CSE 726 : Data Intensive Distributed Computing Seminar

Review

Black-Box Problem Diagnosis in Parallel File System

was presented by : Rishi Baldawa
review by : Hiraksh Bhagat

In this paper, the authors has developed an algorithm for automatically diagnosing different performance problems in parallel file systems by comparing different metrics gathered at every node. It uses Black Box performance metrics for peer comparison to basically do two things (i) to find whether any fault exists in the system and (ii) analyze the metrics to pinpoint faulty resource. The main goals of the author are application transparency minimal false alarms, minimal instrumentation overhead and many specific problem coverage. The paper says very clearly of what it is not looking to achieve here like code-level debugging, pathological workloads and diagnosis of non-peers. The paper demonstrates authors' approach for realistic storage problems injected into different file system bench marks in PVFS and Lustre clusters.

The paper aptly describes why it uses Black Box metrics in peer comparison. It makes various of assumptions like all peer servers have identical software configuration, are synchronized and have a homogenous environment. The problem involving storage and network resources are separated into two classes viz. hog faults and busy or loss faults. Considering a small file system, the paper makes a variety of observations assuming many things which is not entirely true. Based on these observations, the authors developed the diagnosis algorithm. It works in two phases. The first phase finds the faulty server by using PDF on various OS-level metrics. It gives two approaches for this viz. Histogram based approach and Time based approach. Threshold selection is implemented on training data using machine learning algorithms. Phase 2 observes peer divergence in storage and network resources by calculating throughput and latency.

Advantages:

- It gives an algorithm to find non fail stop severs which is difficult to find under normal circumstances.
- Low Overhead
- Low data requirement
- SLOs avoided

Disadvantages:

- The paper has made observations and assumptions on a relatively smaller parallel file system.
- It has a restricted approach in testing in homogenous environment.
- It assumes their approach can be easily scalable to real world large HPC clusters.
- It only finds performance degrading servers.

During the discussion of the paper, various questions were raised on the observations made by the author and the assumptions made to get them. Questions about why a faulty server has an effect on the entire parallel file systems' throughput were tough to understand. The only answer that was drawn out of the discussion was synchronization issues. It also brought out a point where the paper contradicts itself on machine learning algorithms because it uses ML algorithms in threshold selection to judge a faulty server. In the end it was evident that the paper required a lot of future work for the algorithm to be implemented on large scale HPC clusters. The discussion had to end here due to time constraints.