. However, this requires the improvement of OLT CT’s transmission functions to attach burst...
headers according to destinations of the TSs. This leads to an increase in CAPEX for the development to design the hardware circuits of OLT CT.

### 2.2 Proposed sharing BMT method

Fig. 1(b) shows the summary of the proposed method. The proposed method can enable a BMR to receive both $T_d$ and $T_c$ accommodating both resource information and virtual images of CEs transmitted from hypervisors. This is realized by applying an ONU with a BMT to the C-node for VM migrations (ONU-C), making it possible to reduce the number of optical interfaces (IFs) and L2SWs at C-nodes while diverting configurations of the NG-PON2. The ONU-Cs are controlled by the OLT CT at the M-node (CT-M) to avoid collisions between BM signals from A-nodes and C-nodes. The CT-M executes the DBA by collecting the traffic information and measuring the delay between ONU-Cs and sets both MAC-tables and TS-tables to the ONU-Cs. Since the link of the proposed NW consists of passive devices, undesired BM signals arrive at the BMR of OLT CT on the way to the desired destined OLT CT. Therefore, optical gate switches (OSWs) are installed in front of the OLT CT and the ONU-Cs. However, the proposed method has the disadvantage that the efficiency of bandwidth utilization at the BMTs deteriorates if propagation delays between C-nodes are not integral multiples of TS length as described in Fig. 1(b). Fig. 2 shows the configuration of the C-node and the M-node. The C-node can send TSs to the desired destination and receive the desired TSs by controlling the OSW according to the TS-table.
2.3 Sequence of proposed method

The proposed method consists of four steps;

1. Setting local time to each node: The CT-M sets the local time ($T_L$) and synchronizes the clock frequency ($F_{ck}$) to each ONU-C. This is the same way as PON. Simultaneously, the ONU-Cs set the $T_L$ and synchronize the $F_{ck}$ to the OLT CT in the same C-node. Then, the OLT CTs set the $T_L$ and synchronize the $F_{ck}$ to A-nodes. Meanwhile, the A-nodes and the C-nodes send the $T_L$ back to the C-node and the M-node, respectively.

2. Delay measurement between C-nodes: The CT-M measures propagation delays between each C-node and the M-node based on $T_L$ from the C-nodes. The propagation delay between C-nodes is calculated by subtracting the measured delay between the M-node and the C-node from each other.

3. Execution of DBA: The M-node executes the DBA according to the amount of $T_A$ and $T_C$. Then, the M-node sets allocated TSs to each C-node as a TS-table. Meanwhile, the C-node allocates TSs to the A-nodes belonging to this C-node using the TSs excluding the allocated TSs for C-nodes.

4. Switching CE for VM migrations: When a VM migration occurs, the C-node which detects the new MAC address of the VM sends this information to the M-node. Then, the M-node updates and sets new MAC-tables and TS-tables to C-nodes and A-nodes. Finally, the C-nodes and A-nodes simultaneously change the bandwidth and timings of TS transmissions, with the C-nodes changing the On/Off of OSWs and the A-nodes changing the wavelength according to new MAC-tables and TS-tables.

2.4 Mathematical expression for the CAPEX

We define the C-node$j$’s CAPEX in the proposed method as $C^P_j$ and that in the conventional method as $C^C_j$. $C^P_j$ and $C^C_j$ are expressed as follows:

$$C^P_j = \left( \frac{C_{OLT-CT}}{2} + C_{Coupler} + C_{BMR} \right) \left[ B_{TS} \left( \sum_{i \in C\text{-node}, i \neq j} T_{i,j} B_{TS} \right) + \sum_{k \in A\text{-node}} T_{k,j} B_{TS} \right]$$
$$+ (C_{ONU} + C_{CMR} + 2C_{OSW} + 3C_{Coupler} + 2C_{Isolator} + C_{BMT}) \left[ B_{TS} \sum_{i \in C\text{-node}, i \neq j} T_{i,j} B_{TS} \right], \quad (1)$$

$$C^C_j = \left( C_{LSW} + C_{CMR} \right) \sum_{i \in C\text{-node}, i \neq j} T_{i,j} B_{TS} + \left( C_{LSW} + C_{CMR} \right) \sum_{i \in A\text{-node}, i \neq j} T_{i,j} B_{TS}$$
$$+ \left( \frac{C_{OLT-CT}}{2} + C_{CMR} \right) \sum_{k \in A\text{-node}} T_{k,j} B_{TS} + C_{AMP}.$$  

Here, $\lceil \cdot \rceil$ represents the ceiling function and $T_{i,j}$ is a matrix indicating the average traffic volume from node$i$ to node$j$. $B_{TS}$ and $B_W$ are the capacity of a TS and a
wavelength. $C_{component}$ is the relative cost of each component. The terms of Eq. (1) have two kinds of ceiling functions. The outer one represents the unit of wavelength. The inner one represents the unit of TS. The first and second terms of Eq. (1) mean the effectiveness of the proposed method; a BMR can receive both $T_A$ and $T_C$ and a BMT can send $T_C$ to each C-node.

### 3 Preliminary evaluation of NW CAPEX

We evaluated NW CAPEX based on Eq. (1) and (2). We assumed that NW CAPEX is dominated by that of C-nodes. Therefore, we define the ratio of NW CAPEX ($R_C$) as $C^O/C^C$. To evaluate the applicable area for the proposed method, we changed $N_C = 2-10$ and $T_C = 1-10$ Gbps. The relative cost of each component was set in reference to [7,8], and other parameters were set in reference to [3] (Fig. 3(a)). We assumed that $C_{Isolator}$ is equal to $C_{Coupler}$. Each method suitably placed amplifiers according to power budgets. The propagation delay between each node was set to 5 km (25 μs), which is a non-integral multiple of TS as the disadvantage condition of the proposed method. Fig. 3(b) shows the $R_C$ when changing $N_C$ and $T_C$. The smaller the accommodated $N_C$ and the greater the increase in $T_C$, the less effective sharing BMTs and BMRs becomes. Conversely, the greater the accommodated $N_C$, the less efficient bandwidth utilization of TSs from the ONU-C caused by the disadvantage of the proposed method as described in Fig. 1(b) becomes. According to these relationships, $R_C$ at minimum became 0.48 at $N_C = 6$, $T_C = 1$ Gbps. This means the proposed method could reduce NW CAPEX by 52% at maximum. It found that the proposed method can achieve cost-effectiveness under the condition that $N_C$ is below 6 or $T_C$ is below 5 Gbps at the number of A-nodes ($N_A$) = 40; which is sufficient scale for current metro NWs [9].

### 4 Summary

We proposed the future metro NW architecture based on NG-PON2 technologies, which enable a BMT for VM migrations to share between BMRs of different OLT CTs. A simulation result showed that the proposed NW could achieve reduction of NW CAPEX by a maximum of 52%.