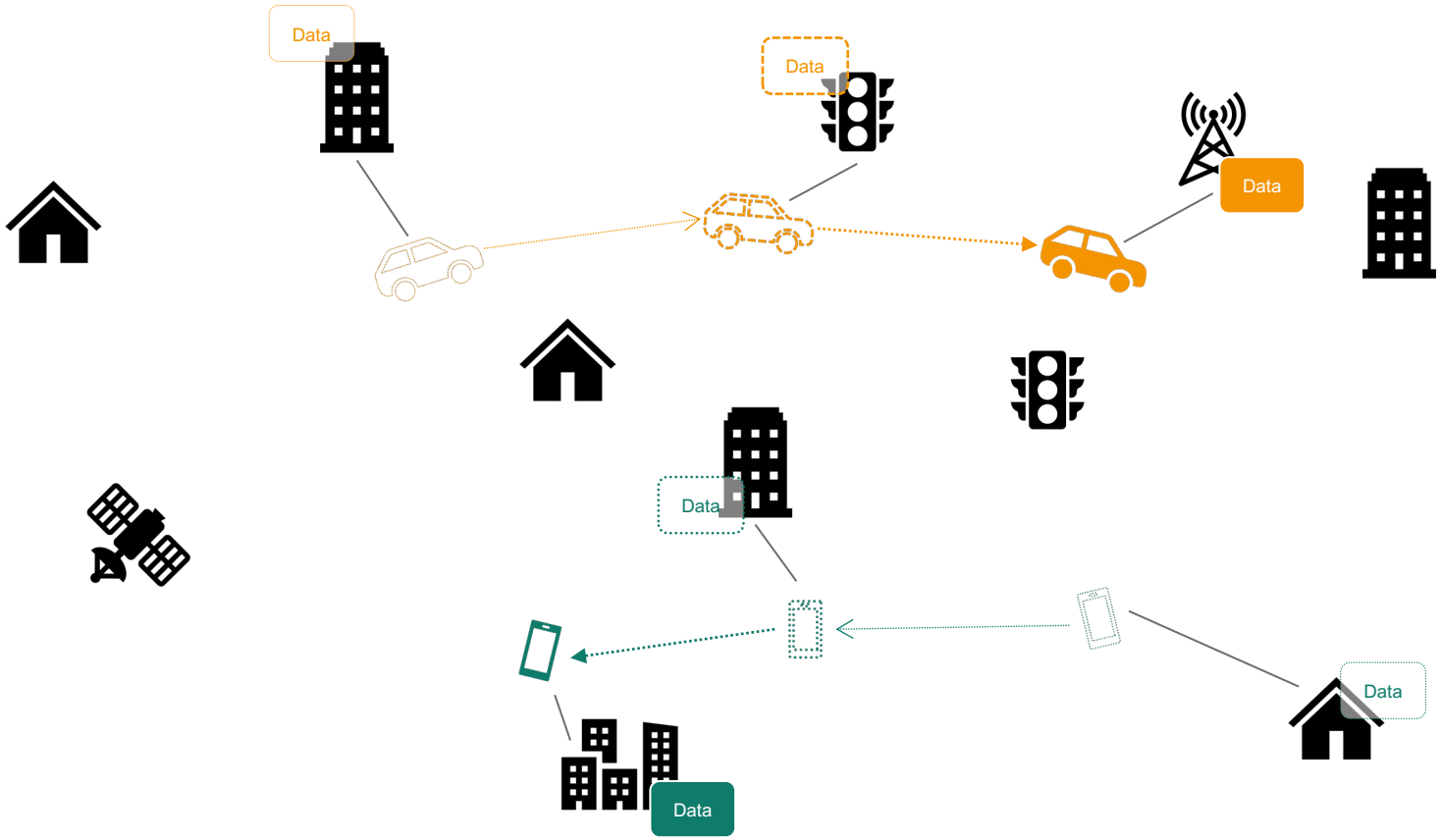
Technische Universität Berlin

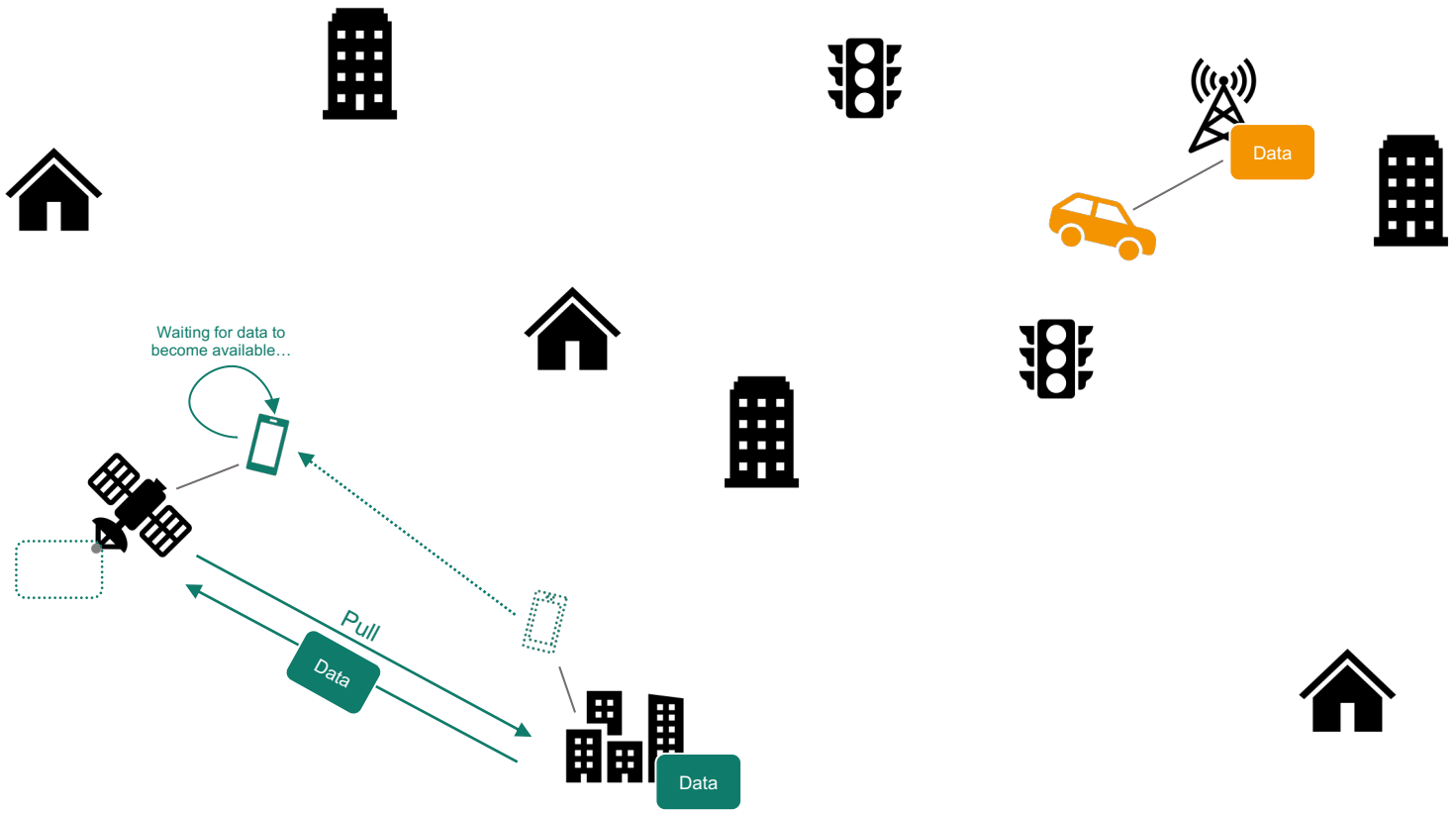
Predictive Replica Placement for Mobile Users in Distributed Fog Data Stores with Client-Side Markov Models

Malte Bellmann, **Tobias Pfandzelter**, David Bermbach | DML-ICC

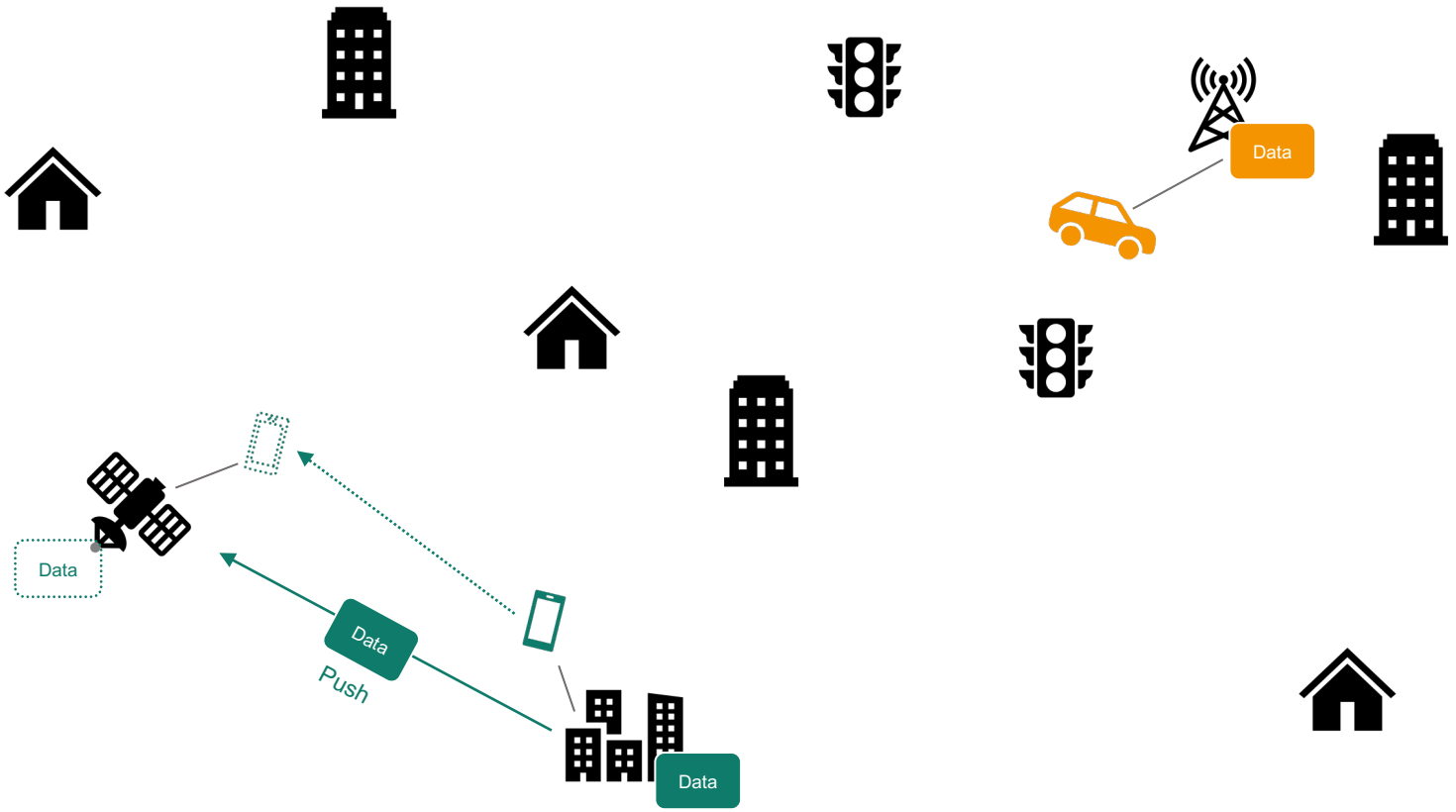
Mobile Users Move Through the Fog and Require Data Access at Their Nearest Fog Node



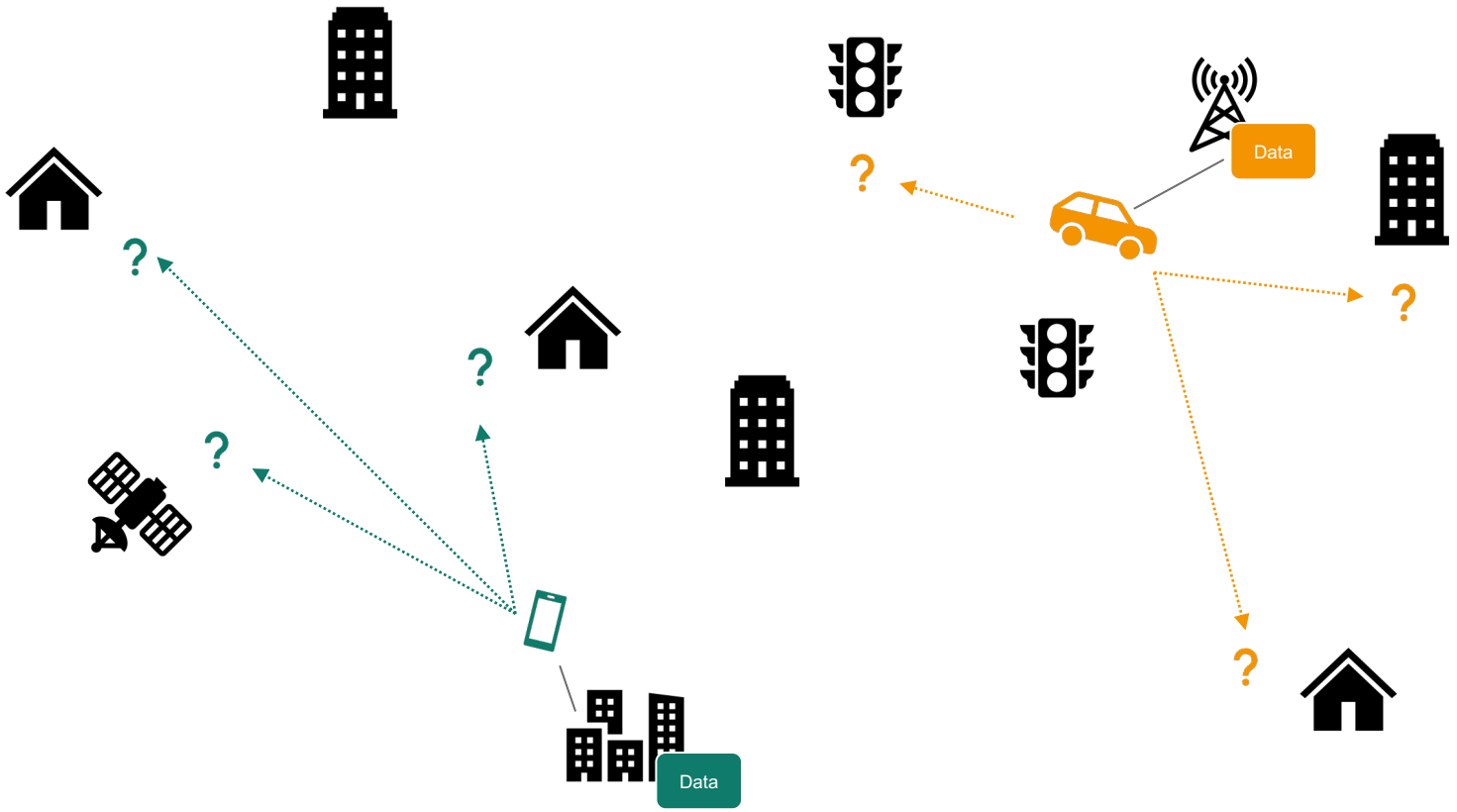
Pulling Data at Every Move Leads to Significant QoS Impairments



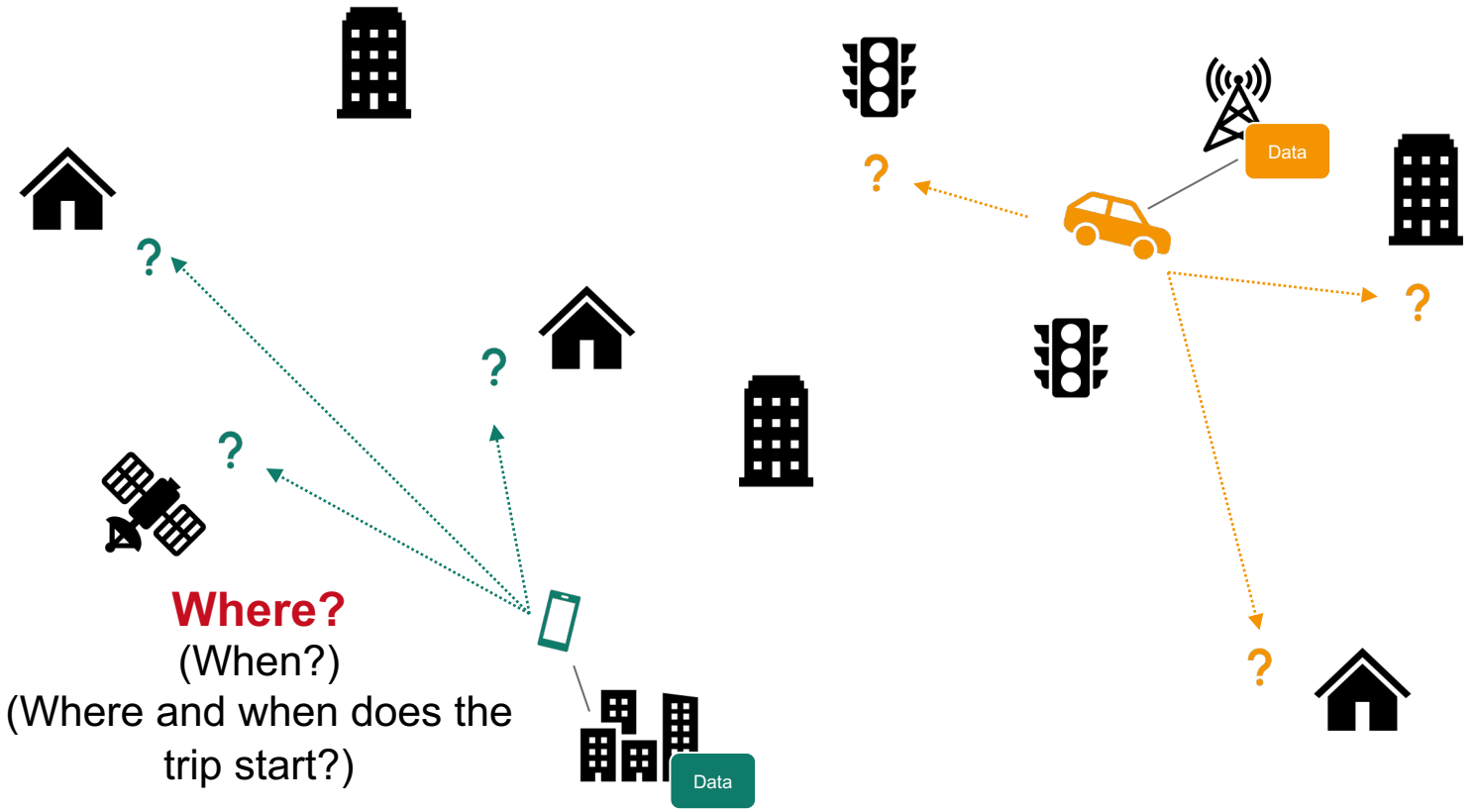
Proactive Data Replication?



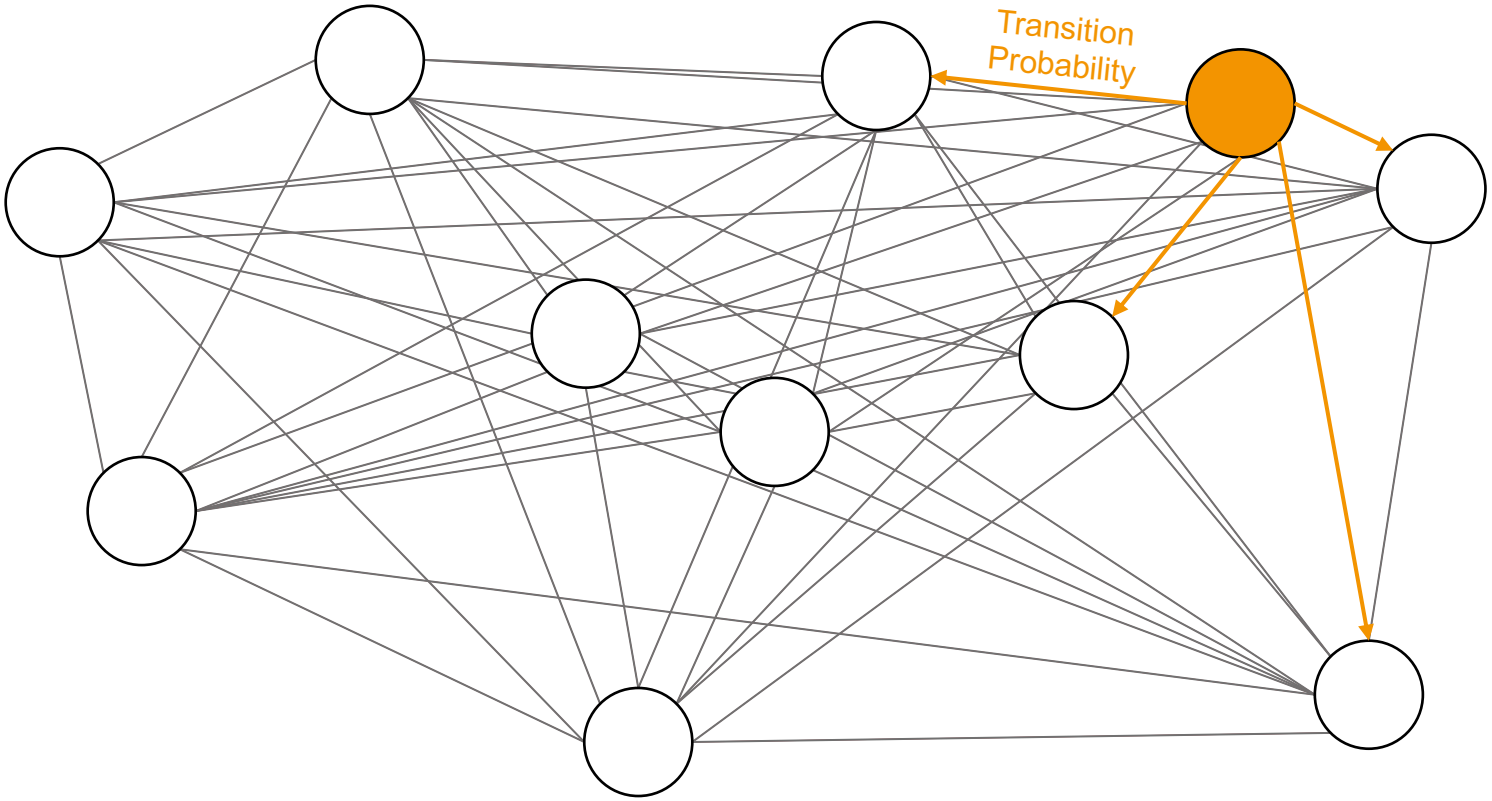
Requires Location Prediction!



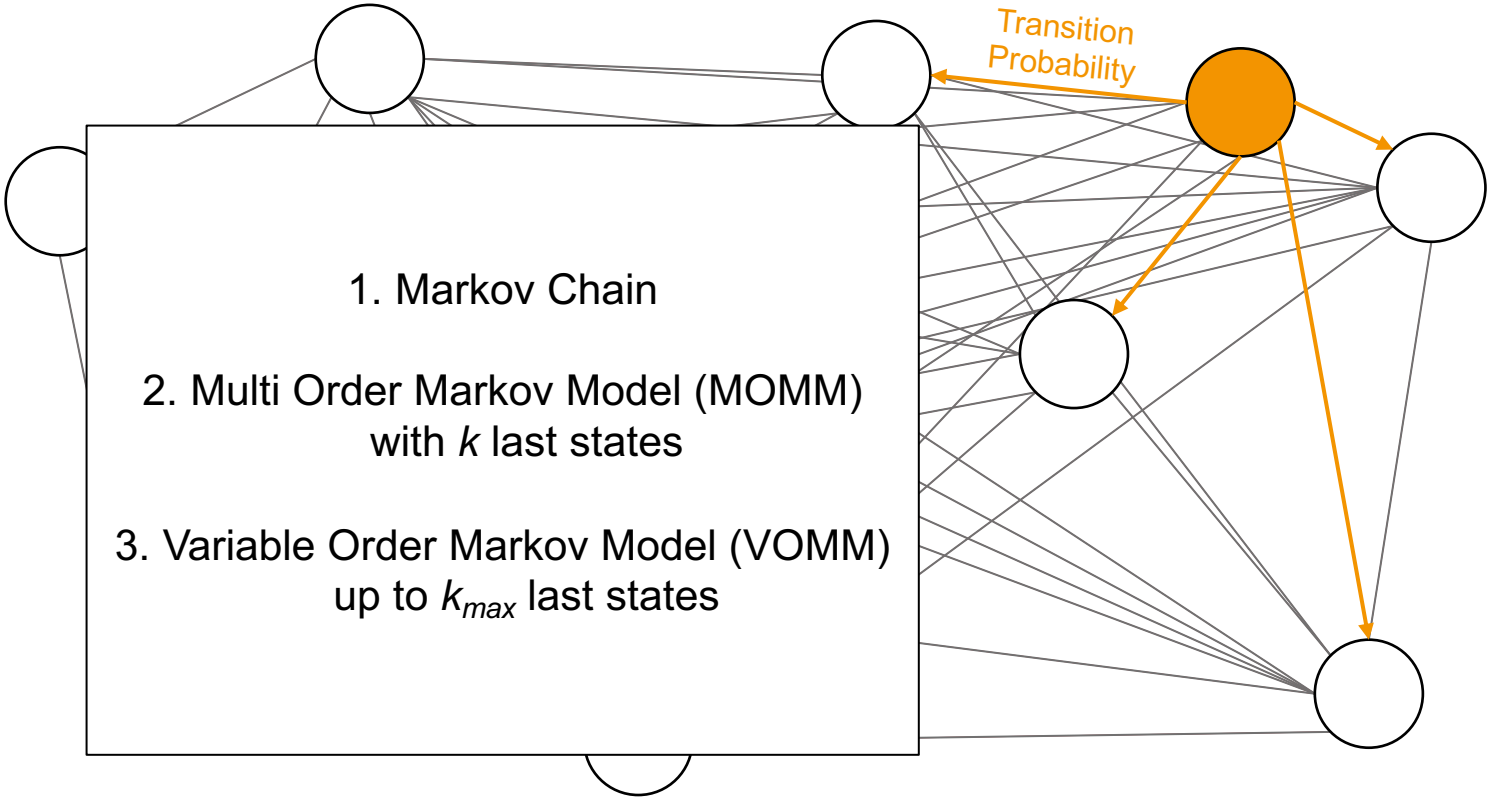
Requires Location Prediction!



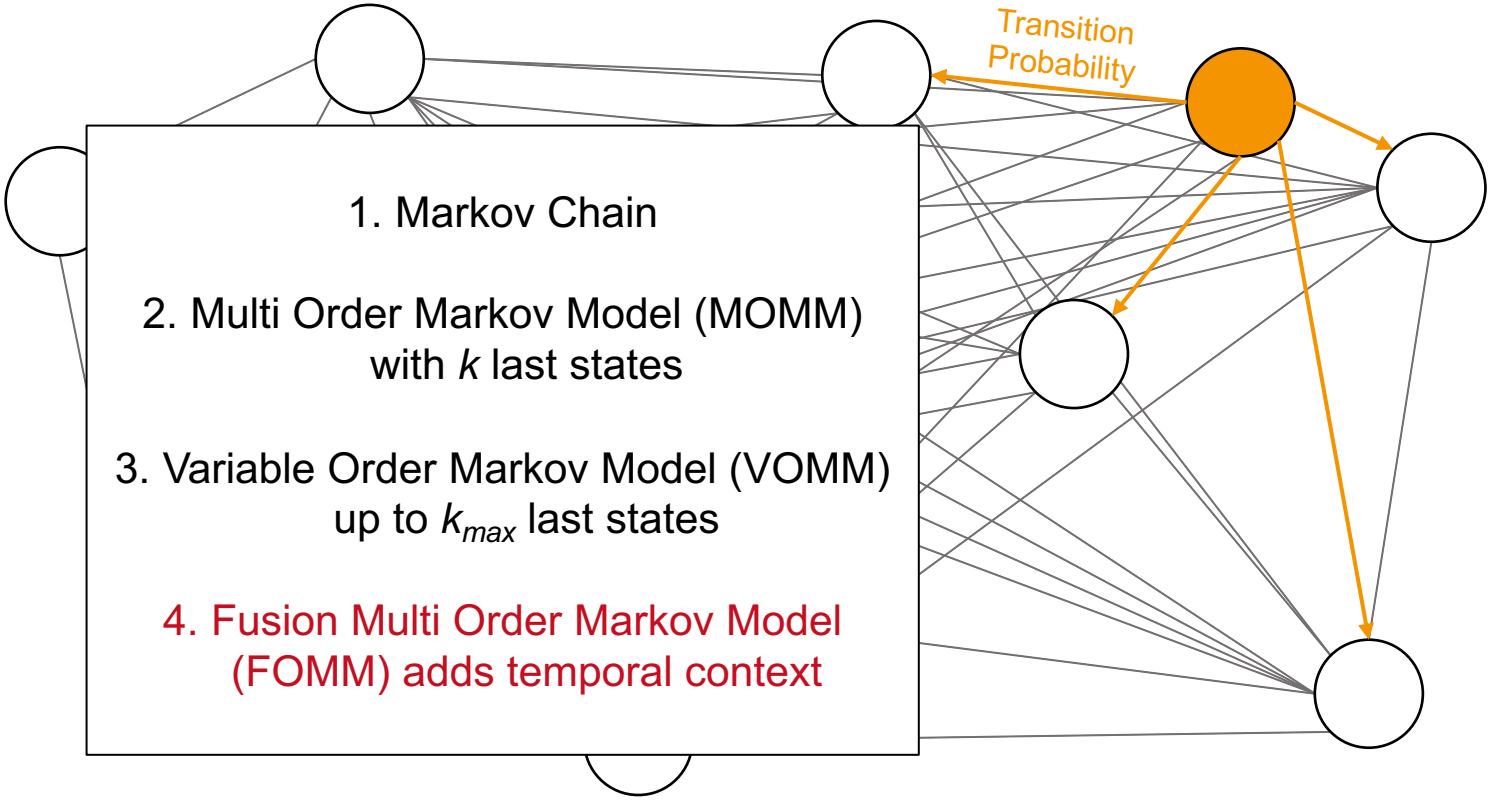
Fog as a Connected Graph



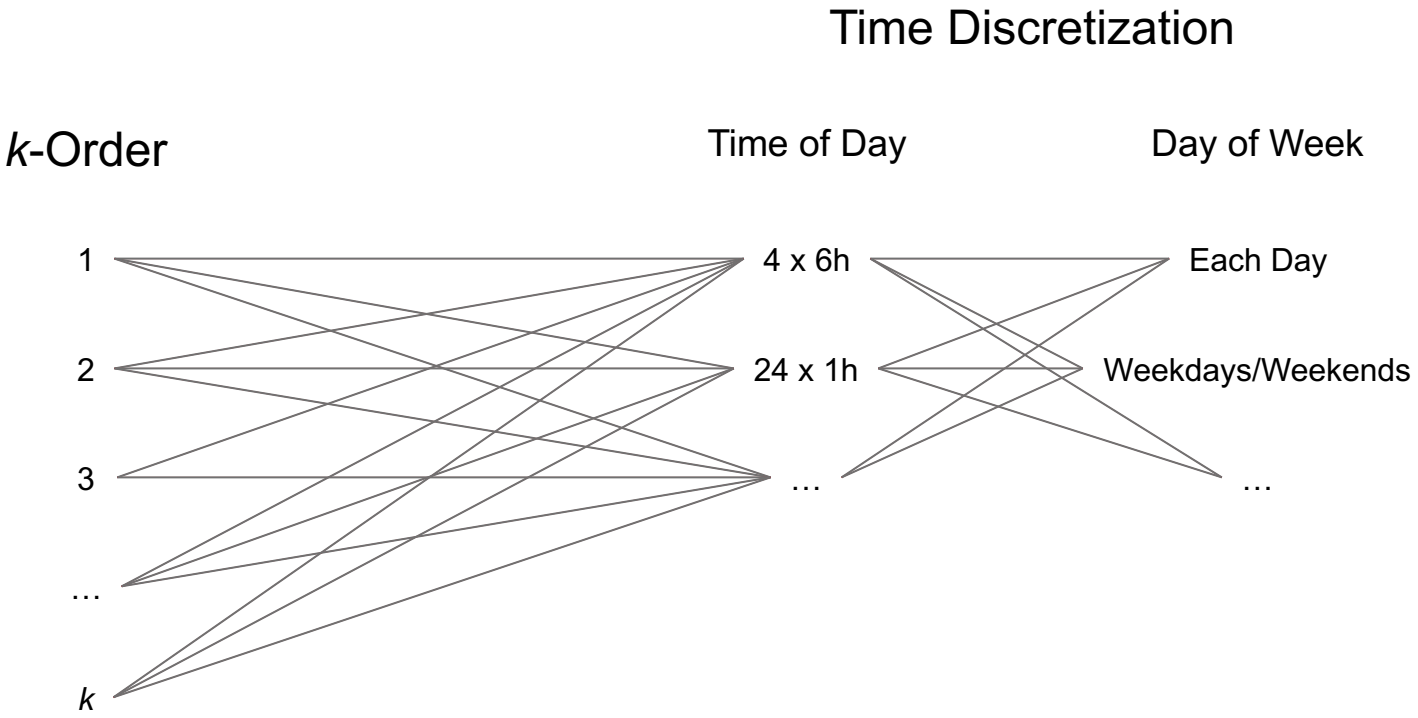
Markov Model Approach



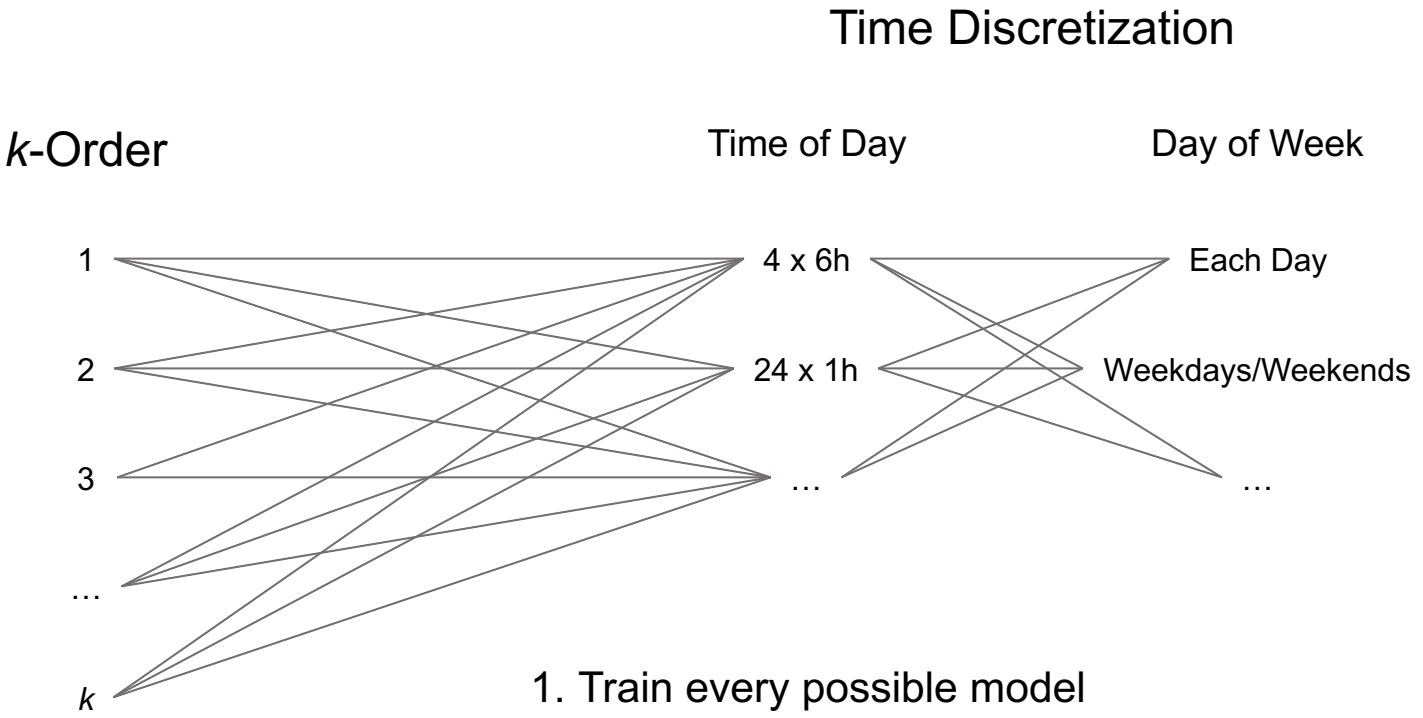
Markov Model Approach



Fusion Multi Order Markov Model (FOMM)



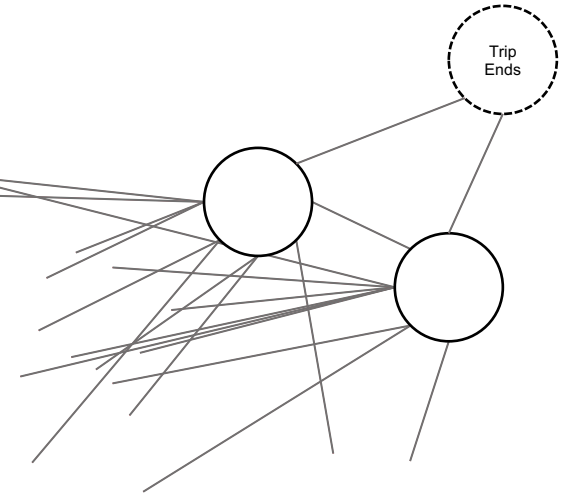
Fusion Multi Order Markov Model (FOMM)



1. Train every possible model
2. Assign weights to models based on specificity
3. Collect available results and return weighted sum

Extensions

End-of-Trip

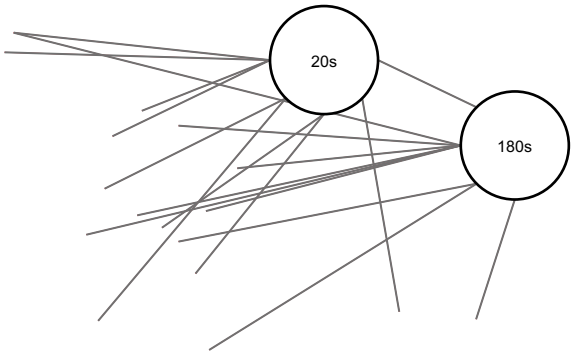


Additional state indicates that trip ends

Dynamic *topN*

Given target probability *P*, return most probable next States so that their Cumulative probability is ≥ *P*

Duration



Model has information on duration of a stay at a node

Client vs Server-Side Models

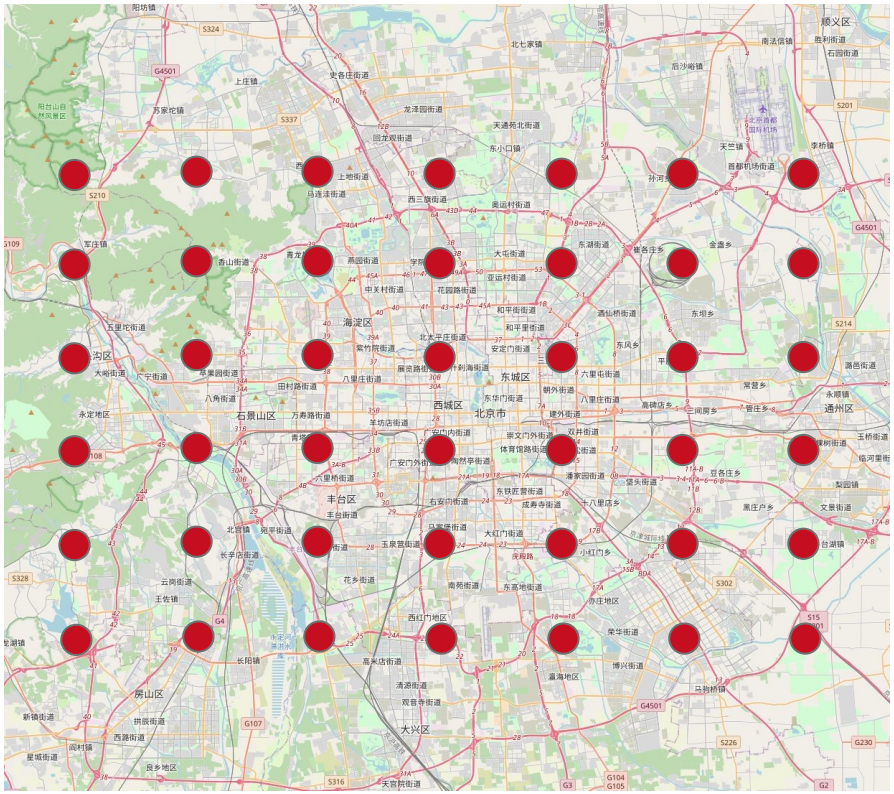
Client-Side	Server-Side
+ Scalability	- Bottleneck
+ Privacy	- Movement data sent to central location
- Increased resource consumption	+ No client resources required

→ Yet all algorithms can also be used as server-side

Evaluation – Simulation

Parameters:

- 10x10 (30x30) Node Grid on Beijing, China
- Fixed Data Transfer Time/Limited Bandwidth
- GeoLife GPS Trajectory Dataset (182 Users) [1-3]



Evaluation – Metrics

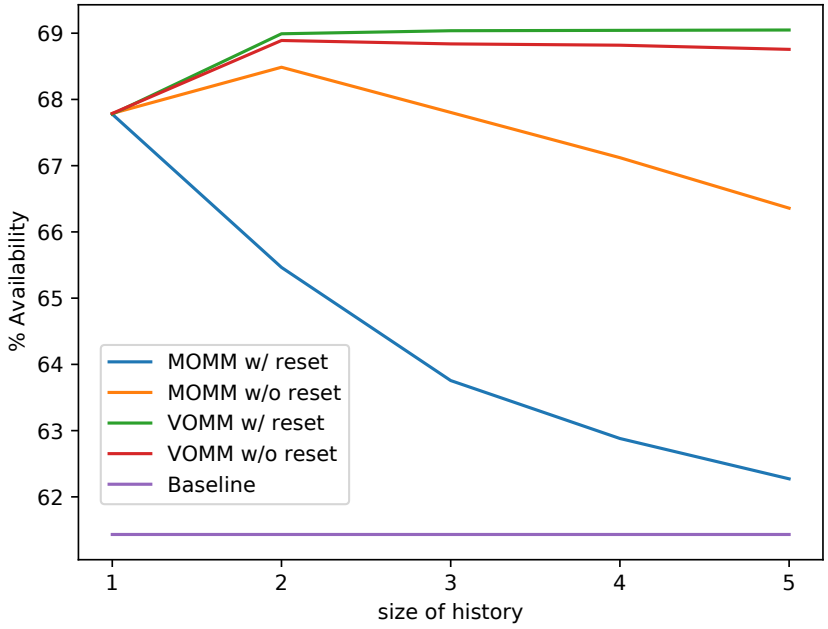
Data Availability

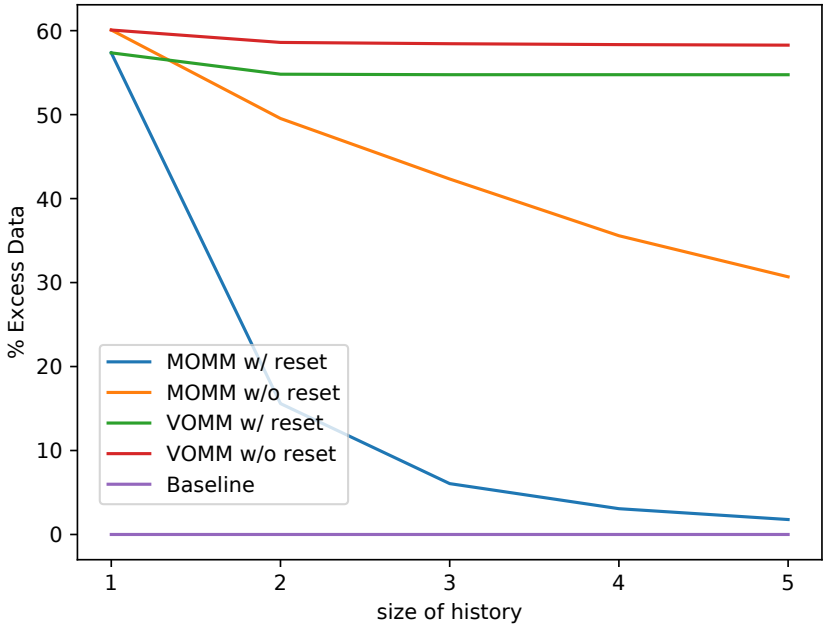
$$\frac{\text{Time Client Can Access Their Data at Nearest Node}}{\text{Total Time}}$$

Excess Data

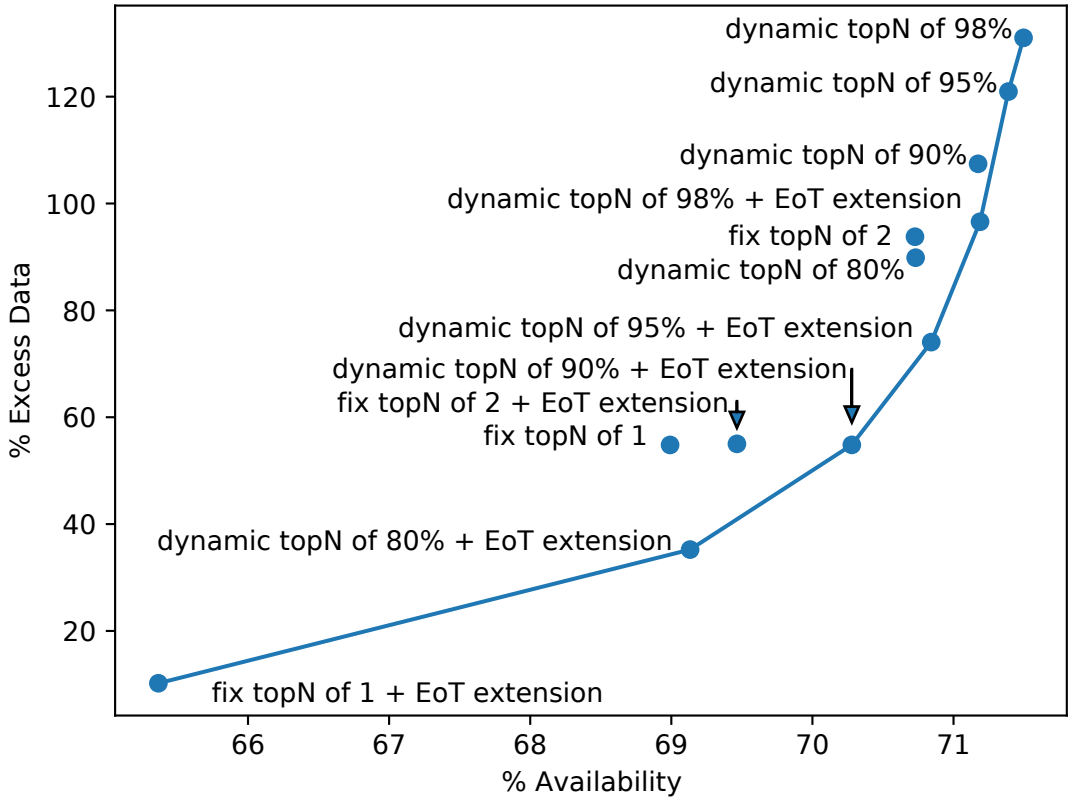
$$\frac{\text{Total Data Stored} - \text{Data Stored at Right Node}}{\text{Total Data Stored}}$$

Evaluation – Baseline, MOMM, VOMM

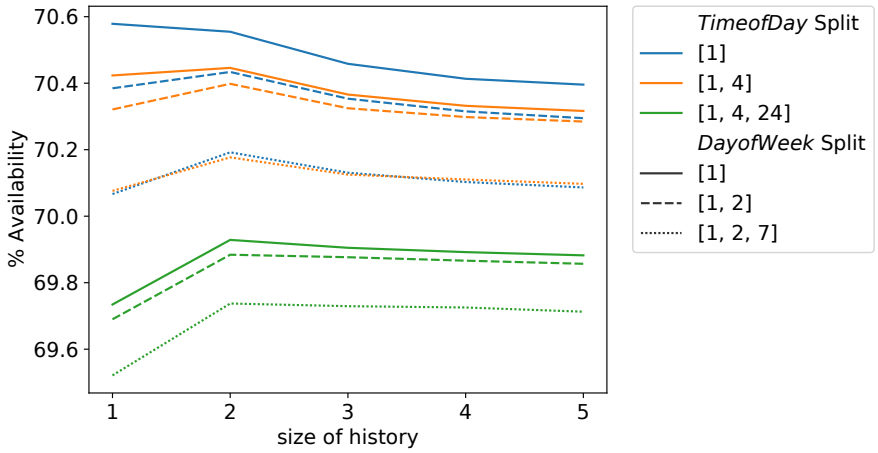


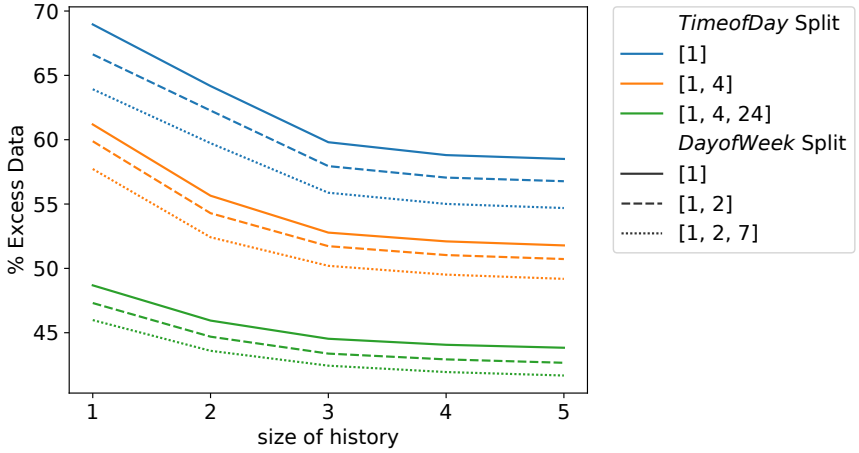


Evaluation – Extensions (on VOMM)



Evaluation – FOMM





→ Better performance than VOMM
→ More metadata can reduce excess data

In our paper…

… Startup/Pause Prediction

… Simulation with a more complex network model

… Combination of Approaches

… Related Work

… Detailed Discussion of our Work

Conclusion & Future Work

- We need strategies for frequent data migration in the fog
- Markov models are inherently suited for this kind of task
- There might not be *one* perfect approach for each kind of application
- We may consider additional data sources, such as *client hints*
- Semantically splitting data/partial data replication could be feasible

Contact: tp@mcc.tu-berlin.de

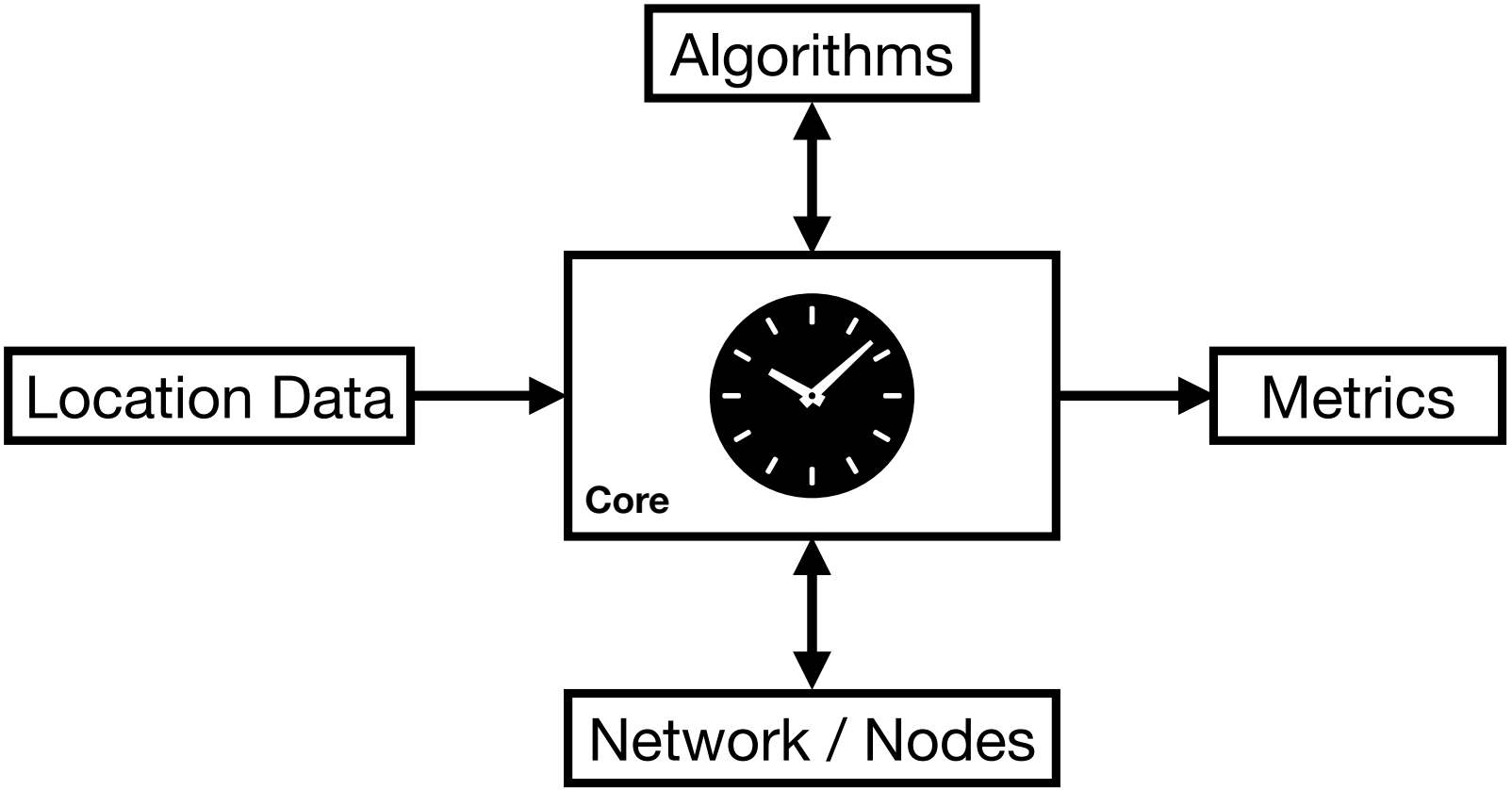
References

[1] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding Mobility Based on GPS Data. In *10th International Conference on Ubiquitous Computing*. 312–321.

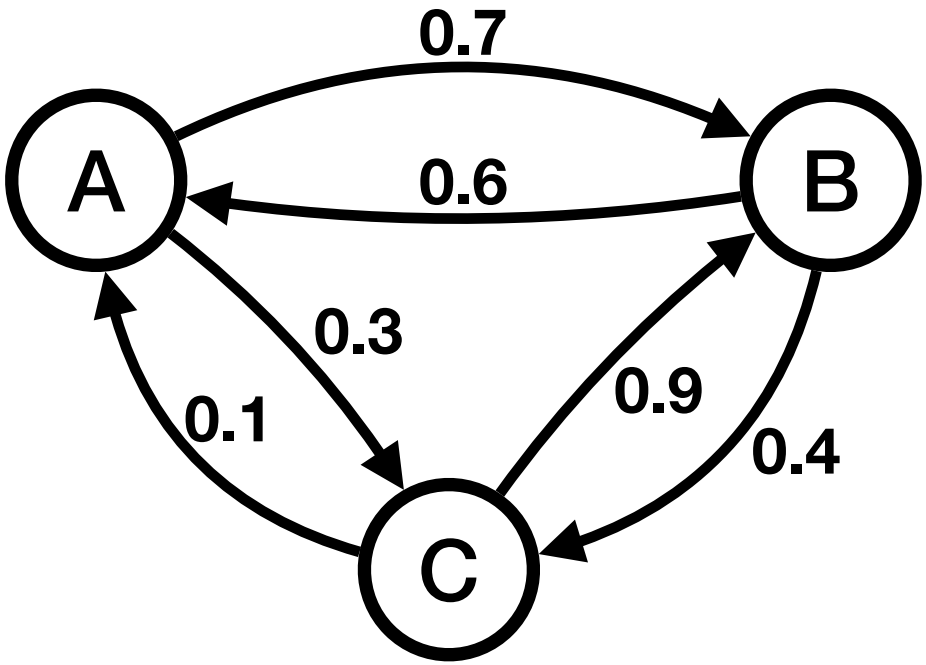
[2] Yu Zheng, Xing Xie, and Wei-Ying Ma. 2010. Geolife: A Collaborative Social Networking Service among User, Location and Trajectory. *IEEE Data Engineering Bullettin* 33, 2 (2010), 32–39.

[3] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *18th International Conference on World Wide Web*. 791–800.

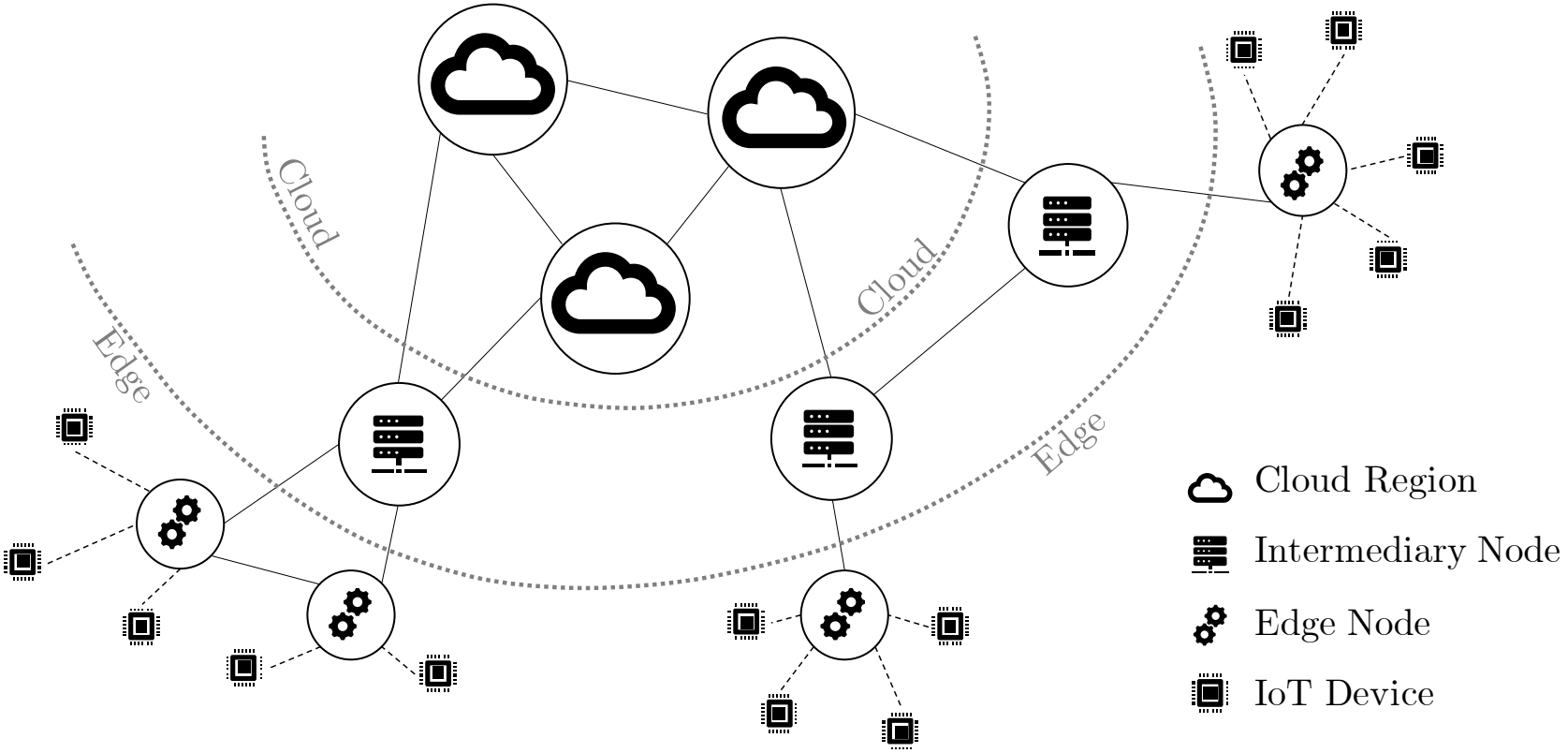
Simulation



Markov Model



Fog Architecture



Performance Over Time