

Towards Predictive Replica Placement for Distributed Data Stores in Fog Environments

Tobias Pfandzelter, David Bermbach
Mobile Cloud Computing Research Group
TU Berlin & ECDF, Germany
{tp, db}@mcc.tu-berlin.de

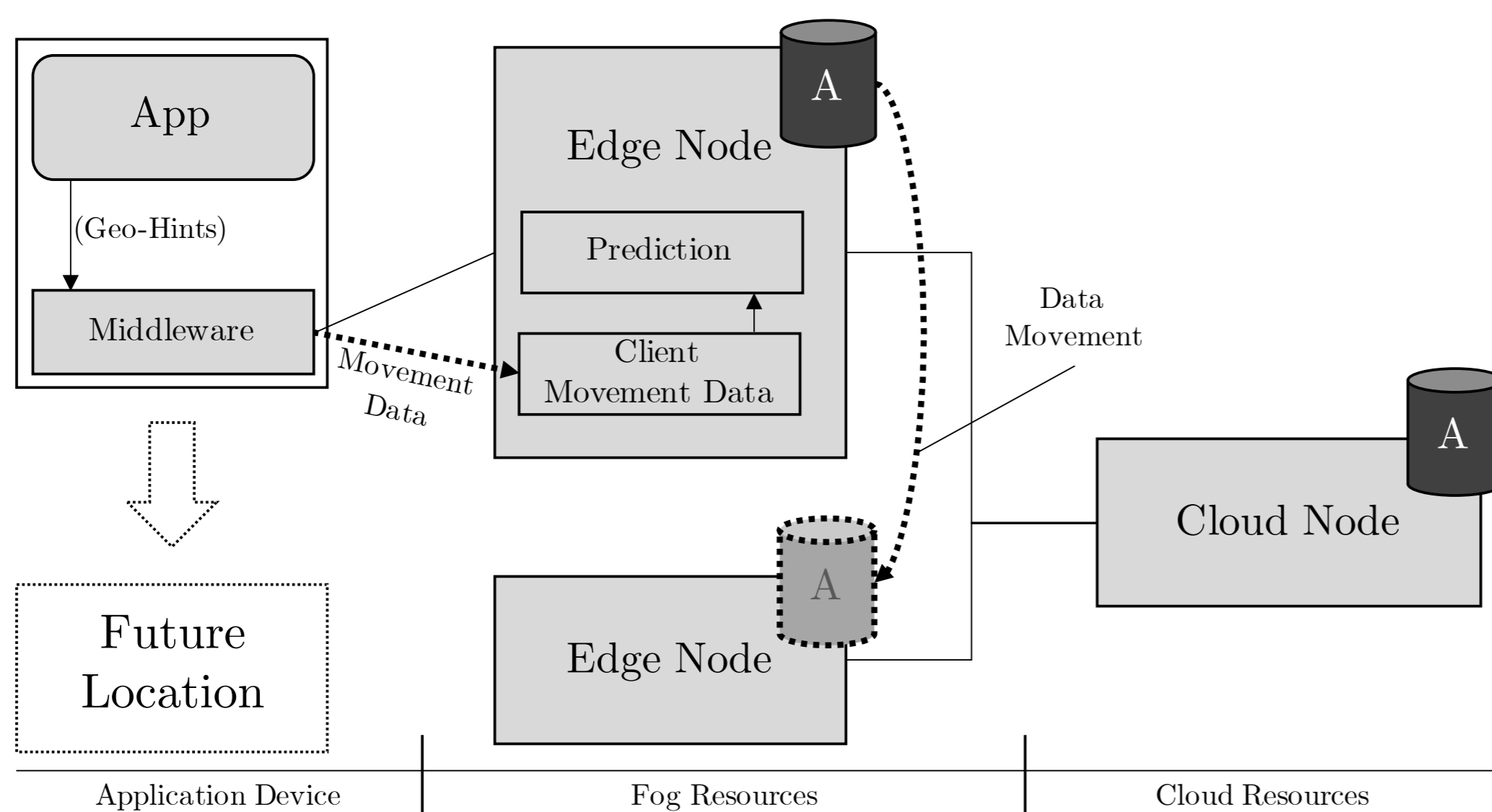
CHALLENGE

Fog data stores replicate datasets near clients for improved access performance. How can the data store decide **where** to replicate that data before it is accessed, i.e., predict future data access from mobile clients?

APPROACH

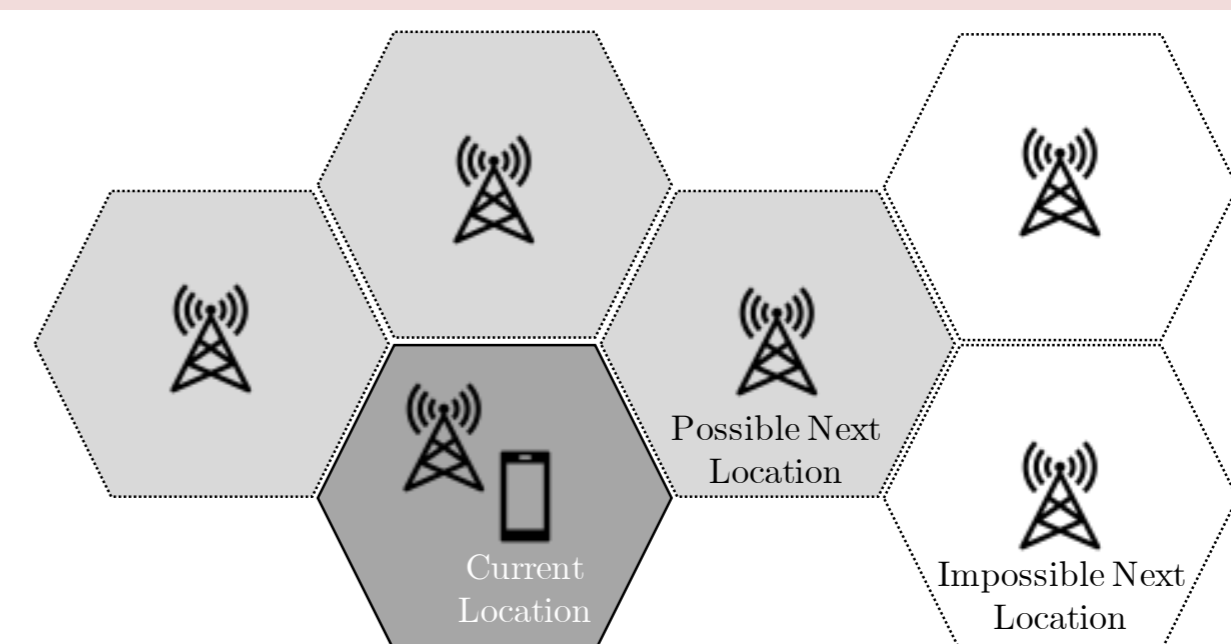
- Use past access patterns + *geo-hints* from application
- Use probabilistic models to predict future data access
- Decentralized for scalability in geo-distributed Fog environments

ARCHITECTURE



- Hybrid approach with custom middleware and prediction on node preserves privacy, but takes load off client
- Nodes leverage probabilistic methods (e.g., Markov models) to predict future client access
- Edge nodes proactively replicate data to peers and remove their own replica

- Topology limits the number of nodes that must be considered as possible next locations
- Data movement policies can be respected



FUTURE WORK

Implementation and Evaluation:

- Different sources of location hints (past access patterns, past location, hinted future locations, etc.)
- Different prediction methods (Markov models, ARIMAX time-series analysis, AI/ML, etc.)
- Compare other replica management approaches: scalability, performance overhead, privacy in context of data movement policies