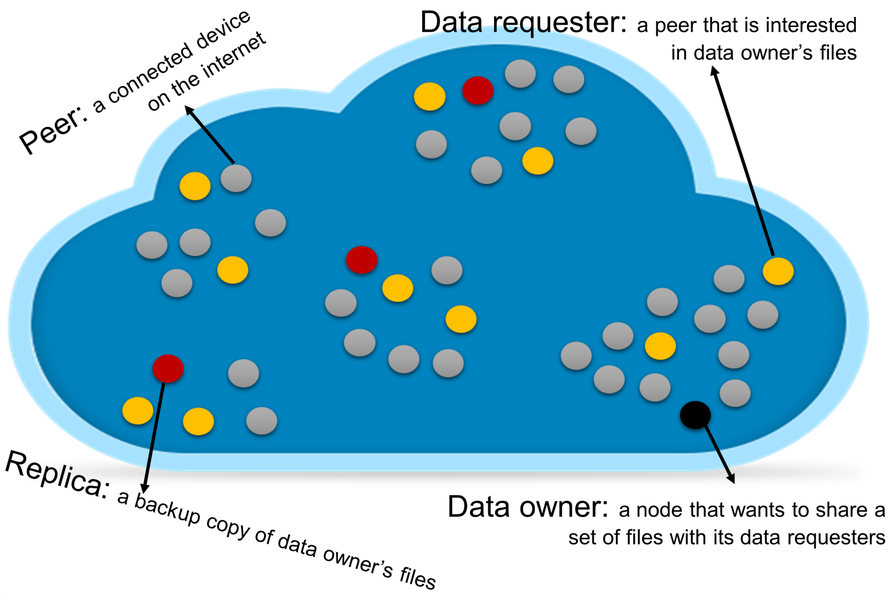
Awake: decentralized and availability aware replication for P2P cloud storage

Yahya Hassanzadeh, Alptekin Küpçü and Öznur Özkasap
Department of Computer Engineering, Koç University, Turkey
{yhassanzadeh13, akupcu, oozkasap}@ku.edu.tr

Abstract:

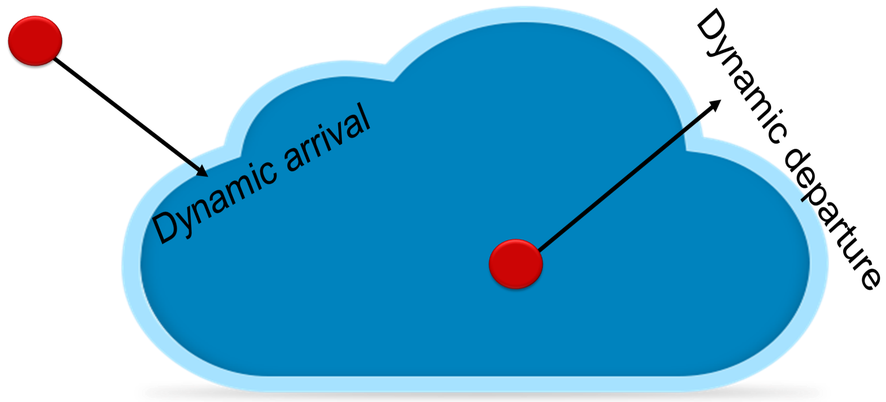
The data availability of a peer-to-peer (P2P) cloud storage is negatively affected by the dynamic arrivals/departures of the nodes to/from the system, which is called churn. We propose a dynamic, fully decentralized, and availability aware replication algorithm named Awake, which maximizes the availability of replicas in the structured P2P cloud storage. Our extensive analysis results show that Awake improves the availability of replicas with an average gain of about 21% compared to the best existing solutions, and it is independent of the churn behavior of the underlying system.

P2P Cloud Storage:

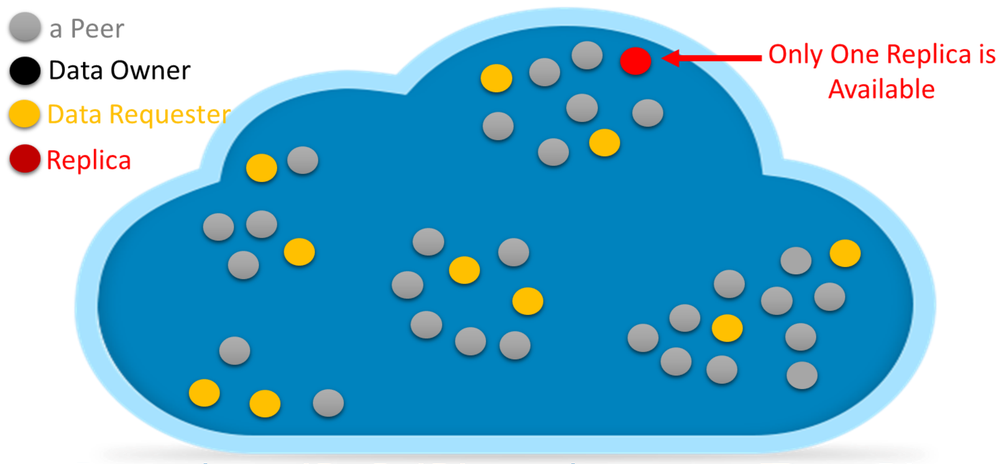


Problem:
Data unavailability because of churn

Churn: Dynamic arrivals and departures of nodes to/from the P2P cloud storage

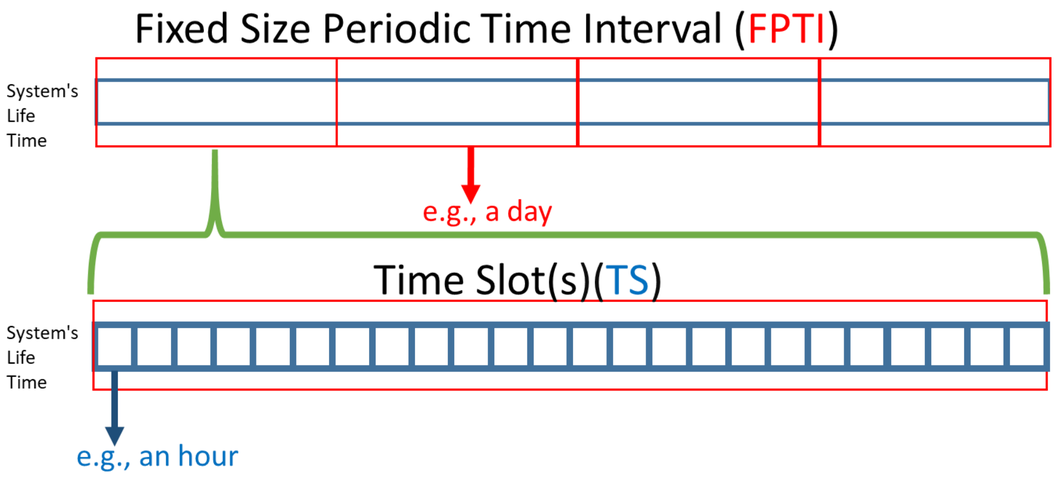


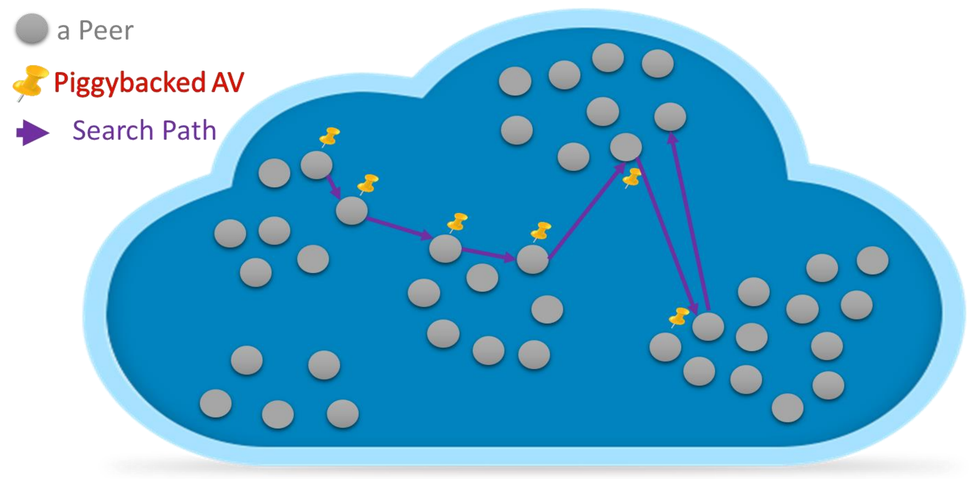
Poor data availability or data unavailability because of churn



Awake: Availability Aware Replication

Awake is employed in a structured P2P system where the availability information of nodes is represented by an availability vector. Each cell of the availability vector corresponds to the availability probability of the node during a certain time slot of a fixed time interval. Nodes piggyback their availability vectors on the query messages they initiate or route. On receiving a message for routing, the node updates its availability view of other nodes using the received availability vectors.





$$\max \sum_{t=1}^{S} \sum_{i=1}^{N} Y_i \times AT_j[i][t] \quad \text{s.t.}$$

$$\sum_{i=1}^{N} Y_i = R$$

$$\forall i, 1 \leq i \leq N \quad Y_i \in \{0, 1\}$$

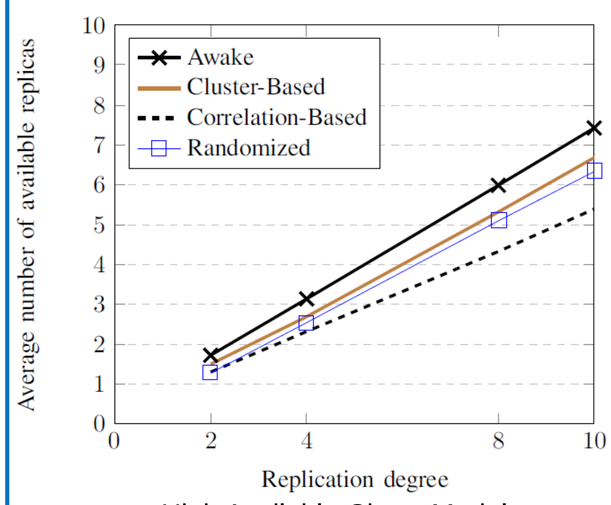
Awake models the availability aware replication as an integer linear programming problem, solves it on the input set and outputs the replica set to the data owner. The data owner then shares the obtained replica set with its set of data requesters.

Related Works:

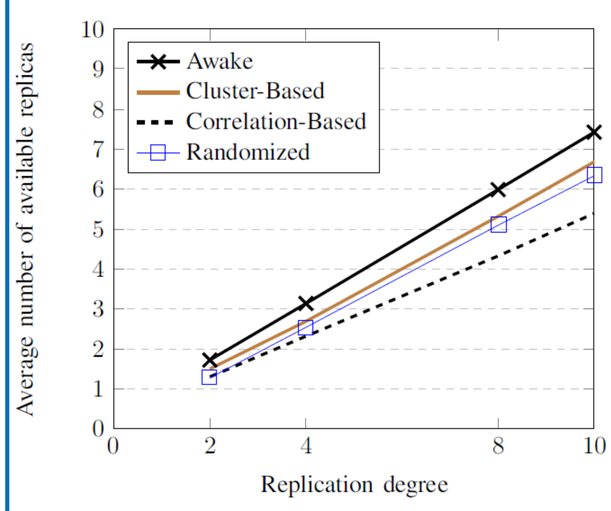
Algorithm	Strategy	Availability Awareness	Targeted Churn Model
Randomized	Randomized assignment of replicas	No	High Available
Cluster-based	Clusters nodes based on common features like time zone, replicates in each cluster.	No	High and Moderate Available
Correlation-based	Replicates on pairs of anti-correlated nodes	No	Low Available
Awake	LP model of availability aware replication	Yes	Independent

A comparison between decentralized availability-based replication algorithms

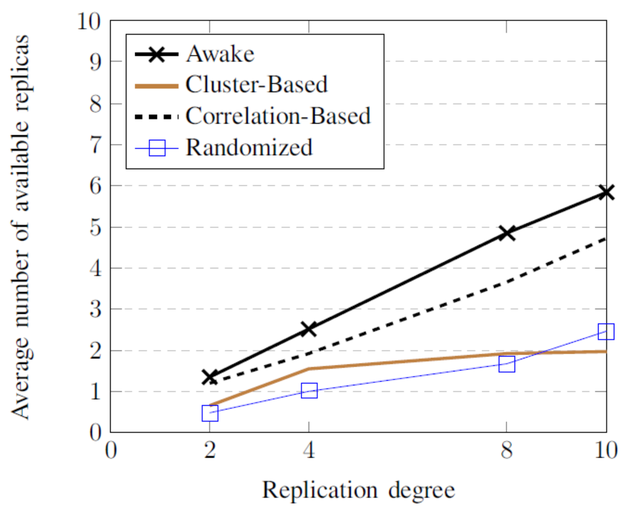
Performance Results:



High Available Churn Model



Moderate Available Churn Model



Low Available Churn Model

Awake outperforms the existing solutions under the high, moderate, and low available churn models with the gain of about 24%, 14%, and 26%, respectively. The simulation results indicate that for a certain number of replicas and a fixed churn model, Awake performs independently of the system size.

Reference

Hassanzadeh-Nazarabadi, Yahya, Alptekin Küpçü, and Öznur Özkasap. "Awake: decentralized and availability aware replication for P2P cloud storage." Smart Cloud (SmartCloud), IEEE International Conference on. IEEE, 2016.

Our open source simulator, SkipSim is available on GitLab:
git@gitlab.com:yhassanzadeh13/skipsim-distribution-bundle.git