Pingmesh: A Large-Scale System for Data Center Network Latency Measurement and Analysis – Public Review

Sujata Banerjee
HP Labs
Palo Alto, CA, USA
sujata.banerjee@hp.com

Fast troubleshooting of network problems presents considerable challenges for the network operations team in production systems. In many cases, it is not obvious that a system issue such as slow response from a server or an unreachable server, is caused by a network problem or not in the first place. Today most of this detection and diagnosis is semi-manual and time consuming, and is particularly burdensome in large scale systems. Often the problem is detected only after the customer has already felt the impact of a failure and has submitted a trouble ticket. When a system level problem is reported, the network operations team must first identify if it is indeed a network problem and if so, locate the problem to correct the situation.

This paper presents a solution to the above problems in large datacenters with thousands of servers and describes their deployment experience. The system is called Pingmesh and the basic approach is to use inter-server network latency measurements as an indicator of network problems, or the lack thereof. To limit the overheads of all-pair server measurements, the Pingmesh system measures inter-server latencies at three levels (Top-of-Rack switch, intra-datacenter and inter-datacenter) to create complete latency graphs at each level. Every server in the datacenter participates in these measurements. The latency data is collected continuously by Pingmesh agents at each server under the direction of a Pingmesh controller that creates the list of servers to *ping*. The goal is to provide as close to complete coverage as possible, while limiting measurement overheads. The system design of both the Pingmesh controller and agents have been developed with the goals of scalability, low overhead and always-on capability. The analysis of the data generated by the Pingmesh system provides information about increased latency over historical baselines and packet drop rates.

What makes this an interesting experience paper is that it demonstrates how to use a simple technique to diagnose complicated datacenter network problems in real world deployments. The overall approach and architecture is straightforward, though it is clear that a considerable engineering effort has gone into the design of the individual system components to have the desired properties. The system has been running in several distributed production datacenters for four years and generates 24 Terabytes and over 200 Billion probes per day. It is being used to detect network SLA violations, black holes, silent packet drops, and in conjunction with other methods, can localize the source of the network problem. The authors present a few anecdotal scenarios in which Pingmesh aided in fault detection and mitigation, which makes for an interesting read.

The Pingmesh system is not without its limitations. As expected, simply measuring inter-server latencies is not sufficient to detect and localize the source of all network problems. For example, in order to identify a faulty spine switch, the basic approach needs to be augmented with additional measurements such as traceroutes. The Pingmesh system continues to be enhanced with new features as new network issues come to light in the deployed datacenters and the original scope is expanded. In spite of the limitations, and some doubts about the applicability of the approach to other deployments, the program committee agreed that learning about the experience of this Pingmesh team would provide value to others in the field.