Machine Learning Assisted Optical Network Resource Scheduling in Data Center Networks

Hongxiang Guo, Cen Wang, Yinan Tang, Yong Zhu, Jian Wu, Yong Zuo

Machine Learning Assisted Optical Network Resource Scheduling in Data Center Networks*

Hongxiang Guo, Cen Wang, Yinan Tang, Yong Zhu, Jian Wu, and Yong Zuo

Beijing University of Posts and Telecommunications, Beijing 100876, China
hxguo@bupt.edu.cn

Abstract. Parallel computing allows us to process incredible amounts of data in a timely manner by distributing the workload across multiple nodes and executing computation simultaneously. However, the performance of this parallelism usually suffers from network bottleneck. In optical switching enabled data center networks (DCNs), to satisfy the complex and time-varying bandwidth demands from the parallel computing, it is critical to fully exploit the flexibility of optical networks and meanwhile reasonably schedule the optical resources. Considering that the traffic flows generated by different applications in DCNs usually exhibit different statistical or correlative features, it is promising to schedule the optical resources with the assistance of machine learning. In this paper, we introduce a framework called intelligent optical resources scheduling system, and discuss how this framework can assist resource scheduling based on machine learning approaches. We also present our recent simulation results to verify the performance of the framework.

Keywords: Machine learning · Optical switching · Data center network · Parallel computing.

1 Introduction

In the age of big data, various large-scale parallel computing frameworks (e.g. Hadoop and Spark) have been widely applied in commercial data centers. Due to the diverse network patterns and the huge traffic volume caused by the parallel computing jobs, the traditional electrical data center networks (DCNs) are currently hard to satisfy their networking demands. In contrast, optical switching enabled DCNs possess advantages of high bandwidth, low latency and multi-dimensional (e.g., space/time/wavelength) switching capability and thus have been regarded as a potential solution for the next generation DCNs.

However, it is still a challenge to effectively schedule the optical network resources to meet the requirement of parallel computing. On one hand, because multiple parallel computing jobs may usually coexist in DCNs, it is hard to individually allocate optimal optical resources for each job. On the other hand,

* Supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61331008 and No. 61471054.

the background applications (e.g. virtual machine migration and data backups) will inevitably appear in DCNs and generate large background flows to compete the network resources with the running parallel computing jobs, leading to a significant slowdown of computing process. As a result, it is difficult to directly adapt the optical resources to the network traffic demands of parallel computing jobs.

Recent advances in machine learning (ML) have achieved remarkable performance in dealing with a variety of pattern recognition and classification problems. In DCNs, the statistical property of the traffic generated by different applications (i.e. background applications and parallel computing applications) shows significant difference. For instance, the parallel computing jobs usually generate socalled Coflows [1], which consist of many parallel flows with a common performance goal (e.g., a common completion deadline). Thus, the ML approach seems highly suitable to act as a traffic pattern recognizer and assist the optical resource scheduling process. Indeed, once the runtime network traffic patterns are accurately recognized before the optical resource scheduling, the multi-dimensional switching capability of optical switching enabled DCNs is expected to be exploited more effectively, and accordingly the upper-layer parallel computing can also be accelerated.

In this paper, we make a discussion on ML assisted optical network resource scheduling in DCNs. We introduce a framework named Intelligent Optical Resources Scheduling System (IORSS), which relies on ML techniques to classify the mixed traffic flows, recognize the traffic patterns of concurrently running parallel computing jobs, and then assist the optical switching enabled DCNs to appropriately scheduling its multi-dimensional network resources. We simply review our recent research works and shed light on the feasibility of our IORSS. We also verify the outstanding performance of IORSS through simulations.

2 Intelligent Optical Resource Scheduling System

The overview and the workflow of IORSS are shown in Fig. 1. The aim of IORSS is to recognize the traffic pattern by examining the collected traffic statistics, and then guide the DCN to reasonably allocate the optical resources for parallel computing jobs. To this end, a flow classification module is necessary to firstly distinguish and extract the network traffic generated by parallel computing applications. In other words, the original mixed traffic matrix R is divided into two parts, a background traffic matrix B and a mixed Coflow matrix C. Obviously, $R = B + C$, and the matrix B is relatively sparse as it is not likely that the large volume of background traffic arises throughout the whole DCN. As for the matrix C, it may contain several concurrent Coflows that are needed to be further decomposed, namely $C = \sum_{k=1}^{m} A^{(k)}$, in which m denotes the total number of Coflows and $A^{(k)}$ represents the traffic matrix of k^{th} Coflow. Therefore, a traffic pattern recognition module is utilized to analyze and recognize the traffic pattern of each individual parallel computing application. Finally, according to the recognized traffic pattern, an optical resources allocation model is required

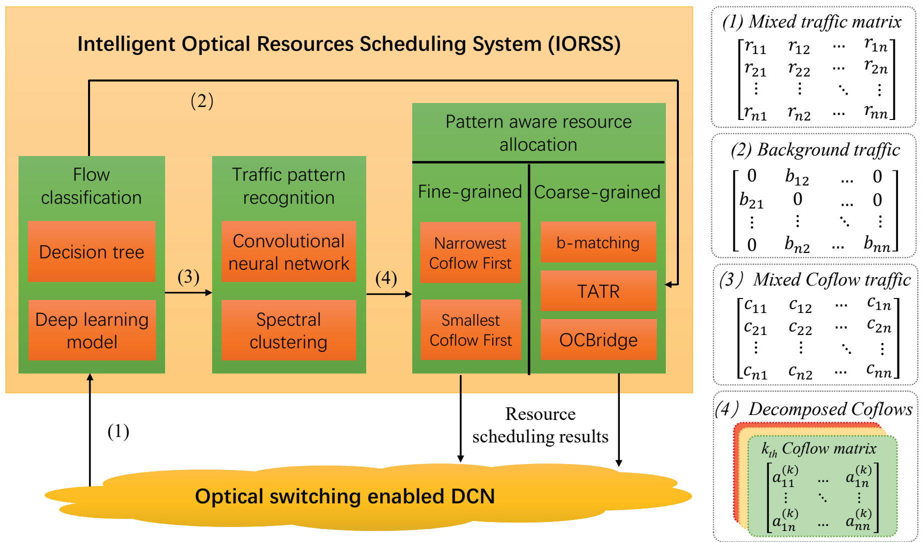

Fig. 1. Intelligent optical resource scheduling system (IORSS) for optical switching enabled data center network. The figure also presents several example algorithms about how to realize each module in IORSS.

to adapt the optical resources to the running parallel computing applications as well as background flows.

2.1 Flow Classification Module

In DCNs, because the network traffic demands of different applications are usually mixed together, the flow classification module needs to identify the type of each flow according to the collected traffic statistics, and then classify it into different groups (i.e., whether or not the flow is generated by parallel computing jobs). The commonly used statistics may include the flow level information such as five-tuple, and the packet level statistics such as time stamp, average packet size and packet size variation.

Currently, machine learning algorithms have shown outstanding performance in solving the classification problem. Similarly, given the sufficient amount of related network statistics, we may anticipate a well-trained machine learning model (e.g., deep learning model) can provide an accurate mapping from the input statistics to the flow type. One example is our previously proposed decision tree model [2], which is capable of identifying the type of flow according to the packet size variation.

2.2 Traffic Pattern Recognition Module

Different from the point-to-point flows produced by background applications, the "all-to-all" network traffic in parallel computing is actually much more complicated. As for the resource scheduling in the case of multiple computing jobs, several heuristic approaches such as Narrowest-Coflow-First [3] and Smallest-Coflow-First [3] have proven their superiority, but the priori knowledge of the traffic pattern (e.g. which flows belong to the same computing job) is necessarily required. However, because the networking demands generated by concurrent computing jobs are also mixed up, the individual traffic pattern of the mixed flows are hard to be recognized and separated, leading to the difficulty in directly allocating optical resources for the mixed traffic of parallel computing jobs. Fortunately, the traffic flows of each job are usually clustered and show a potential correlation, which provides an opportunity to discover the traffic pattern by using ML techniques.

To obtain the traffic pattern, after extracting the statistics of the flows generated by parallel computing jobs, an intuitive way is to directly feed these flow statistics into a fullyconnected deep learning model and find a mapping which outputs the traffic pattern. Nevertheless, because the basic neural network cannot effectively deal with the input with 2D pattern (i.e., matrix), it is not suitable to recognize the traffic pattern of a specific computing job from mixed flows. Besides, the huge amount of flow statistics in DCN will definitely cause many difficulties in training a large scale fully-connected learning model. In our recent research [4], we provided an alternative solution, which applies a convolutional neural network (CNN) to learn the number of parallel computing jobs running concurrently in DCNs, and then uses a spectral clustering method to obtain the set of worker nodes within each parallel computing job. By this way, the traffic pattern can be clearly figured out to guide the optical resource allocation process.

2.3 Optical Resource Allocation Module

Considering the traffic characteristics in data centers, it is beneficial to construct an optical switching enabled DCN by combining the coarse-grained optical circuit switching (OCS) and the fine-grained time-slotted optical switching. For the OCS resource allocation, the main principle is to offload the heavy stable network traffic in DCNs. According to the background flows classified by the flow classification module, the OCS reconfiguration strategies, such as b-matching [5], traffic adaptive topology reconstruction (TATR) [6] and OCBridge [7], can be simply employed. For the fine-grained optical resources, after obtaining the traffic pattern learnt by the traffic pattern recognition module, a simple but effective heuristic algorithm such as Smallest-Coflow-First [3] can be applied as the principle of allocating the time-slotted optical network resources.

3 Performance Evaluation

In order to verify the feasibility and evaluate the performance of IORSS, we conducted a simulation on the OpenScale [8], which is a small-world topology based optical DCN architecture. Moreover, to validate the applicability of IORSS to different optical switching enabled DCN architectures, we further extended our IORSS into a Fat tree DCN architecture with OCS capability.

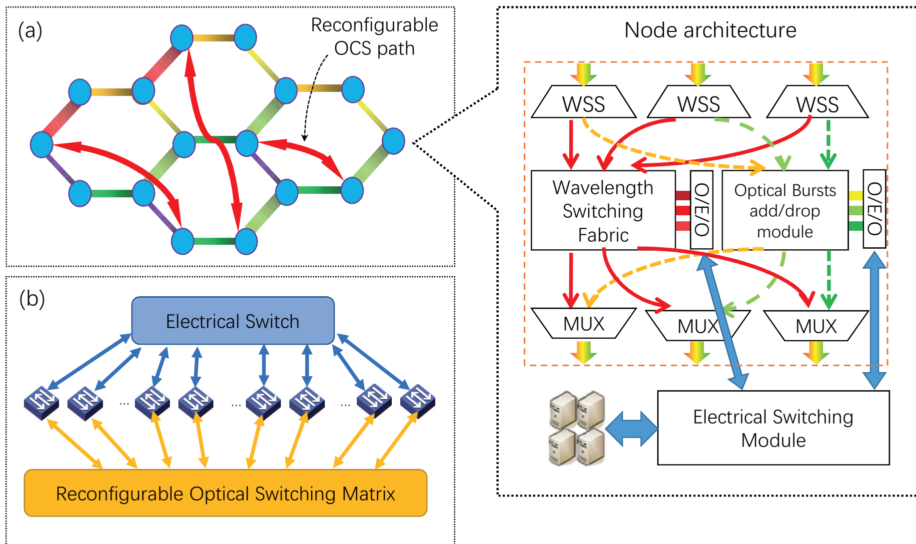

Fig. 2. (a) OpenScale architecture and (b) Fat tree architecture with OCS capability.

3.1 DCN Architectures and IORSS Implementation

The architecture as well as the node design of OpenScale is depicted in Fig. 2(a). In OpenScale, several hexagon rings are organized into a regular lattice structure. The nodes in each ring operate in a time-slotted switching manner to provide logical full-mesh connections among the adjacent nodes. Each node also has wavelength switching capability, so it can build a number of OCS paths to connect remote nodes directly. In summary, OpenScale is a typical optical DCN architecture which relies on both coarse-grained and fine-grained optical resources allocation. In our simulation, the OpenScale architecture consisted of 216 nodes. Each node was capable of building one OCS connection, and the bandwidth of each OCS path was set as 10 Gbps. We simplified the time-slotted switching in each ring as a time-slot based static allocation process. More specifically, each node was statically allocated 1 Gbps to communicate with other nodes within the same ring.

The Fat tree architecture is depicted in Fig. 2(b), where all nodes are connected with a high-radix electrical switch as well as a reconfigurable optical switching matrix. In the simulation, there were also 216 nodes included in the Fat tree architecture. The bandwidth of optical link and electrical link were set to 10 Gbps and 1 Gbps, respectively. Similar with the OpenScale, each node also had the capability to build one OCS connection to communicate with another node.

Regarding the implementation of IORSS, a decision tree based method [2] was deployed as the flow classification module, in order to extract Coflows [1] (i.e. all the correlated flows generated by parallel computing jobs) from the mixed traffic. For the traffic pattern recognition module, we used the method proposed in [4]. Finally, we employed TATR [6] as the coarse-grained OCS resources scheduling algorithm. We deployed Smallest-Coflow-First (SCF) algorithm [3] to guide the fined-grained time-slotted resources, which means the bandwidth of time-slotted optical resources were firstly assigned to the smaller Coflows.

3.2 Simulation Setup and Results Analysis

In our simulation, the raw traffic requests were generated by mixing background flows and the traffic flows of multiple (from 1 to 6) Coflows. The size of background flows followed a uniform distribution from 10^5 Mbytes to 10^7 Mbytes, while the total size of each Coflow followed a uniform distribution from 1 Mbytes to 10^5 Mbytes. We randomly generated 200 groups of raw traffic requests and recorded the average Coflow completion time (CCT) under the ML assisted resource scheduling of IORSS. We compared the CCT result with the baseline scheduling method which randomly built OCS connections among the nodes and applied the round-robin algorithm to schedule the time-slotted optical resources.

As shown in Fig. 3(a), it can be found that when multiple Coflows coexisted in DCNs, our IORSS could provide an effective resource allocation in both OpenScale and Fat tree architectures. With the increasing number of concurrent Coflows, the performance promotion became more significant. According to the results shown in Fig. 3(b), we can find that with the help of SCF, the smaller Coflows had a better CCT promotion. As a result, the average CCT could be effectively reduced, and accordingly the parallel computing jobs were expected to be accelerated.

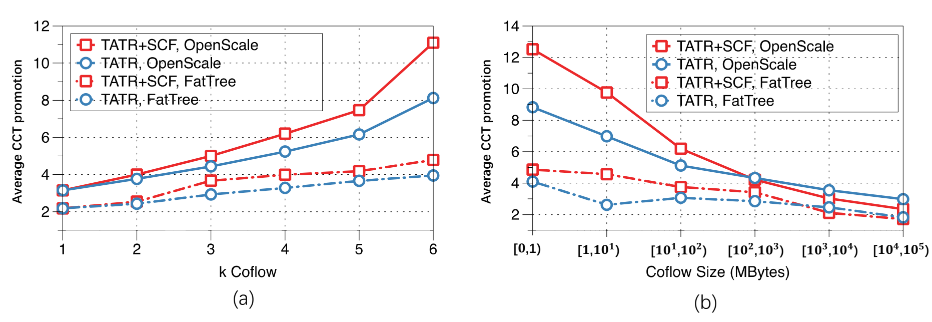

Fig. 3. Simulation results. The promotion ratio is defined by dividing the absolute CCT value under the IORSS by the CCT of baseline scheduling method.

4 Conclusion

Parallel computing has become one of the most popular applications in modern DCs. To fully take the advantage of optical switching enabled DCN architectures, a machine learning assisted intelligent optical resources scheduling system (IORSS) is introduced and discussed in this paper, and its applicability to different DCN architectures has also been successfully verified by simulations. We believe our IORSS can drive the optical switching enabled DCN architectures to more effectively scheduling their optical network resources, and improve the performance of the upper layer parallel computing applications.

References

1. Chowdhury, M., Stoica, I.: Coflow: A networking abstraction for cluster applications. In: Proceedings of the 11th ACM Workshop on Hot Topics in Networks, pp. 31–36. Washington (2012)
2. Wang, C., Cao, H., Yang, S., Guo, J., Guo, H., Wu, J.: Decision Tree Classification based Mix-flows scheduling in Optical Switched DCNs. In: Proceedings on Optical Fiber Communication Conference, paper W1C.4. San Diego (2018)
3. Chowdhury, M., Zhong, Y., Stoica, I.: Efficient coflow scheduling with Varys. In: Proceedings of the 2014 ACM Conference on SIGCOMM, pp. 443–4546. Chicago (2014)
4. Wang, C., Guo, H., Gao, X., Chen. Y., Tang, Y., Wu, J.: Efficient Topology Reconstruction via Machine Learning based Traffic Patterns Recognition in Optically Interconnected Computing System. IEEE Access **7**, 28548–28558 (2016)
5. Yang, T., Zhang, D., Guo, H., Wu, J.: Topology reconstruction strategy with the optical switching based small world data center network. In: 21st OptoElectronics and Communications Conference, paper WA2-108. Niigata (2016)
6. Zhang, D., Yang, T., Guo, H., Wu, J.: Enabling traffic optimized topology reconstruction with the optical switching based small world data center network. In: 41st European Conference on Optical Communication, paper P.6.11. Valencia (2015)
7. Tang, Y., Guo, H., Zhu, Y., Yuan, T., Gao, X., Wang, C., Wu, J.: Effectively Reconfigure the Optical Circuit Switching Layer Topology in Data Center Network by OCBridge. Journal of Lightwave Technology **37**(3), 897–908 (2019)
8. Zhang, D., Guo, H., Yang, T., Wu, J.: Optical switching based small-world data center network. Computer Communications **103**, 153–164 (2017)