Volley: Automated Data Placement for Geo-Distributed Cloud Services

Tuesday, May 4, 12pm EST
CIC, Room 2101

Abstract

As cloud services grow to span more and more globally distributed datacenters, there is an increasingly urgent need for automated mechanisms to place application data across these datacenters. This placement must deal with business constraints such as WAN bandwidth costs and datacenter capacity limits, while also minimizing user-perceived latency. The task of placement is further complicated by the issues of shared data, data inter-dependencies, application changes and user mobility. We document these challenges by analyzing month-long traces from Microsoft’s Live Messenger and Live Mesh, two large-scale commercial cloud services.

Volley is a system we built to address these challenges. Cloud services make use of Volley by submitting logs of datacenter requests. Volley analyzes the logs using an iterative optimization algorithm based on data access patterns and client locations, and outputs migration recommendations back to the cloud service. To scale to the data volumes of cloud service logs, Volley is designed to work in SCOPE, a scalable MapReduce-style platform; this allows Volley to perform over 400 machine-hours worth of computation in less than a day. We evaluate Volley on the month-long Live Mesh trace, and we find that, compared to a state-of-the-art heuristic that places data closest to the primary IP address that accesses it, Volley simultaneously reduces datacenter capacity skew by over 2X, reduces inter-datacenter traffic by over 1.8X and reduces 75th percentile user-latency by over 30%.

Bio

Sharad Agarwal is a researcher in the Networking Research Group at Microsoft Research. His research has recently focused on geographically scaling datacenter services and he works closely with some of Microsoft’s online properties. Prior to the last 5 years at MSR, he spent 2 years at Sprint Research Labs working on routing and traffic management systems for IP backbones. Sharad has a PhD, MS and BS from UC Berkeley. More details are at:

http://research.microsoft.com/~sagarwal

Copyright (c) 2010 Carnegie Mellon CyLab

For more information on the CyLab Seminar Series, please visit www.cylab.cmu.edu.
For more information on the CyLab Seminar Series, please visit www.cylab.cmu.edu.