Demonstration of time-wavelength co-allocation (TWCA) problem in novel dynamic wavelength scheduled WDM-PON for distributed computing applications

Min Zhu, Wei Guo, Shilin Xiao, Benoit Geller, Weiqiang Sun, Yaohui Jin, Weisheng Hu

Demonstration of Time-Wavelength Co-Allocation (TWCA) Problem in Novel Dynamic Wavelength Scheduled WDM-PON for Distributed Computing Applications

Min Zhu[ab], Wei Guo*[a], Shilin Xiao[a], Benoit Geller[b], Weiqiang Sun[a], Yaohui Jin[a], Weisheng Hu[a]
[a]State Key Lab of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, China,
[b]Laboratory of Electronics and Computer Engineering, École Nationale Supérieure de Techniques Avancées (ENSTA), 32 Boulevard Victor 75739 Paris Cedex 15, France

ABSTRACT

This paper studies the problem of the implementation of distributed computing applications in local area networks. We propose a novel Dynamical Wavelength Scheduled Wavelength Division Multiplexing Passive Optical Network (WDM-PON) architecture, in which the number of the available upstream wavelength channels is greatly less than that of optical network units (ONU). And we experimentally demonstrate the feasibility of the proposed WDM-PON, which enables dynamically scheduling upstream data in the time division multiplexing (TDM) and WDM scheme from ONU to optical line terminal (OLT). The Time-Wavelength Co-Allocation (TWCA) Problem is defined in the proposed WDM-PON architecture to aggregate large files for distributed computing applications and three scheduling algorithms are presented to solve this problem. The significant improvement compared with the conventional TDM-over-WDM PON is illustrated through simulations.

Keywords: Optical Network, Distributed Computing Applications, WDM-PON, Scheduling

1. INTRODUCTION

In local areas with the scope of dozens of kilometers, such as a campus or a science and technology park, geographically distributed computing and storage resources are involved in large-scale distributed e-scientific computing applications such as geosciences, biomedical informatics, and nuclear physics [1][2][3]. The distributed computing applications consist of a set of tasks with dependent relationships. These tasks are allocated to different computational and storage resource nodes for computing, analyzing and other processing. Thus, each geographical distributed computational and storage resource node has the potential to generate extremely large capacity transmission such as gigabytes or terabytes size [4][5]. These data should be aggregated again to the remote supercomputer at data center for final data-processing and visualization within several seconds or several tens of seconds. We consider Passive Optical Network (PON) [6][7] as an optimal local area network infrastructure to satisfy the requirements of multi-gigabits network connection with low loss, low latency, and minimal jitter. However, existing Wavelength Division Multiplexing Passive Optical Network (WDM-PON) [8][9] architectures require expensive wavelength specified optical sources in ONUs so that it hasn't been largely deployed. As for the existing Time Division Multiplexing Passive Optical Network (TDM-PON) [10][11], the traffic loads from different ONUs must be upstream-transmitted in one wavelength channel with a Time Division Multiple Access (TDMA) scheme and the upstream wavelength per ONU is fixed. Due to high traffic burstiness and unbalanced traffic pattern in distributed computing applications, the utilization of wavelength becomes poor and transfer latency is very large with the above two PON architectures.

In this paper, we propose a novel Dynamical Wavelength Scheduled WDM-PON architecture, in which the number of the available upstream wavelength channels is greatly less than that of optical network units (ONU). The proposed PON architecture enables dynamically scheduling upstream data from different ONU to OLT in the combined TDM and WDM scheme for real-time data aggregation in distributed computing applications. It would not only employ the same

* wguo@sjtu.edu.cn; phone 86 21 34204597; fax 86 21 34204597;

ONU equipped with a small-scale wavelength-tuneable laser, but also allows all ONUs to share all wavelength resources across the PON. Thus, each ONU can obtain upstream bandwidth on demand in the wavelength or finer sub-wavelength granularity and better quality of service (QoS).

Meanwhile, to improve effectively the overall PON system throughput and to minimize the transfer latency for data aggregation, we consider the problem of bandwidth reservation for dedicated wavelength channels and for future time slots, where a wavelength channel is operated in a TDM fashion. We define the above problem as a Time-Wavelength Co-Allocation (TWCA) Problem to schedule these large files over their respective wavelengths and timeslots as soon as possible. We will prove that TWCA is NP-complete, propose three scheduling algorithms to solve this problem. And finally we evaluate the proposed PON system performance compared with the conventional TDM-over-WDM PON under different traffic loads condition.

The rest of the paper is organized as follows. Section 2 presents the proposed WDM-PON topology and parameters. Experimental demonstration is given in Section 3 to verify the total connectivity between OLT and ONUs. In section 4, we model the data aggregation problem as the Time-Wavelength Co-Allocation (TWCA) Problem and propose three scheduling algorithms based on some sorting schemes to solve it. The numerical results are presented and discussed in Section 5. At last, conclusions are given in Section 6.

2. DYNAMIC WAVELENGTH SCHEDULED WDM-PON

The proposed dynamical wavelength scheduled WDM-PON architecture has N ONUs, N downstream wavelength and m upstream wavelength (m<N), as shown in Fig.1. Multiple Wavelength Laser (MWL) and WDM receiver are deployed in the OLT for virtual point-to-multipoint downstream connection and upstream reception, respectively. The Remote Node (RN) consists of an arrayed waveguide grating (AWG), an optical coupler (OC) and N optical circulators, where OC is used to combine different upstream wavelengths λ_{11}~λ_{1m} from their respective ONU to the OLT. Each ONU is equipped with a small-scale wavelength-tuneable laser for dynamic upstream-wavelength-scheduled transmission and a fixed receiver for downstream data.

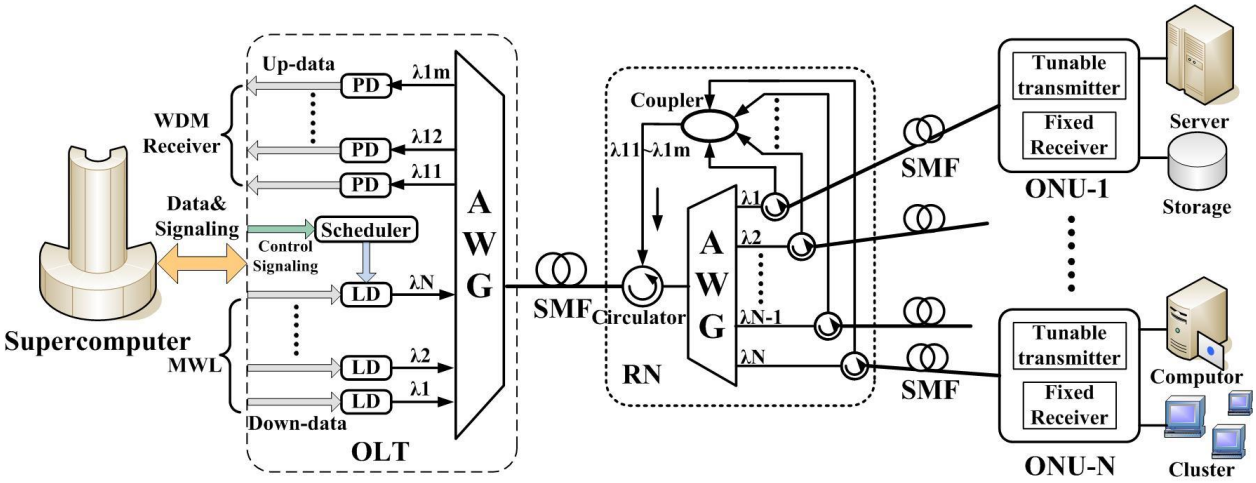

Fig.1 Schematic diagram of the proposed dynamic wavelength scheduled WDM-PON

The centric supercomputer is attached to the OLT, and other computing and storage resources are located in their respective ONUs. The proposed WDM-PON provides a multi-gigabits network connection between the OLT and ONUs for large data transfer in distributed computing applications.

The center scheduler in OLT has an overall knowledge of the amount of data that each ONU will send by the last round of polling scheme. In the scheme, every ONU reports the amount of data to be transferred to the OLT in up-controlling signalling (usually referred to as a Report message in Ethernet-PON protocol). Based on the overall information, the OLT scheduler decides to schedule upstream data on which wavelength channel and at which time slots for each ONU, in order to minimize the completion time of the upstream transmission. Then the scheduling results of wavelength

channel and timeslots (usually referred to as a Grant message in Ethernet-PON protocol) are carried on the each downstream WDM wavelength λ_1~λ_N and transmitted to the respective ONU. Upon reception of the Grant message, the ONU adjusts electronically its wavelength-tuneable laser to send both up-data and up-controlling signalling in designated wavelength channel and timeslots.

The transmission of both up-data and up-controlling (Report) signals are separated in time domain as shown in Figure 2. The up-data are followed closely by up-controlling signal using the same wavelength channel for each ONU. The up-controlling (Report) signals are received by the supercomputer at the data center, and then fed back to the centric scheduler in the OLT for scheduling processing. The up-controlling signalling (Report) contains the amount of data to be transferred from ONUs to the OLT.

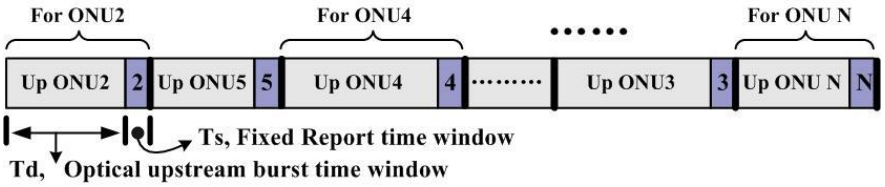

Fig. 2 Time-Division-Multiplexing for up-data and up-controlling (Report) signals

Figure 3 shows the wavelength allocation plan used in the experiment network, and divides the C-band in two, with one half carrying downstream channels (λ_1~λ_N) and the other half carrying upstream channels (λ_{11}~λ_{1m}). The two bands are separated by a relatively broad guard band.

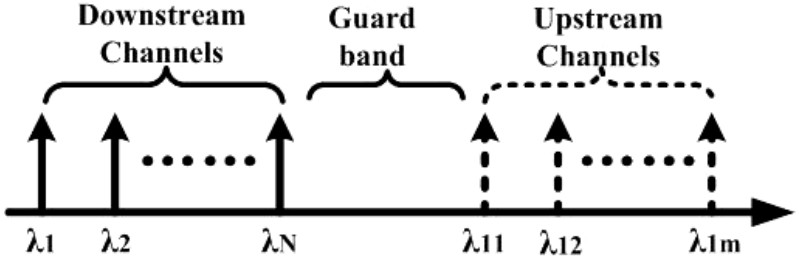

Fig. 3 Wavelength Allocation for up- and down-stream and control signaling

The dynamic wavelength scheduled WDM-PON offers some attractive features. First, compared to the conventional WDM-PON where a pair of down and upstream wavelengths for each ONU are employed, identical ONUs equipped with a small-scale wavelength-tuneable laser at each node avoid expensive wavelength specified optical sources in ONUs and difficult maintenance. Secondly, the dynamic wavelength scheduled scheme satisfies the requirements of the high traffic burstiness and of the unbalanced traffic pattern in distributed computing application, which brings division multiplexing scheme into the conventional TDM-PON and consequently has higher upstream transmission capacity. Thirdly, it is both a cost-effective and a practical solution for high-level distributed computing application by making a trade-off between TDM-PON and WDM-PON.

3. EXPERIMENTAL SETUP AND RESULTS

In downstream (OLT-ONU) direction, the proposed WDM-PON architecture with N ONUs and N downstream wavelength channels has virtual point-to-multipoint downstream connection as same as conventional WDM-PON. The novelty of the proposed WDM-PON is that, in upstream (ONU-OLT) direction, it schedules upstream data both in time division multiplexing (TDM) and WDM fashion and has flexible wavelength or sub-wavelength upstream bandwidth. So we perform an experiment to only demonstrate the effectiveness of TDM and WDM of upstream wavelength, to verify the operation principle of the proposed WDM-PON, as shown in Fig.4.

In the ONU, two upstream data carriers at the same wavelength (λ_1=1556nm) are fed into two single drive Mach-Zehnder modulators (MZMs) in separate time slots, which are respectively modulated by a 2.5-Gb/s (2^{31}-1) pseudorandom binary sequence (PRBS), generated by a pulse pattern generator (PPG). The pulse pattern generator (PPG)

was set to a "zero substitute" mode to generate a burst packet consisting of 408 bits of random data followed by 616 '0's. Then the two 2.5-Gb/s non-return-to-zero (NRZ) data streams are amplified by an erbium-doped fibre amplifier (EDFA) and filtered by a tuneable band-pass filter (BPF) in the two upstream paths, respectively. After transmission of 12.5km single mode fibres (SMF), the two data streams are combined with an optical coupler. The TDM waveforms as shown in Fig. 4a)~c) confirm that contention among the up-data streams traffic can be avoided by proper scheduling. The combined data streams are then received by a photodiode (PD) at the OLT after another 12.5km SMF. For the WDM of upstream wavelengths, another upstream data carriers at the wavelength channel (λ_2=1548.8nm) in another upstream optical paths is also add into this experiment setup in the same way as above. The spectra of multiplexing signals are inserted in Fig.4d) to demonstrate WDM of λ_1=1556nm and λ_2=1548.8nm upstream wavelengths.

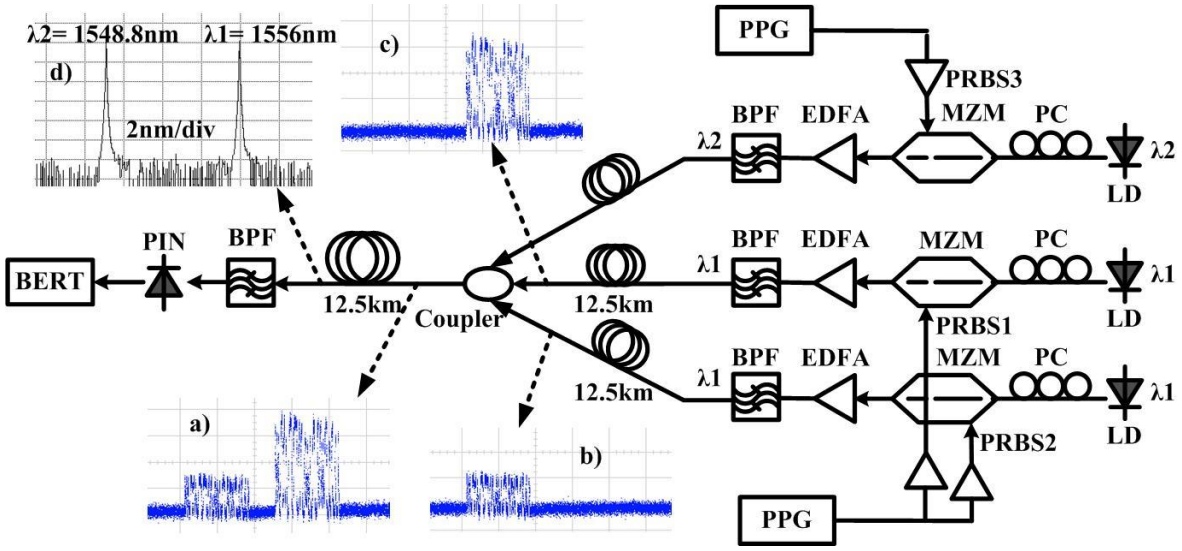

Fig. 4 Experimental setup and the performance of TDM and WDM of upstream wavelength λ1and λ2

The bit-error-rate (BER) performance and the eye diagram of the NRZ upstream data before and after the 25kms SMF are shown in Fig. 5 and the small power penalty is only about 0.8 dB.

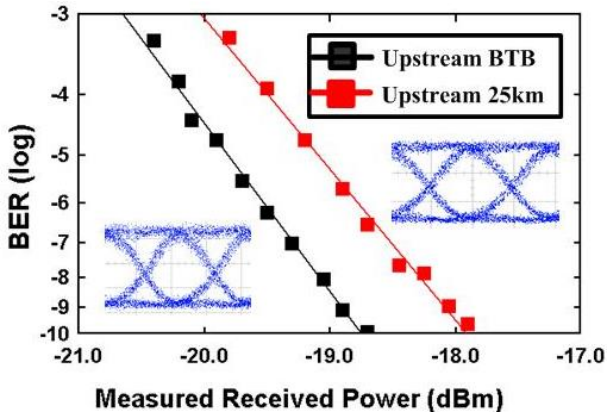

Fig. 5 BER curves and eye diagrams for NRZ upstream data

4. SCHEDULING ALGORITHM

We assume that the OLT scheduler has an overall knowledge of the amount of data that each ONU will send by last round of polling communication between the OLT and ONUs. The objective of our proposed scheduling algorithm is to determine 1) the upstream wavelength channel on which is the file transfer, and 2) the time slots at which a connection must be reserved for the corresponding file, in order to make the transmission of many large data files from each ONU to

the centric supercomputer (OLT) as soon as possible. Each file f_i is characterized by its file size S_i, its transfer time T_i, its release time R_i that is a time at which a file becomes ready for transfer, and its maximum transfer delay deadline D_i. Transfer time T_i is estimated as $T_i = S_i / BW + P_i$, where BW denotes the bandwidth capacity of wavelength channel, and P_i is the ONU-to-OLT propagation delay for file f_i. We assume that each file is large enough to make the value of S_i /BW much greater than that of P_i, so we ignore the ONU-to-OLT propagation delay P_i. This means that the transfer time T_i only depends on the file size S_i.

We define the above problem as Time-Wavelength Co-Allocation (TWCA) Problem in WDM-PON. We term N independent files originating from ONU as N independent jobs to be processed, and term M available identical wavelength channels as M parallel machines, because of the same bandwidth capacity (say B=2.5-Gigabit). This is just the famous Multiprocessor Scheduling Problem (MSP) $P_m \parallel C_{max}$, which is a NP-hard problem [12], where P_m denotes identical machines in parallel, C_{max} denotes makespan (the maximum completion time to process all jobs), defined as $\boldsymbol{max}(C_1,\ldots\ldots,C_n)$, which is equivalent to the completion time of the last job to leave the system.

So we propose three greedy algorithms to solve the TWCA problem based on list scheduling algorithm [13]. First, we reorder N independent files based on the Longest File First scheme (LFF) and Random scheme in the list, respectively. The Longest File First scheme (LFF) is to place the shorter files toward the end of the schedule and to balance transmission load on all channels. Secondly, we take files in turn from the above list to allocate it to suitable wavelength channel using three greedy algorithms, such as The Least File-Load first (LFL), The First-Fit first (FF) and The Best-Fit first (BF). To show the performance of our greedy algorithms, we perform simulations using three greedy algorithms compared with the conventional TDM-over-WDM PON, which has fixed upstream wavelength channel for each ONU.

4.1 The Least File-Load first (LFL)

This greedy algorithm is based on the cognition that the largest file (having the largest transfer time) is the bottleneck for scheduling, because it requires more resources in terms of the amount of time required to be free on the links to be transferred. The LFL algorithm aims at scheduling the largest files at first so that they get the priority to be scheduled earlier. LFL first assigns the M largest files to the M wavelength channels. After that, whenever a wavelength channel is free, the next largest file among those not yet processed is put over the wavelength channel, until all the files have been processed.

4.2 The First-Fit first (FF)

The idea of this greedy algorithm is that an optimal solution for TWCA problem is to balance totally all files into all wavelength channels. So FF firstly establishes a lower bound (LB) on the maximum completion time as an optimal solution for TWCA. The LB is then simply calculated as shown in Eq. (1), which is the average transfer time of N independent files on the M wavelength channels.

错 误 ！不 能 通 过 编 辑 域 代 码 创 建 对 象 。

(1)

Then each file is taken in turn from the above list and allocated to the first fit wavelength channel onto which it will fit without exceeding the LB. If the file doesn't fit this wavelength channel because of exceeding this value LB, then the file try to be allocated to the second wavelength channel. If the file does not fit onto any channel, it is allocated to the least loaded channel. The above steps are repeated until all the files have been processed.

4.3 The Best-Fit first (BF)

This heuristic is a slight variant of FF. This greedy algorithm is based on the intuition that putting a file into least load wavelength channel without exceeding the LB will be a best choose for scheduling. BF similarly allocates each file out of the file list in turn, but instead of allocating it to the first channel on which it will fit, it is allocated to the channel on which it will fit 'best', which has least load among non-empty channels (without exceeding the LB) and have enough gap for the new file within the LB. If the file does not fit onto any channel with exceeding this value LB, it is allocated to the least loaded channel. The above steps are repeated until all the files have been processed.

5. NUMERICAL RESULTS

To show the benefits of dynamic scheduling of upstream wavelength, we perform simulations by comparing the proposed WDM-PON with a conventional TDM-over-WDM PON in terms of the maximum completion time of

upstream data transmission from all ONUs to the OLT. The simulations are performed in a 16-ONUs WDM-PON. In the case of TDM-over-WDM PON, they are grouped by 4 TDM-sub-PONs each containing 4 ONUs and a fixed upstream wavelength, and in the proposed WDM-PON, 16 ONUs share 4 upstream wavelengths by using a tuneable laser. The traffic demand from each ONU follows a negative exponential distribution, and the slot duration and the data rate are assumed to be 2s and 2.5-Gb/s, respectively. The network load, which is defined as the ratio of the mean number of traffic slots over the total traffic slots and interval slots, is variable from 0.1 to 0.9. It simulates different amount of files from different ONUs in the upstream transmission.

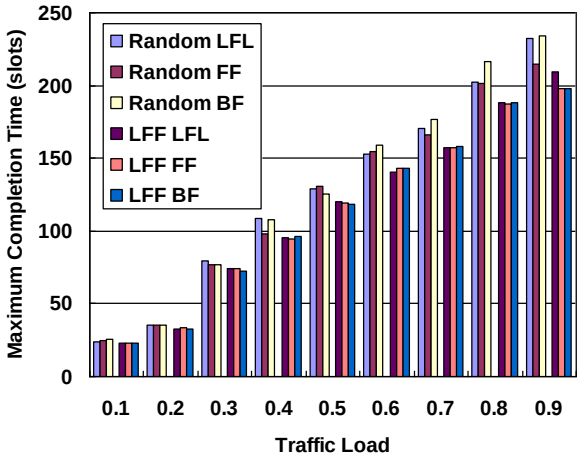

Fig. 6 Performance comparison of the LFF scheme with Random scheme for three scheduling algorithms

Fig.6 compares the performance of the three greedy algorithms both in the case of the Longest File First (LFF) scheme and the Random scheme. As we expect, since the LFF scheme schedules larger files earlier on a channel and then fills up the gap left with the smaller jobs, leads files of different sizes to be effectively combined on a channel and finally well balances transmission load on all channels. Less maximum completion time is exhibited than when using Random scheme.

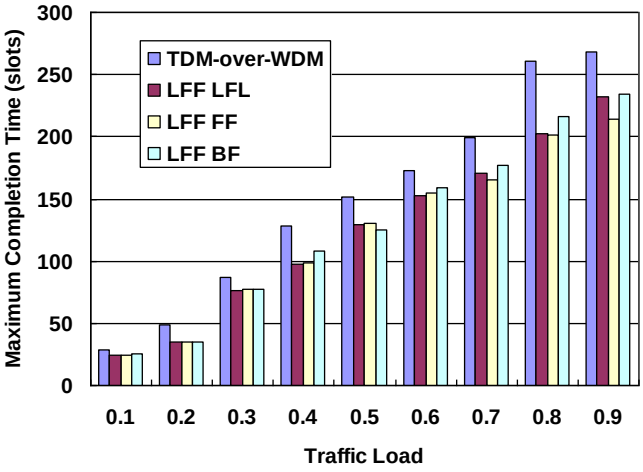

Fig. 7 Performance of the three scheduling algorithms in the proposed WDM-PON compared with the conventional TDM-over-WDM PON based on the LFF scheme

Figure 7 shows the performance of the three scheduling algorithms based on the LFF scheme in the proposed WDM-PON, compared with the conventional TDM-over-WDM PON. The maximum completion time does not vary much with the network load for LFL, FF and BF algorithms, but it significantly outperforms the conventional PON. This is primarily due to the sharing and flexible scheduling of wavelength resource among all ONUs. We also note that BF is clearly more complex than FF, but, surprisingly, BF performs slightly worse than FF. The difference is more perceptible when the network load becomes heavier. An interesting trend is the performance of LFL compared with FF. For the lower network load, LFL performs slightly better than FF. As the network load becomes heavier, FF's performance also

improves, and eventually surpasses LFL. This is possibly attributed to the following reason. When the network load is lower, the number of files is small, and FF could not find well-balanced solutions among all available channels. With the increasing network load, the number of files of different sizes is much more than that of channels, which helps in equally distributing files into all channels. Therefore, we conclude that LFF-FF algorithm should be the preferred heuristic algorithm.

6. CONCLUSIONS

In this work, we propose and experimentally demonstrate a novel WDM-PON architecture for the implementation of distributed computing applications in local area networks. The proposed Dynamical Wavelength Scheduled WDM-PON offers a cost-effective and practical solution by making a trade-off between TDM-PON and WDM-PON, in which the number of the available upstream wavelength channels is greatly less than that of optical network units (ONU). The proposed WDM-PON enables dynamically scheduling upstream data from ONU to OLT both in flexible wavelength or sub-wavelength bandwidth. It satisfies the requirements of the high traffic burstiness and unbalanced traffic pattern in distributed computing applications. We define the Time-Wavelength Co-Allocation (TWCA) Problem for large data aggregation from geographical distributed ONUs in the proposed WDM-PON and present three scheduling algorithms to solve this problem. The significant improvement in the maximum completion time is compared with the conventional TDM-over-WDM PON under different traffic loads through simulations.

ACKNOWLEDGEMENT

This work is jointly supported by National Natural Science Foundation of China (NSFC) grant No.60672016, No. 60632010, and No. 60572029, and the National “863” Hi-tech Project of China.

REFERENCES

1. T. DeFanti, C.D. Laat, J. Mambretti, K. Neggers and B.St. Arnaud, “TransLight: A Global-Scale LamdaGrid for e-Science,” IEEE Communications of ACM, 46(11), pp. 34–41(2003)
2. W. Guo, “Distributed Computing over Optical Networks,” in Proceedings of IEEE Optical Fiber Communication Conference and Exposition (OFC) (San Diego, California, 2008), OWF1.
3. W. Guo, “Resource Allocation Strategies for Data-Intensive Workflow-Based Applications in Optical Grids”, in Proceedings of 10th IEEE International Conference on Communication systems, (Singapore, 2006), pp. 1 – 5.
4. Amitabha Banerjee, Wu-chun Feng, “Algorithms for Integrated Routing and Scheduling for Aggregating Data from Distributed Resources on a Lambda Grid,” IEEE Transactions on Parallel and Distributed System. 19, pp. 24-34 (2008).
5. Andrew J. Page, Lukas Ahrenberg, and Thomas J. Naughton, “Low memory distributed reconstruction of large digital holograms”, Optics Express, 16(3), pp. 1990-1995(2008)
6. Giuseppe Talli, Paul D.Townsend, “Hybrid DWDM–TDM Long-Reach PON for Next-Generation Optical Access,” J. Lightwave Technol. 24(7), pp. 2827–2834(2006).
7. Y. Shachaf, C. H. Chang, P. Kourtessis, “Multi-PON access network using a coarse AWG for smooth migration from TDM to WDM PON”, Optics Express, 15(12), pp. 7840-7844(2007)
8. Sil-Gu Mun, Sang-Mook Lee, Katsunari Okamoto, and Chang-Hee Lee, “A multiple star WDM-PON using a band splitting WDM filter”, Optics Express, 16(9), pp. 6260-6266(2008)
9. M. S. Rogge, Y.-L. Hsueh, and L. G. Kazovsky, “A novel passive optical network with dynamic wavelength allocation,” in Proc. OFC Tech. Dig.,2004, Paper FG1.
10. N.Genay et al, “Colourless ONU module in TDM-PON and WDM-PON architecture for optical carrier remote modulation,” ECOC2005, VOL. 2, Sep. 2005.
11. M. P. McGarry, M. Maier, and M. Reisslein, “Ethernet PONs: A Survey of Dynamic Bandwidth Allocation (DBA) Algorithms,” IEEE Commun. Mag., vol. 42, no. 8, Aug. 2004, pp. S8–S15.

12. Amitabha Banerjee, Wu-chun Feng, “Algorithms for Integrated Routing and Scheduling for Aggregating Data from Distributed Resources on a Lambda Grid,” IEEE Transactions on Parallel and Distributed System. 19, 24-34 (2008).
13. Radulescu, A., van Gemund, “On the complexity of list scheduling algorithms for distributed-memory systems,” in Proceedings of the 13th international conference on Supercomputing, (New York, NY, USA, 1999), pp. 68–75.