End-to-End Roundtrip Latency vs # Invocations
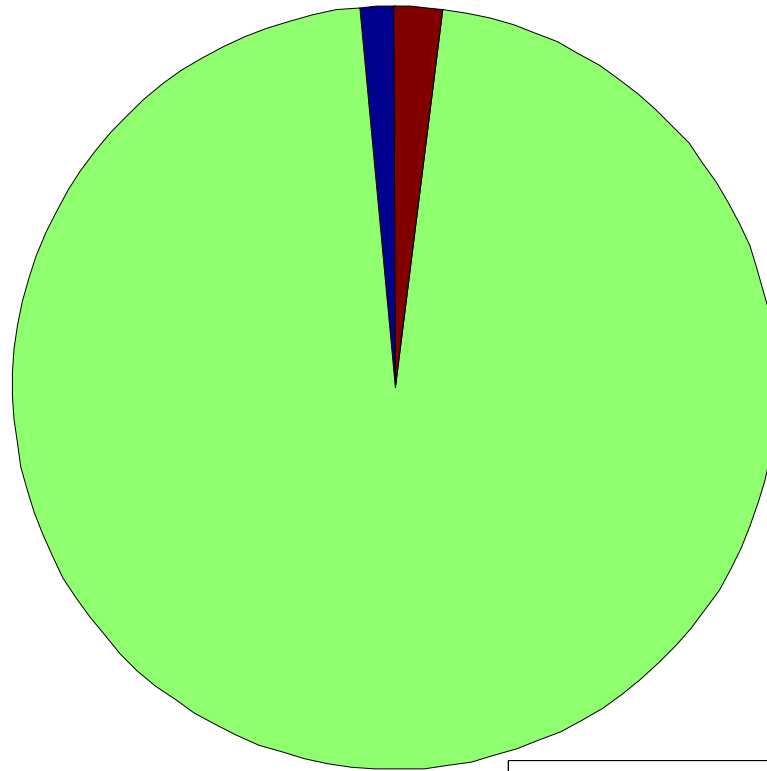
(Note: y-axis is in microseconds)
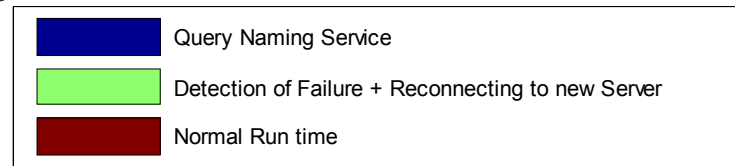Maximum jitter: ~700ms
Minimum jitter: ~300ms

Failover time vs # Failures Injected
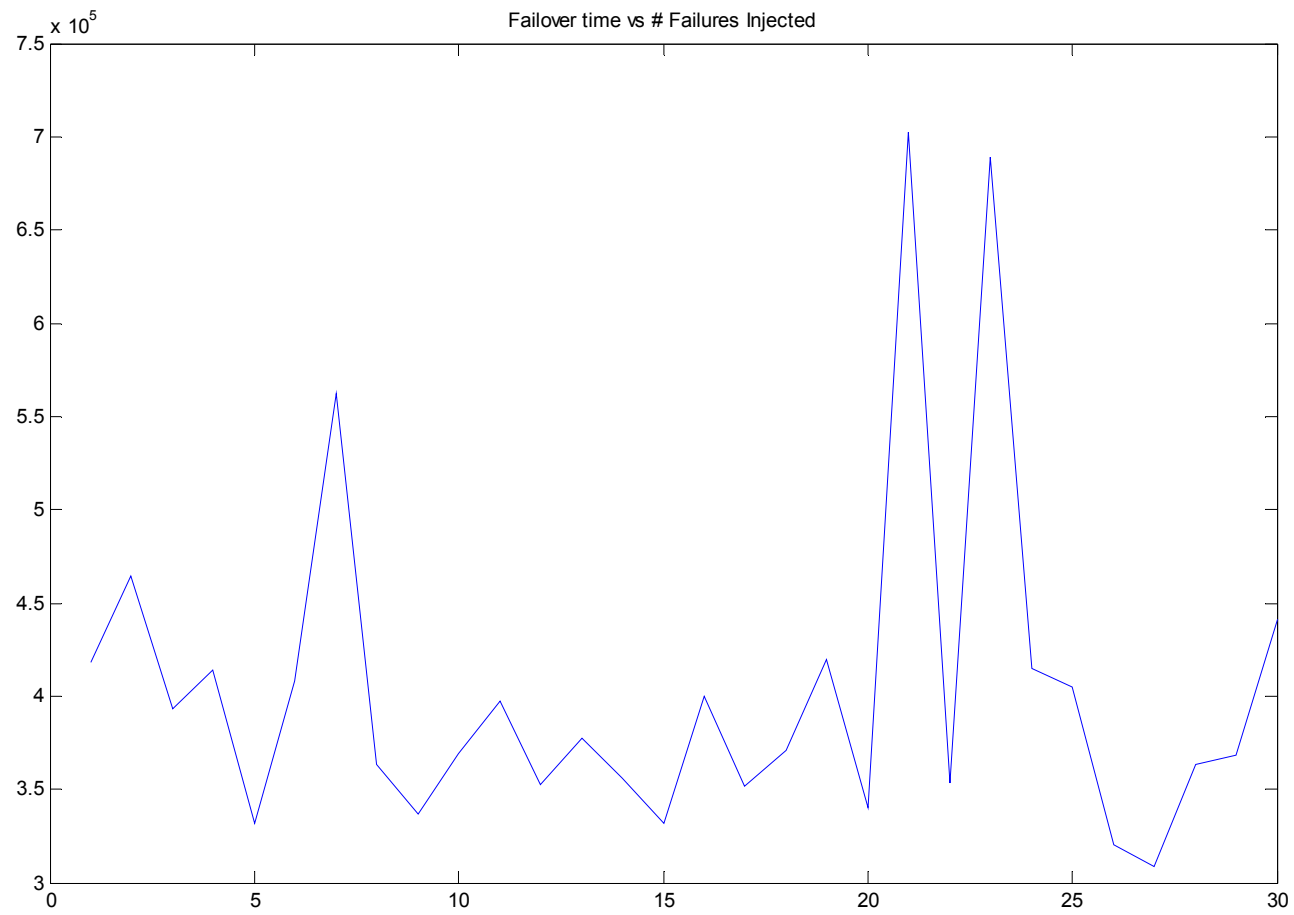
(Note: y-axis is in microseconds)
Average failover time: ~ 404ms

Failure Induced
1%   2%

97%

Legend:
- Query Naming Service
- Detection of Failure + Reconnecting to new Server
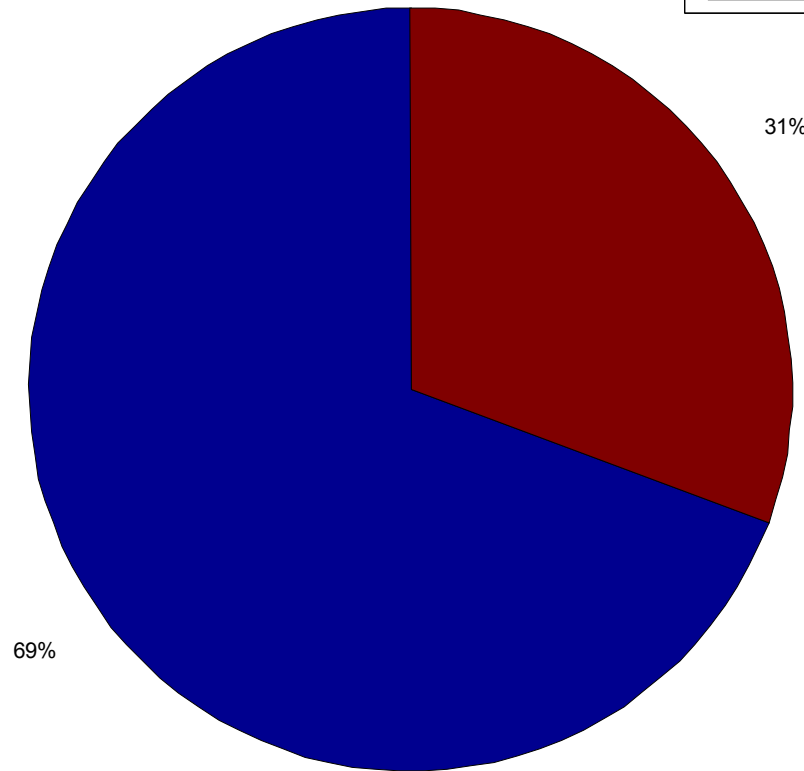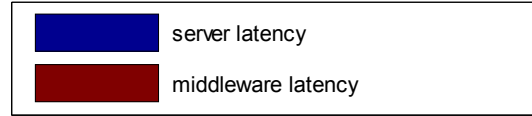- Normal Run time

FaultFree

server latency
middleware latency

31%

69%

Strategies to reduce spikes

1. To keep the naming server query time low, the client will keep a list of available servers. When the client cannot connect to any of them, it'll ask the replication manager for an updated list.
2. Before the client reissues its invocation, it creates another instance of HostBean on the new server that it failed over to. The failover time consists of 1) connecting to a new HostBean, and 2) reissuing invocation and waiting for reply. To reduce the time for 1), we'll pre-create HostBean's on servers that are on the "list of available servers."
3. To reduce the time for 2) (reissuing invocation and waiting for reply), we can implement caching on the server side.
4. We can implement active replication for faster failover.