Reconfigurable Virtual Data Centre Networks Based on OpenFlow-enabled OPS

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, Bristol, UK
[3] Universitat Politècnica de Catalunya, Barcelona, Spain
[4] Nextworks, Pisa, Italy

We demonstrate a reconfigurable virtual data centre network by utilizing scalable and flow-controlled optical packet switching system. Flexible virtual network reconfiguration conducted by the centralized SDN controller is decoupled from the sub-microsecond hardware switching time scale. OpenFlow messages enabling control communication between SDN controller and OPS switch are validated. Results show QoS could be guaranteed by priority assignment and load balancing for applications in virtual networks.

Introduction

Data centres (DCs) are facing the rapid development of ICT markets, providing a broad range of emerging services and applications[1]. One of the key requirements of the next generation DC is the multi-tenancy which makes efficient utilization of existed resources. As the key enabler for supporting multi-tenancy[2], DC network (DCN) virtualization supports diverse services and applications running on top of multiple virtual networks (VNs) sharing of the heterogeneous DCN resources (switches, ports, wavelengths, etc.). In addition, the established VNs need to be reconfigurable to adapt to the dynamic applications requirements.

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

Optical switching technologies based on space, time, and wavelength statistical multiplexing could implement fast and large port-count switches[4]. 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[5]. Although the OPS is able to support statistical multiplexing of sub-microsecond traffic flows, conventional control plane frameworks prevented the on-demand resource previsioning and guarantee of QoS due to the limited intelligence and awareness of network topology.

In this paper, we present and demonstrate a reconfigurable virtual optical DCN architecture that decouples the OPS sub-microsecond flow switching from the software defined network (SDN) based control plane. Through the OpenFlow-based interface, the VN could be reconfigured by SDN controller and the QoS could be guaranteed for statistical multiplexing with proper priority assignment.

System operation

Figure 1(a) shows the proposed reconfigurable virtual DCN architecture. 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[5]. The optical flow control is implemented between the ToRs and the OPS. Optical flows generated and transmitted by the ToR includes an optical label that, according to the OPS look-up table (LUT), determines the forwarding output port at the OPS. The OPS provides positive ACK signals 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.

The control plane is composed by a centralized SDN 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. OpenFlow has been implemented to facilitate the communication between the SDN controller and the switch controllers[6]. A VN 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 (tens of 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. Moreover, the control plane, that has also the vision of the network devices could track the status of the ToRs buffers and initiate in advance a load balancing procedure to diverge the traffic towards less used resources for providing higher QoS capabilities.

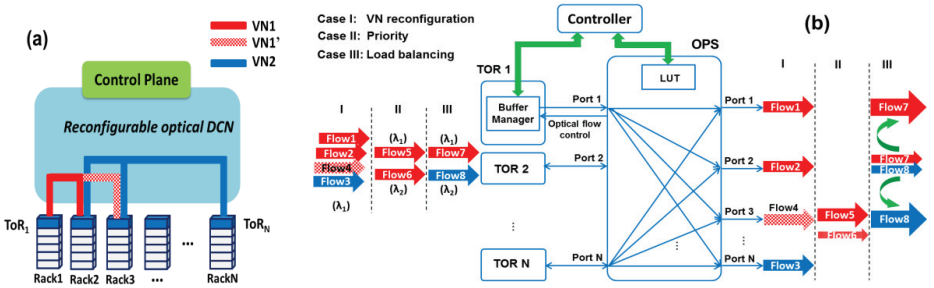

Fig. 1: (a) Architecture for reconfigurable optical DCN; (b) System validation set-up

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

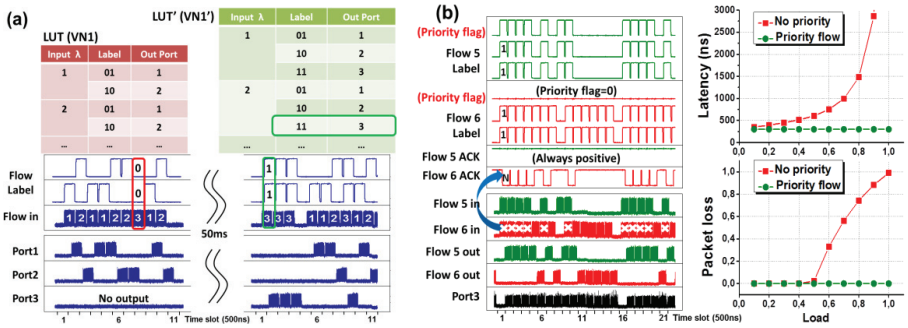

Fig. 2: (a) VN1 reconfiguration; (b) Time traces of Flow5 & Flow6 and packet loss & latency for different input load

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 from 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.

Case I has demonstrated the VN reconfiguration and resource sharing exploiting statistical multiplexing. 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). The controller updates the LUTs of the ToRs and the OPS, accordingly. Fig. 2(a) shows the LUT update for the original VN1 and the reconfigured VN1' (LUT'). The traces in Fig. 2(a) shows flows with ToR3 as the destination were dropped since no matched label will be found at the OPS. On the contrary, after VN1 reconfiguration and the LUT update, the flows towards ToR3 can be properly delivered. It has been measured that it takes about 50ms for the update procedure and 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.

In Case II we have demonstrated 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. Fig. 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 and latency 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.

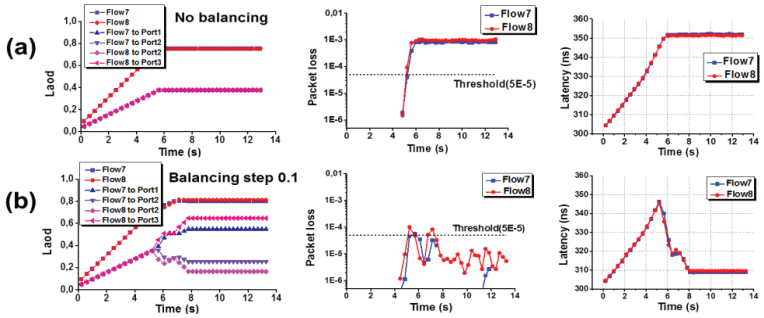

Fig. 3: Load, packet loss and latency without adjusting (a) and with load balancing with step of 0.1 (b)

In Case III we have demonstrated load balancing operation for the flows belonging to different VNs. Flow7 in VN1 and Flow8 in VN2 have the common output Port2 among two potential destinations. 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. 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. 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. The dynamic adjustment of controller would balance the load at Port2 to Port1 (for Flow7) and Port3 (for Flow8). 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 architecture based on scalable and flow-controlled OPS has been presented and experimentally validated. Flexible VN reconfiguration operated 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 by leveraging statistical multiplexing, and the QoS can be effectively guaranteed with flow priority assignment and load balancing.

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. 3, 311-336 (2011).

[2] M. Meeker and L.Wu, “2013 internet trends,” Kleiner Perkins Caufield & Byers, Technical Report (2013).

[3] P. Gill et al., "Understanding network failures in data centers: measurement, analysis, and implications," ACM SIGCOMM 2011, pp. 350-361.

[4] C. Kachris, K. Bergman and I. Tomkos, Optical Interconnects for Furture Data Center Networks (Springer, 2013), Chap. 1.

[5] W. Miao et al., “Novel flat datacenter network architecture based on scalable and flow-controlled optical switch system,” Optics Express 22, 2465-2472 (2014).

[6] F. Agraz et al., “Experimental assessment of an SDN-based control of OPS switching nodes for intra-data center interconnect,” ECOC 2014, paper We. 2. 6.5.