Christopher Poenaru 20409287
Anthony Clark 20434297

ECE 454 Design Document – Assignment 1

Sync vs Async

We designed this implementation using synchronous protocols and using HsHa servers for both the FE and BE nodes. We decided to use synchronous requests due to the elegant simplicity of handling network connections. The HsHa model is used to provide multiple worker threads where computation needs to be done, however only requires one networking thread.

Load Balancing

To perform load balancing, we implement a very simple probabilistic solution. To do this, we choose the backend server to forward a request to based on a weighted probability, where the weight is determined by the number of cored of the backend server. The number of cores that a backend server has is broadcast by the backend when it joins the cluster. If a server has 6 cores, and the total cores from all backend servers in the cluster is 40, then the probability of sending the request to that particular backend server is 6/40. While this method doesn't provide guarantees about balancing the load fairly, it offers a fair amount of confidence without significant management overhead to poll for resources and usage.

Crash & Failure Detection

To handle the detection of crashes and failures we make use of the heartbeat mechanism from our BE nodes. Every 100ms a BE node should send a heartbeat to each FE seed node, and those will further propagate this information to each of the FE nodes. If a seed node does not receive a heartbeat from a backend server for over one second then it will remove the node from it's list of BE nodes and propagate the new list to the other FE servers.

BE Node Join

When a BE Node joins the cluster, it will broadcast to each of the seed nodes that it exists, giving its hostname, number of cores, service and management ports. The FE non-seed nodes run an RPC request to the seed nodes every 500ms to obtain an updated list of the BE servers in the cluster. This means that FE nodes should have up to date lists of the BE nodes, including newly joined BE nodes within 500ms (plus networking latency and overhead).