Facebook Analytics - Structured Abstract

Context:

To improve Facebook's approach to load balancing on shorter timelines of a few seconds, even though it already has effective traffic engineering on longer durations of minutes to hours.

Objective:

To identify the timeframe between packets of which type out of three classifications we made in the technical report contribute more to the traffic intensity in both clusters.

Method:

Categorize the data – intra cluster and datacenter, inter cluster and intra datacenter, inter cluster and datacenter. Plot the line graph with timestamp as x-axis and packet length as y-axis with the classification we made to categorize the data. Need to perform this step for both database and Hadoop cluster separately. Identify the packets which causes traffic in shorter timeframes from each classification in both clusters.

Results:

The packets which contribute more to the traffic was identified and these packets are located with the plots along with the timeframe. Since the traffic will be dynamic and will not be repetitive, we are not able to predict efficiently each time when there are some heavy hitters in load balancing but with the segment of data, we can identify the heavy hitters and be efficiently balanced.

Novelty:

Many publications and articles about the Facebook data center stated that the Facebook setup with Auto scale technique with custom load balancer using modified round robin algorithm while this project suggests using Stackelberg Game theory model for its effective resource utilization and low error rate.

Database cluster - Packet length over the period of time

Hadoop cluster - Packet length over the period of time