Peer-to-Peer Wikipedia

Rain Gu and Nick Bradley

October 9, 2016

1 Introduction and Motivation

Wikimedia's mission is to bring free educational content to the world.(1) One way they do this is through Wikipedia an online collaborative encyclopedia hosted by Wikimedia. Having a single organization manage the content has several implications for unbiased content. For eaxmple, if there are negative comments about a company that company could pay to have it removed which they do this is by supporting the infrastructure required to host online content such as Wikipedia. This is one of the largest costs and has opportunities for content to become biased. In the client-server model used by Wikipedia, full control of the documents lies with the server managers.

We are going to investigate the feasiblity of a peer-to-peer Wikipedia-style document sharing system.

Wikipedia spends $$$ on hosting costs per year.

2 Background and Assumptions

Our primary focus is on making

Peer-to-peer systems are $^{.Theyallow}{}_{Ournetworkhas}{}_{nodesthathaveaMTBFof}{}_{(anodeisconsideredfailedifitdoesnotacceptmessagesfromt}$

the content of a new document will be available to any client once it is committed by the leader.

the content of a document will not be available to any client if it is deleted on any connected node.

3 Proposed Approach

We start under the strongest guaratee that any document must be available if any node is online. This requires every document to be stored on every node. We can implement our application protocol on top of Raft.

We may then investigate how to relax the availablity guarantee to allow for scaling to a larger network. May be able to give a probabilistic bound

based on uptime: since uptime is independent, could have that the average uptime of the nodes that a document is on must be greater than some threshold.

For example, if a document is on 10 nodes with uptimes of about 1 hour, then it would be equivalent to a document being on 5 nodes with uptimes of 2 hours.

Question: how do you pick whch nodes the document goes on?

Under this approach, need a search method to find the document.

4 Evaluation Methodology

Simulate a variety of network setups to evaluate performance

Make sure or guarantees hold on all edge cases 1) will the document propgate if a leader election happens?

5 Timeline

6 References

1) Wikimedia https://www.wikimedia.org/