Institut National des Sciences Appliquées de Lyon

Routing in Black Box

Ritvik Kapila

Supervisor: Prof. Frédéric Le Mouël

A report submitted in fulfillment of the requirements of

IST-4-NET2 Networks *in the*

Information Science and Technology Department

Dec 20, 2019

Modularized Load Balancing for Multipath Data Center Networks

By Fujie Fan, Bing Hu and Kwan L. Yeung

1 Introduction

The purpose of this paper is to propose a load balancing algorithm for large scale, multi-tier fat tree based data centre networks. Several load balancing algorithms which are widely used face many challenges such as efficiency and scalability, due to the ever increasing data centre network size. In this paper, performance of different flowlet-based load balancing algorithms and MLAB (Modularized Load Aware Balancing) are compared based on better delay performance under all traffic patterns. It also concludes that the new proposed algorithm, the Modularized Load Balancing, can achieve perfect load balancing in multi-path networks and gives a mathematical proof that it is more efficient than any existing algorithm in performance.

2 Context

Load Balancing is the process of distributing network traffic across multiple servers thus ensuring that no single server bears too much demand, in-turn providing maximum bisection bandwidth. Efficient Load balancing strategies are necessary as more and more commercial applications such as web services, search engine and network disc are deployed on the cloud. If handled improperly, load balancing can introduce packet re-ordering and degrade the throughput. To avoid packet reordering, flowlet switching is used, which means that the packets belonging to the same burst will be forwarded by the same path; different bursts can be routed by different paths. The paper also proves theoretically that MLAB has perfect load balancing property, and that with a little cost of deployment, it can significantly improve the performance of modern data center networks.

3 Contribution

Two major contributions are made in this paper:

3.1 Fast Congestion Feedback Mechanism

Fast Congestion Feedback Mechanism of MLAB is a completely distributed and data-driven scheme for load notification without any need of control packets or central controllers. The paper implements the load balancing algorithm in an in-network and distributed way. The paper focuses more on 2-tier networks as an example to explain Fast Congestion Feedback Mechanism. Feedback is used to transmit the global load information to the sources. To implement an in-network algorithm, the feedback process only happens among switches. The main objective of any data centre is to achieve high performance switching. Also, it is made sure that the protocol stacks on end hosts are not affected. Such an in-network algorithm is implemented by using encapsulation technologies such as VXLAN: Virtual extensible local area network, and NVGRE: Network virtualization using generic routing encapsulation – a packet is encapsulated after it is received by a switch and decapsulated before it is sent to a server.

3.2 Design of Modularized Architecture for high-tier Networks

In this section, the author explains a new scalable Modularized Architecture for high-tier networks, that is designed and implemented, where the routing domains are isolated. The new load balancing algorithm is much simplified for deployment and upgrade. The aggregation tier and the core tier is modularized into several routing domains. A domain consists of all aggregation switches and core switches that are on the same path between two access switches. From the perspective of access switches, a domain is a black box that can switch their data to another access switch. The black box is more like a single switch than a network. Modularized Architecture comes in really handy while upgrading networks as well as finding and replacing faulty ones.

4 Discussion

The paper explains everything in a lot of detail along with examples and mathematical proofs. After giving examples and proving the hypothesis in theory, the author also evaluates the performance of the same using experimental results in each section.

Summarizing, MLAB has three advantages to improve the performance of data centers: First, it improves the flexibility and scalability of data centers. With the modularized architecture, it is very convenient to upgrade a low-tier network to a higher-tier one. Second, MLAB has natural desynchronizing characteristic due to the encapsulated organization, which can avoid potential performance degradation in many cases. The new algorithm can effectively desynchronize the traffic from different sources and has far less communication overhead than many other algorithms. Third, the new load balancing algorithm is compatible with existing applications, and is transparent to end hosts. Each server can treat the upper network as a single large-scale switch so that the protocol stacks are not affected.