Experimental Demonstration of One-to-Many Virtual Machine Migration by Reliable Optical Multicast

Payman Samadi, Junjie Xu, Keren Bergman

Lightwave Research Laboratories, Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA, ps2805@columbia.edu

Abstract We present one-to-many Virtual Machine migration leveraging optical multicast. A control plane provides connectivity between the application and data plane layers. We evaluated the prototype on a datacenter testbed and report improvement on the migration time compared to conventional methods.

Introduction

Traffic pattern in cloud-based data centers has significantly shifted from north-south to east-west\(^1\). Many services actively communicate among racks and pods to perform services such as load-balancing, maintenance or fault recovery. Virtual Machine (VM) migration is a major application used by many services in cloud-based data centers. Large number of VMs in range of gigabytes are migrated regularly and contribute to the east-west load of the data center interconnection network. Conventional migrations are point-to-point only however, VMs share gigabytes of identical pages such as the running OS and the applications.

Optics is widely used in data centers as point-to-point links, however the switching substrate is still electronics. Optical switching has higher energy efficiency compared to the electronic counterpart. Moreover, leveraging optical switching in data centers can further improve the energy consumption by avoiding unnecessary optical/electrical/optical conversions at the switches.

In\(^2\) authors introduced optical multicast on an Optical Circuit Switching (OCS) substrate using passive optical splitters. Compared to Internet Protocol (IP) multicast\(^3\), it does not require complex configurations on all the switches/routers and has superior energy efficiency. Also the bit rate transparency makes it future proof.

In this work, we present one-to-many VM migration using physical layer optical multicast. We implemented a reliable multicast protocol for guaranteed data delivery, necessary for VM migration. Our proposed architecture leverages an OCS network for data transmission and an Electronic Packet Switched (EPS) network for the receiver's acknowledgments. The software architecture consist of a 3-layered control plane to manage migration requests from the application layer and configurations on the data plane.

We demonstrate live migration of VMs from one server to multiple receivers on our data center testbed. Unlike conventional methods on the electronic network, the reliable optical multicast transmits VMs simultaneously to all receivers, independent to the multicast group size and has 1.1 – 8 order lower transmission time. The system can be used in cloud data centers for the east-west network.

Architecture and Testbed

Fig. 1 demonstrates the hardware architecture of the reliable optical multicast to perform one-to-many VM migration. It is a hybrid architecture\(^4\) that Top-of-Rack (ToR) switches are simultaneously aggregated by both OCS and EPS networks. Flow rules on the ToRs are set in a way that the data transmission from the sender to the receivers are on the OCS and the acknowledgment packets are sent on the EPS network. This architecture is to address unidirectional nature of the optical links used in data centers. Acknowledgment packets are minimal and does not add considerable load to the electronic network.

A reliable optical multicast works as following: data from the sender is forwarded to the OCS network. The switching substrate forwards the
data to the input port of the splitter and the output ports are connected to the receivers. Passive optical components physically split the data and multicast to the receivers. To support reliable data transmission, receivers are also connected to the sender through the EPS network by configuring flow rules on the ToRs.

The reliable multicast protocol is implemented using the open-source libraries of NACK-Oriented Reliable Multicast (NORM)\(^5\). NORM uses a selective, negative acknowledgement (NACK) mechanism for transport reliability. We have implemented both single thread and multithread transmission to support higher bit-rates. The latter achieves higher throughput but the implementation does not support live migration yet.

The control plane architecture is shown in Fig. 2(a). The northbound API is implemented by Redis\(^6\) that is a fast pub/sub messaging system and the Southbound API is Floodlight. In this 3-layered architecture, requests from the application layer are sent to the control plane layer. Each request consist of the sender and receivers addresses and the name of the VMs. Requests are scheduled based on the optical circuit and splitter availability using an Integer Programming model solved by a heuristic method\(^2\). For the scheduled requests, ToRs and OSS are configured using southbound APIs and transmission is started. Pages that become dirty during the migration are re-transmitted at the end. At each receiver, once the transmission is finished, the VM is started.

Fig. 2(b) demonstrates our data center testbed. The OCS network is implemented by Calient S320 that is an Optical Space Switch (OSS) and the EPS network by Juniper EX4500, L2/L3 Ethernet switch. For ToRs, a Pica8 P-3920 OpenFlow switch is divided into 8 bridges that operate as 8 separate ToR switches. For each rack, there is a server equipped with a dual-port Network Interface Controller (NIC), an Intel Xeon E5-2430 6-core processor and 24 GB of RAM. The control plane software is run on a separate controller server. Electrical to optical conversions are performed at the ToRs using SFP+ transceivers.

**Evaluations**

In the first set of experiments, we evaluate the performance of the reliable multicast protocol on the OCS network of the testbed. Single thread implementation enables live VM migration with lower throughput. Multithread implementation provides higher throughput but since it splits the VM image to multiple chunks, any update during migration is not transmitted to the migrated machine.

Fig. 3(a) demonstrates the average throughput for 1:4 receivers transmission. In both single and multithread implementations, the throughput...
Conclusions and Future Work
Live one-to-many VM migration using reliable optical multicast was presented in this paper. We built a hybrid data center testbed and developed a software system and a protocol. Experimental results demonstrate lower transmission time independent to the number of receivers compared to conventional migration methods on electronic networks. Our future work includes improving the throughput of the reliable multicast protocol and enabling one-to-many VM migration over inter-data center networks.

Acknowledgments
This research was supported in part by CIAN NSF ERC under grant EEC-0812072 and NSF CNS-1423105. We would also like to thank Juniper Networks and Calient Technologies their generous donations to our testbed.

References