Demonstration of Reconfigurable Virtual Data Center Networks Enabled by OPS with QoS Guarantees

W. Miao[(1)], S. Peng[(2)], S. Spadaro[(3)], G. Bernini[(4)], F. Agraz[(3)], A. Ferrer[(1,3)], J. Perello[(3)], G. Zervas[(2)], R. Nejabati[(2)], N. Ciulli[(4)], D. Simeonidou[(2)], H.J.S. Dorren[(1)] and N. Calabretta[(1)]

[(1)] COBRA Research Institute, Eindhoven University of Technology, w.miao@tue.nl [(2)] University of Bristol [(3)] Universitat Politècnica de Catalunya [(4)] Nextworks

Abstract *We demonstrate a reconfigurable virtual datacenter network by utilizing statistical multiplexing offered by scalable and flow-controlled optical switching system. Results show QoS guarantees by the priority assignment and load balancing for applications in virtual networks.*

Introduction

Data centers (DCs) are facing the rapid development of ICT markets, providing a broad range of emerging services and applications. The next-generation DCs are required to provide more powerful IT capabilities, i.e. more bandwidth, storage, shorter time to market for new services to be deployed[1]. One of the key requirements of the future DC is the multi-tenancy. DC network (DCN) virtualization is the key enabler for supporting multi-tenancy[2]. By taking advantage of virtualization, multiple coexisting virtual DC networks (VN) will be created allowing for the efficient sharing of the heterogeneous DCN resources (ToR switches, DCN switches, wavelengths, etc.) to support diverse services and applications running on top of VNs. As the demand of users or applications has been changing, the established VNs need to be reconfigurable and adaptive to the dynamic applications requirements.

Current electronic switches in DCN support statistical multiplexing and could allow for an efficient sharing of the DCN resources to implement a large number of VNs with QoS guarantee. However, there are hardware and control issues. Today's multi-tier tree-like DCN architecture built up on multiple switches, each with limited ports and speeds, suffers of an intrinsic scalability issues in terms of bandwidth and latency[3]. Moreover, the proprietary operating system of those switches prevents multi-vendors equipment interacting, and therefore the creation of VNs.

Optical switching technologies based on space, time, and wavelength multiplexing could implement fast reconfiguration and large port count switches[3]. A flat DCN architecture based on optical packet switch (OPS) with sub-microsecond end-to-end latency and scalable port count has been recently demonstrated[4]. Although the OPS is able to support statistical multiplexing of sub-microsecond traffic flows, the slower operation of conventional control plane frameworks (in the order of tens of milliseconds) prevented the implementation of DCN based on OPS technology with QoS guarantee based on statistical multiplexing.

This paper presents and demonstrates a reconfigurable virtual optical DCN architecture that decouples the OPS sub-microsecond flow switching from the control plane operation to achieve QoS guarantee and statistical multiplexing at the sub-microsecond flow level.

System operation

The proposed reconfigurable virtual DCN architecture is shown in Fig. 1(a). It is composed by a reconfigurable flat DCN and a unified control plane. The flat DCN is based on scalable and modular reconfigurable OPS architecture with optical flow control[4]. Optical flows generated and transmitted by the ToR include an optical label that, according to the OPS look-up table (LUT), determines the forwarding output port at the OPS. An optical flow control is implemented between the ToRs and the OPS.

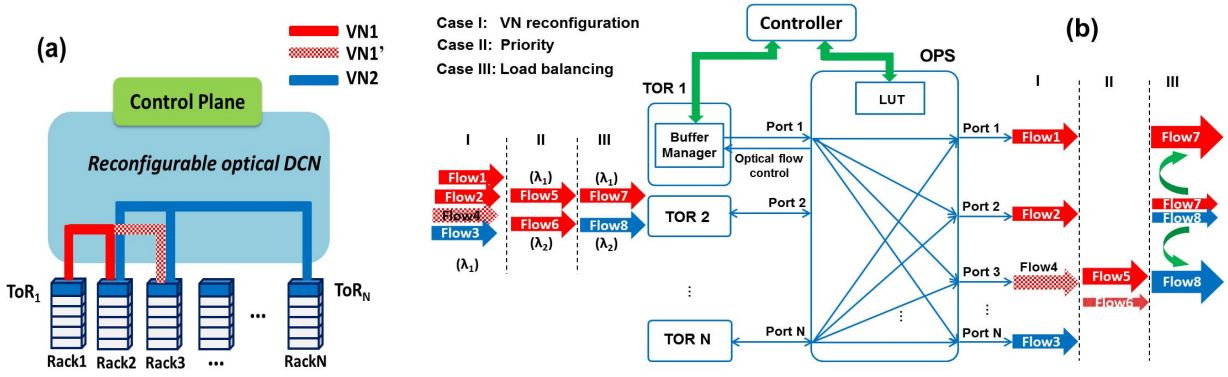

Fig. 1: (a) Architecture for reconfigurable optical DCN; (b) System validation set-up

The OPS provides positive ACK signals to the TOR for successful delivered packets. In case of contention (due to statistical multiplexing and resources sharing), packets has to be retransmitted by the ToR. Different priority policies can be included in the label field to implement different classes of QoS. Sub-microsecond hardware implementation of the optical flow control (flow generation, LUT check, QoS check, ACK generation at the OPS, and (re-)transmission) has been measured[4].

The control plane consists of a centralized controller deployed on top of the flat DCN, and is responsible to provision the VNs by configuring the LUTs of both OPS and ToRs switches. A VN configured by the controller is composed by a set of entries in the LUTs that match multiple traffic flows among a well-defined set of ToRs. As an example, Fig. 1(a) shows that VN1 connects ToR1 with ToR2 while VN2 connects ToR2, ToR3, ToRN, with ToR2 belonging to both VNs. Note that once the VNs are provisioned, the application flows are exchanged between the ToRs of the VNs at sub-microsecond level, thus decoupling the slow control plane (milliseconds time scale) and the fast data plane (sub-microsecond time scale).

VNs reconfiguration is possible by updating the LUTs of the ToRs and the OPS. As an example in Fig. 1(a), if the DC operator wants to include ToR3 in the VN1, the control plane is able to flexibly reconfigure the network topology by updating the LUT content. Moreover, exploiting the statistical multiplexing introduced by the OPS, the bandwidth/wavelength resources sharing is possible either within a single VN or among multiple VNs. This enables the dynamic creation and reconfiguration of multiple VNs with optimization of DCN resources utilization, which can be provided by dedicated VNs computation algorithms implemented by the control plane. Moreover, the control plane, hosting a global view of the network status, can monitor the ToRs buffers and initiate in advance a load balancing procedure to diverge the traffic towards less used resources for providing higher QoS capabilities.

Experimental validation

Figure 1(b) illustrates the experimental setup with three selected cases to validate the reconfiguration, statistical multiplexing with QoS guarantee, and load balancing operation of the proposed virtual optical DCN architecture, respectively. The aggregated traffic flows which include sequences of packets from a source to certain destinations are statistically multiplexed at ToR and then transmitted to OPS node. The buffer manager inside the ToR handles the flow destination and generates the labels which will be carried by the flows including the forwarding information and the class of priority. Buffer manager stores the label information and implement the (re-)transmission according to the ACK sent by the OPS node. The gate used for controlling the transmission of packetized 40Gb/s NRZ-OOK payloads (460ns duration and 40ns guard time) is triggered by the buffer manager to emulate the (re-)transmission. The controller is hosted in a PC interfaced with OPS and ToRs switches through a USB link, following an SDN approach.

Case I demonstrates VN reconfiguration and resource sharing exploiting statistical multiplexing. Assuming that the centralized controller has provisioned the VN1 comprising ToR1 and ToR2, and the VN2 comprising ToR2, ToR3, and ToRN (Fig. 1(a)). Flow1 and Flow2, belonging to VN1, and Flow3 belonging to VN2 are statistically multiplexed on the same wavelength to different destinations. As resource competition may exist between flows from same ToR, the controller has been given the authority to assign different priority levels for each flow. A reconfiguration of VN1 is now required to include in VN1 also connectivity with ToR3 (to support Flow4 forwarding). To do this, the controller updates the LUTs of the ToRs and the OPS, accordingly. Fig. 2(a) shows the LUT

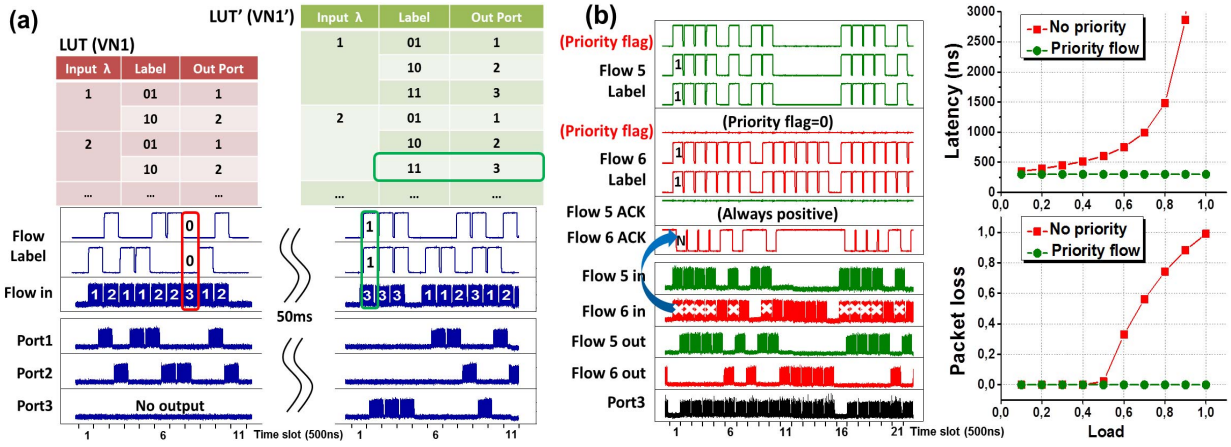

Fig. 2: (a) VN1 reconfiguration; (b) Time traces of Flow5 & Flow6 and packet loss & latency for different input load

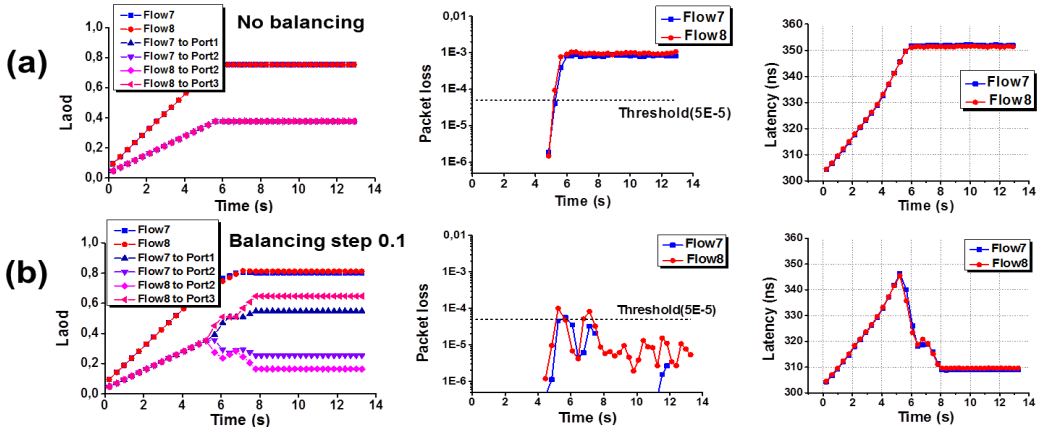

Fig. 3: Load, packet loss and latency changes without adjusting (a) and with load balancing with step of 0.1 (b)

update for the original VN1 and the reconfigured VN1' (LUT'). The traces in Fig. 2(a) show that before the reconfiguration, flows with ToR3 as the destination were dropped since no matched label is found at the OPS. On the contrary, after VN1 reconfiguration and the LUT update, the flows towards ToR3 can be properly delivered. It takes about 50ms for the update procedure (control communications between controller, ToR, and OPS); after that, flows are statistically multiplexed and switched at sub-microseconds time scale. It is worth to note that the other flows destined to ToR1 and ToR2 perform hitless switching during the VN reconfiguration time.

Case II demonstrates statistical multiplexing flows operation under different classes of QoS. Flow5 and Flow6 are heading to the same output port (Port3). Flow5 has been assigned higher priority. One label bit has been used as priority flag. In case of contention at OPS, Flow5 will be forwarded to the output Port3 to avoid packet loss and large latency caused by retransmission. Figure 2(b) shows the label and flow traces, the packet loss, and the latency performance for the two contented flows. Note that the ACK signals for Flow5 are always positive (always transmitted), while the negative ACK signals for Flow 6 indicate retransmission. The packet loss curves confirm no packet loss for Flow5, while the buffer employed at the ToR prevents packet loss up to load < 0.4 for Flow6 and then it increases linearly with the load. Similar behavior is observed for the latency.

Case III demonstrates load balancing operation for the flows belonging to different VNs. Assuming that Flow7 in VN1 and Flow8 in VN2 have common output Port2 among two potential destinations. The contention at Port2 would cause high packet loss for both flows. The contention probability for the output ports indicating the destination usage could be collected from optical flow control signal. Upon reception of the real-time status of the packet loss per flow and the occupancy of each alternative port from the ToR, the controller can enable the balance of the load to output port with less usage. Performance improvement by such load balancing strategy between two flows from different VNs has been reported in Fig. 3. In the experiment, the load of Flow7 and Flow8 has been increased from 0 to 0.8, with 50% probability that a contention occurs at Port 2 at the beginning. Firstly, Fig. 3(a) shows that a packet loss larger than 1E-3 for both flows is measured, in case the controller does not take any action. Once the load balancing is triggered by the controller, a target packet loss threshold of 5E-5 has been set. Above this threshold, the dynamic adjustment of controller would balance the load at Port2 to Port1 (for Flow7) and Port3 (for Flow8) with a step of 0.1. It can be observed that the packet loss is kept less than the threshold and the latency goes down which guarantee the QoS meeting the requirement.

Conclusion

A reconfigurable virtual optical DCN based on OPS has been presented and experimentally validated. Flexible VN reconfiguration by a centralized controller is decoupled from the sub-microsecond hardware switching time scale. Evaluation for the selected cases shows that multiple VNs share the limited resources with guaranteed QoS, by leveraging statistical multiplexing, flow priority assignment, and load balancing.

Acknowledgements

The authors would like to thank the FP7 LIGHTNESS project (n° 318606) for supporting this work.

References

[1] S. Sakr et al., "A survey on large scale data management approaches in cloud environments," IEEE Com. Sur. & Tut., Vol. **3**, no. 13, p. 311 (2011).

[2] M. Faizul Bari et al., "Data Center Network Virtualization: A Survey," IEEE Com. Sur. & Tut., Vol.**15**, p.909 (2013).

[3] C. Kachris et al., Optical Interconnects for Future Data Center Networks (Springer, 2013), Chap. 1.

[4] W. Miao et al., "Novel flat datacenter network architecture based on scalable and flow-controller optical switch system," Optics Express, Vol. **22**, no. 3, p. 2465 (2014).