HW5 Report, Chordy: A Distributed Hash Table

David Fischer

October 4, 2025

1 Introduction

This assignment's goal was the implementation of *Chordy* a DHT based on a variation of *Chord*, a distributed lookup protocol designed to efficiently identify the node in a peer-to-peer that stores a particular data item. *Chord* addresses many problems distributed systems face such as load balancing, decentralization, scalability, availability, and flexible naming.

2 Implementation

2.1 Ring Structure

To ease validation and development, instead of the usually intended consistent hashing function, simple numerical IDs were used for both nodes and later on keys. This still provides the random distribution of keys to nodes, which should result in even storage sizes between nodes. As opposed to the *finger table*, used in the original *Chord* paper, which allows for $O(logN)$ lookup times, a simpler routing table was implemented. Each node only keeps track of its successor and predecessor in the ring. This results in $O(N)$ lookup times. The ring structure is periodically stabilized through a `stabilize/3` function, which gradually heals the ring when new nodes join by having each node ask its successor for their predecessor and adjusting its own records accordingly. As the structure will have to wrap around at some point, the `between/3` function was implemented in a way to handle `From` being larger than `To` by checking if the `Id` is either larger than `From` or smaller than `To`.

2.2 Storage

Storage was implemented by having each node keep a map of key-value pairs which can be added and looked up by sending corresponding messages to the nodes. These messages are routed through the ring structure until they reach the node responsible for the given key, at which point the operation is executed and the client is notified through a message including the reference

it provided using `make_ref/0`. Additionally, to support new nodes joining the ring, a handover message containing the keys split from the predecessor, based on the two nodes keys, is sent to the new node.

2.3 Failure Detection

To detect failed nodes, Erlangs failure detection mechanism was used to have each node monitor their predecessor and successor. Nodes keep a record of their successors successor, which is provided during each stabilization round. This is done to maintain structure should their original successor fail. Should the predecessor fail, nodes set it to `nil` and wait for the predecessors predecessor to notify them. This approach is however not ideal if two neighboring nodes fail concurrently.

2.4 Data Replication

As data is only stored in memory, node failures result in data loss. To mitigate this, nodes keep a replica of their predecessors data which is built concurrently with the main storage. The client is only notified of a successful add operation once both the main storage and the replica have been updated. The handover process now also includes a split version of the replica data sent to new nodes joining the ring. When a predecessor dies, the node merges the replica storage with its main storage.

3 Main problems and solutions

3.1 Functions and Parameters

TODO: amount of functions and parameters often lead to confusion and failing tests due to missmatches and bugs

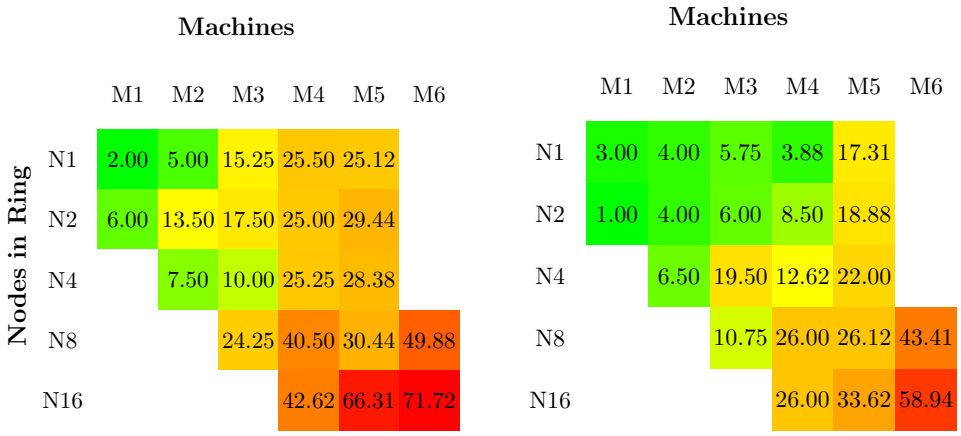
3.2 Replication

TODO: difficulty in implementing replication correctly, especially during handover and failure scenarios

4 Evaluation

4.1 Performance

xxx something on scale affecting the performance in both directions. A major bottleneck in these tests also stems from the initial ring member being the single point of entry for all lookup and add operations.



(a) Add Operations (b) Lookup Operations

Figure 1: Performance comparison in milliseconds for 1000 operations per machine

4.2 Ring Maintenance

When a ring is stabilized on a short delay, the convergence is sped up and thus lookup and add operations get to the correct node faster. Key responsibility and handovers are also sped up. This however causes the system to use more resources both in terms of compute and network usage, specifically nodes generating messages for *request*, *notify*, and *status* each round.

5 Conclusions

The advantages of this distributed storage come both in performance, through load balancing, and fault tolerance, through self healing and the lack of a single point of failure. Fault tolerance should be however prioritized, especially in productive applications, as data loss is often not acceptable.

Future improvements to the system include routing table improvements with a solution as the *finger table* to improve lookup timings and considering network latency, although this would significantly increase complexity. A better replica mechanism which is either periodically synchronized through the stabilize function, or persisted on disk which nodes automatically restarting and recovering their state. Replicas could also be used to improve read performance, although this requires dealing with consistency issues. Finally, either an eventual consistency model could be applied to improve availability, or two-phase commits to ensure consistency, both excluding the other due to the CAP theorem.