Deconstructing Data Center Load Balancing with Network Traffic Profiles

Saim Salman
Brown University

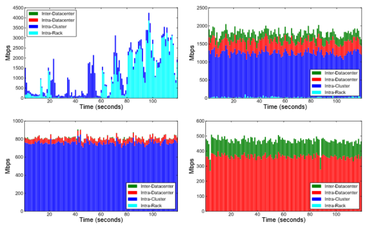
Antonio Marsico
British Telecommunications

Gianni Antichi
Queen Mary University of London

Theophilus Benson
Brown University

GOAL

Dynamically configurable load balancer.

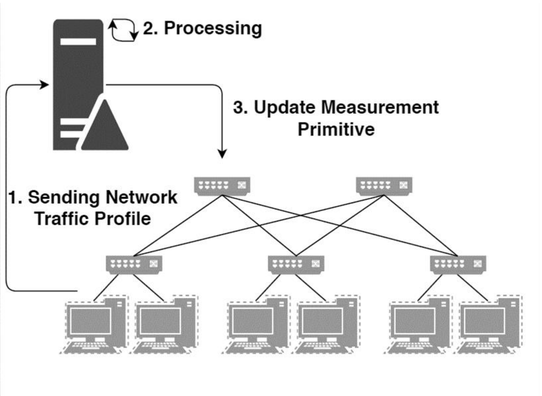


Link Utilization
Queue Occupancy
Heavy Hitter

Can adapt to traffic conditions

Use the correct measurement metric

VISION



- Programmable and extensive load balancer framework
- Each data center host regularly sends its ***network traffic profile to the centralized controller***
- Reconfigures dynamically based on traffic patterns, communication patterns and the topology.

MOTIVATION

Although all data center load balancing techniques try to **lower flow completion time (FCT)** and keep the **network links as utilized as possible** they leverage **different design choices**. This is because **data-centers are plagued with uncertainties**:

Traffic Patterns

Traffic can be highly dynamic in production level data centers.

Communication Patterns

There is high variability in the communication patterns in Data Centers.

Topology

Expansion + failures can cause asymmetries in the network topology.

Scheme	Granularity	Measurement Primitive	Designed For
ECMP	Flow	***None***	Generic topologies
Let It Flow	Flowlet	***None***	Asymmetric Topologies
Hedera / MicroTE	Flow	***Heavy Hitter***	Optimizing FCT for Elephant Flows
DRILL	Packet	***Queue Occupancy***	Best Performance when network load higher than 80%
Conga	Flowlet	***Link Utilization***	2-Tier Topologies
Hula	Flowlet	***Link Utilization***	Scalable Topologies

Takeaway: **No one measurement primitive is solely superior to the other.**

EXPERIMENTS

		Intra-Pod	Inter-Pod	
			Neighbors	Network-Wide
Web Search	Hula			
	Hula-OQ		✔	✔
	Hula-HH	✔		
Data Mining	Hula		✔	✔
	Hula-OQ			
	Hula-HH	✔		

Conducted experiments using the Hula NS2 codebase and varied the *measurement primitive*, *traffic* and *communication pattern*. The table summaries our findings from out experiments.

Scheme

HULA: An in-network load balancer that uses *link utilization* as measurement primitive
HULA-OQ: Used *queue length as measurement primitive*
HULA-HH: Used *heavy hitter as measurement primitive*

Traffic Patterns

Web Search: Characterized by a prevalence of short flows
Data Mining: More long flows than Web Search.

Communication Patterns

Intra-Pod: Most of the traffic resides in the same pod.
Inter-Pod: Majority of the traffic would be directed towards a **neighbor** or have to traverse the entire data center (**network-wide**)

Takeaway: **The choice of measurement primitive is predicated on the traffic patterns, communication patterns and network topology.**