Large-Scale Mobility Data Analytics

Distributed Convoy Pattern Mining

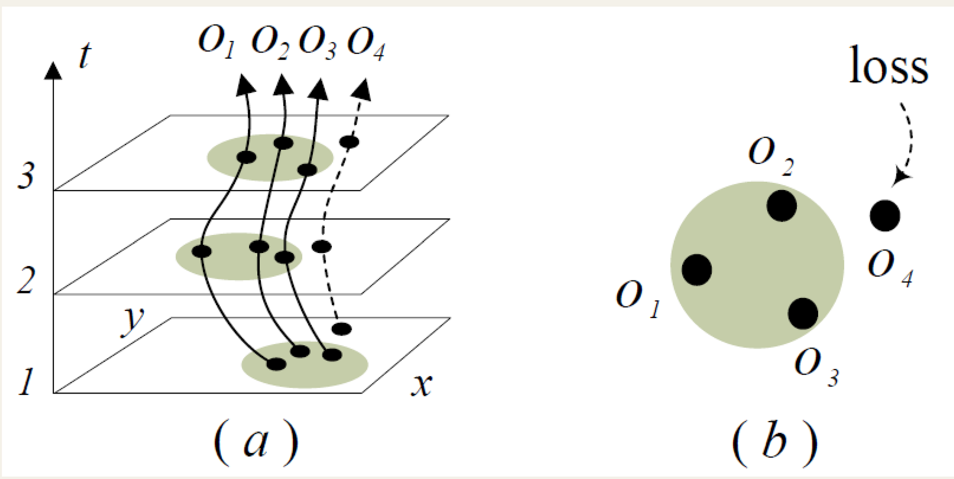
Faisal Orakzai, Thomas Devogele and Toon Calders

Universite' Libre de Bruxelles and University Francois Rebelais Tour

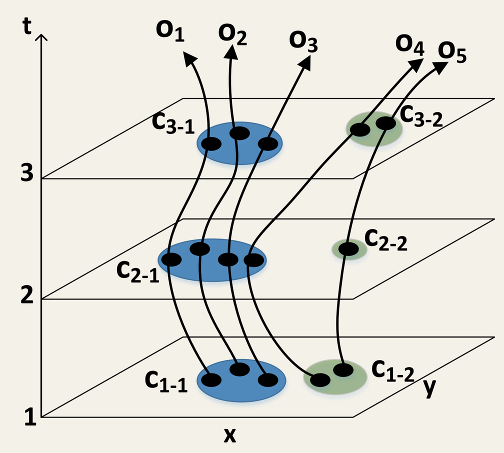
Absctract

Due to the wide spread of mobile devices equipped with location sensors, the amount of mobility data being generated is enormous. Mining this data to reveal interesting behavioral patterns has gained attention in recent years. Various mobility patterns have been proposed which describe collective mobility behaviour. One such pattern is the convoy pattern which consists of at least m objects moving together for at least k consecutive time instants where m and k are user-defined parameters. Existing algorithms for detecting convoy patterns, however do not scale to real-life dataset sizes. Therefore in this paper, we propose a generic distributed convoy pattern mining algorithm and show how such an algorithm can be implemented using the MapReduce framework. Our experimental results show that our distributed algorithm is scalable and more efficient than the existing single instance convoy pattern mining algorithms.

Flock Pattern



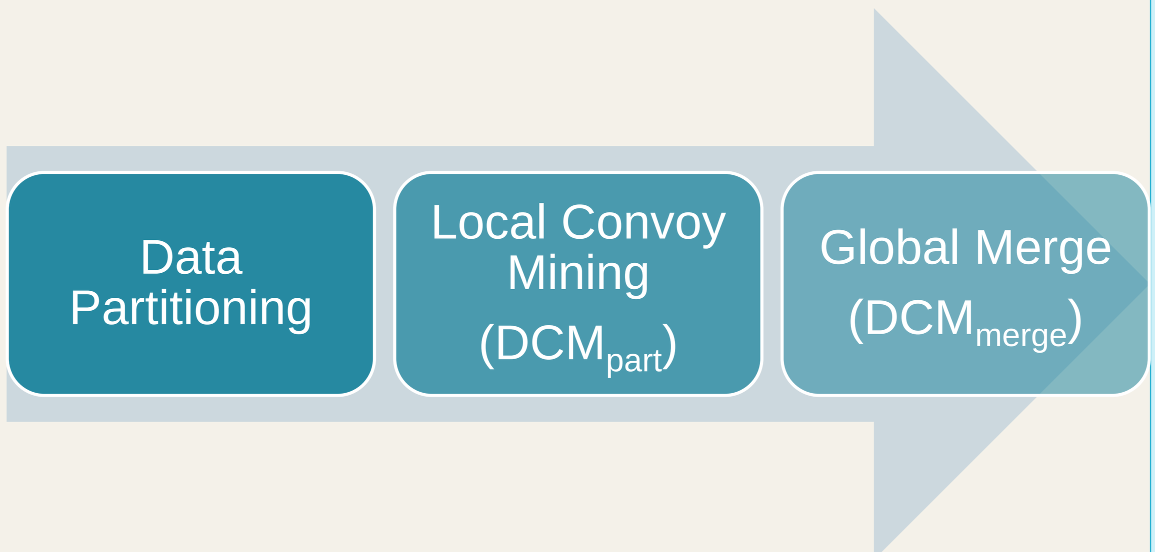
Convoy Pattern



Existing Algorithms:

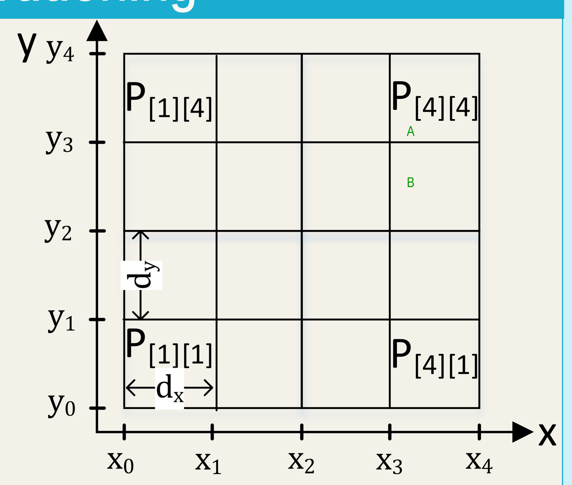
1. Coherent Moving Cluster (CMC)
2. Convoy Mining using Trajectory Simplification (CuTS, CuTs+, CuTS*)
3. VCoDA & Partially Connected Convoy Discovery (PCCD)

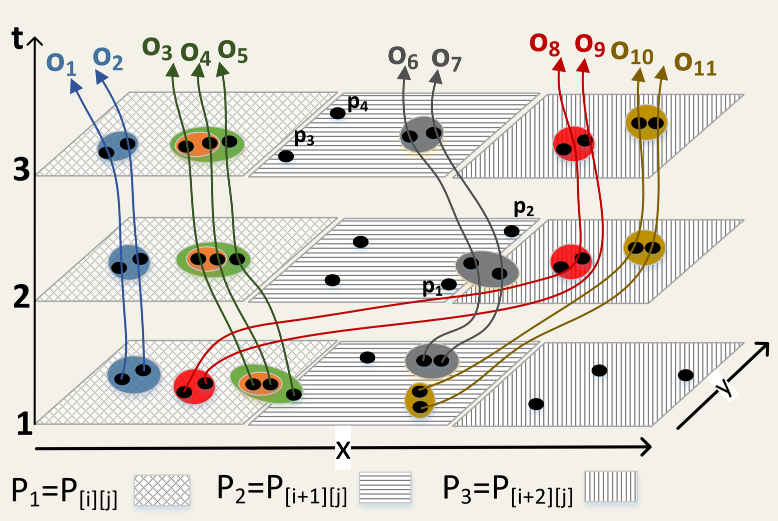
Distributed Convoy Mining



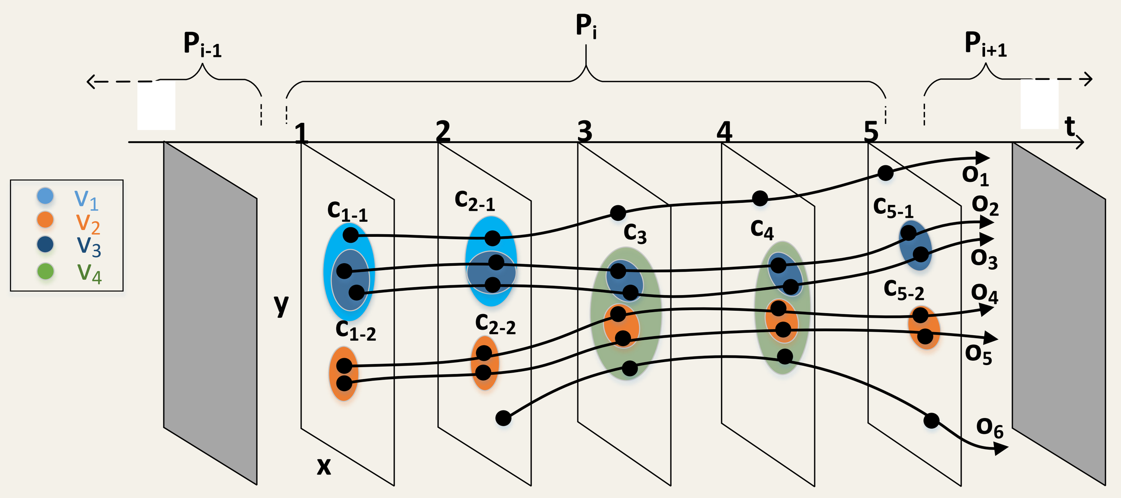
Spatial Partitioning

1. Partial or missing clusters
   a) Overlapped Partitions
   b) Reduced m
2. Partial convoys
3. Fast moving objects
4. DBSCAN cost





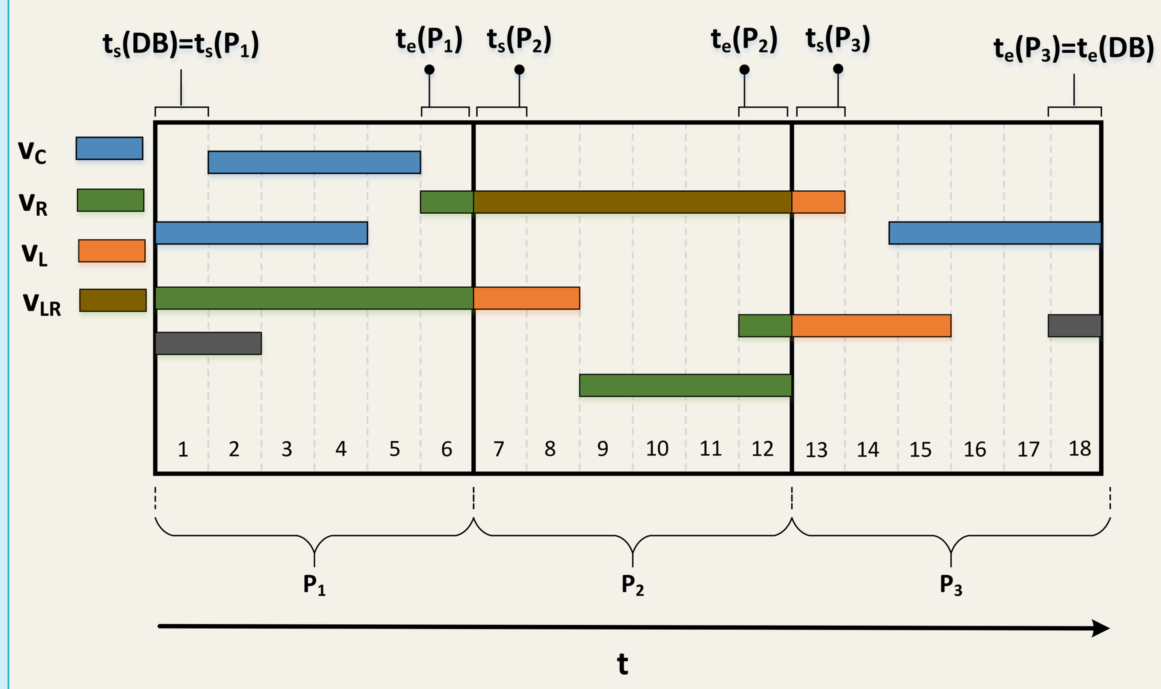
Temporal Partitioning



DCM_{part}

Types of Convoys:

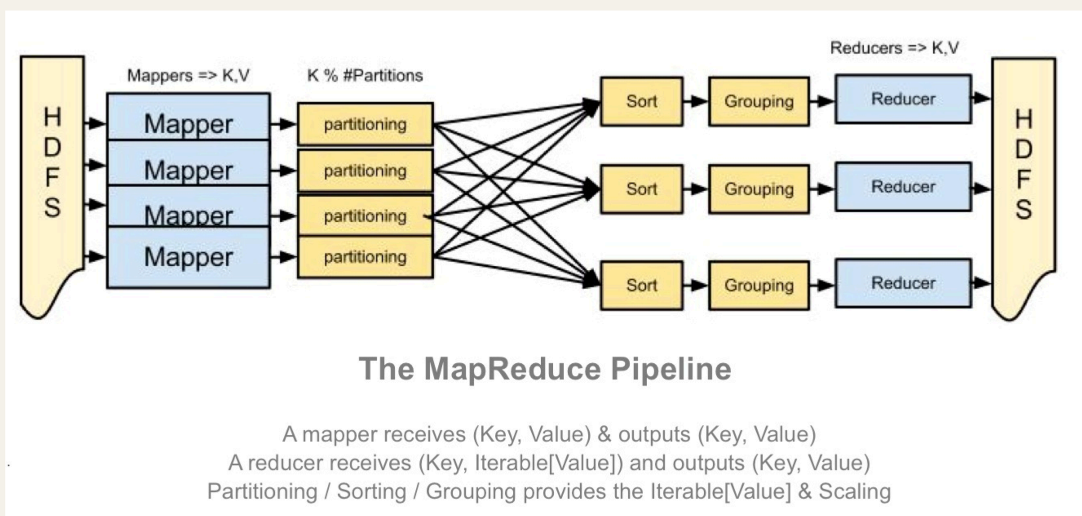
1. Closed convoys
2. Left-open convoys
3. Right-open convoys
4. Left-right-open convoys.



DCM_{merge}

1. Runs on a central node
2. Sorts the data based on merge-time
3. Merges the open convoys to form closed true convoys

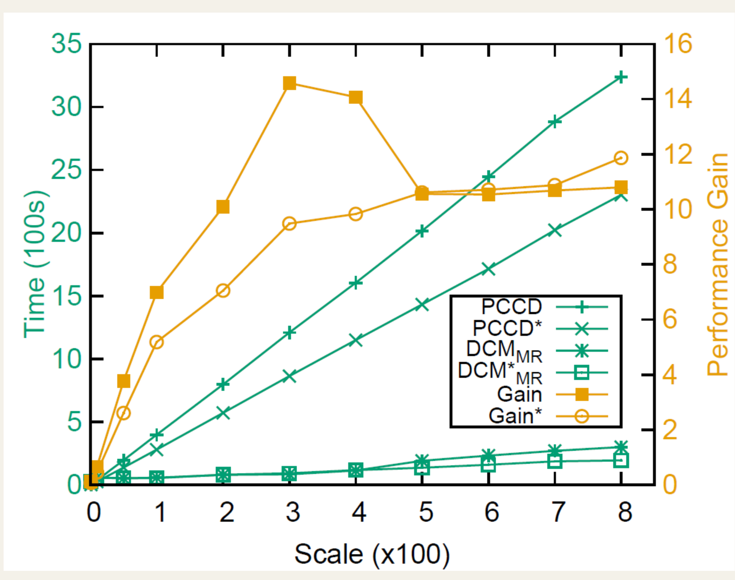
DCM_{MR}



1. Partitioning based on HDFS block size
2. Custom Practitioner and Group Comparator for accuracy
3. Single key to force one reducer
4. Secondary sorting on merge-time

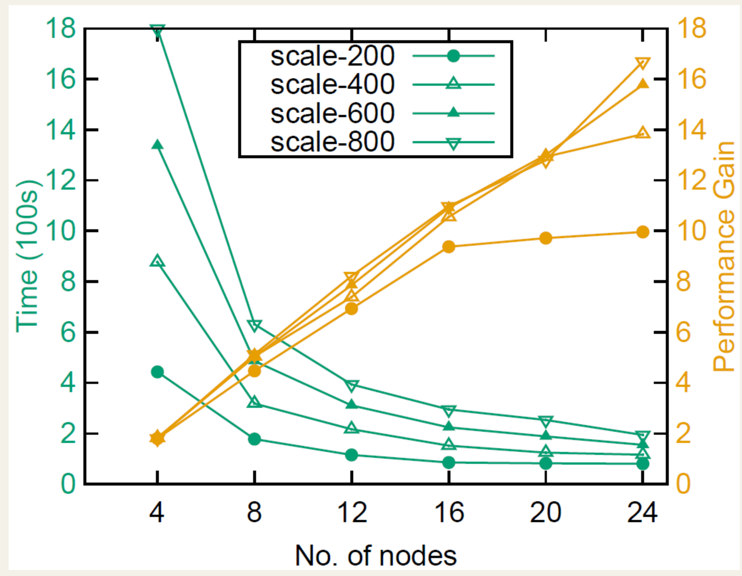
Experimental Results

Data Scalability

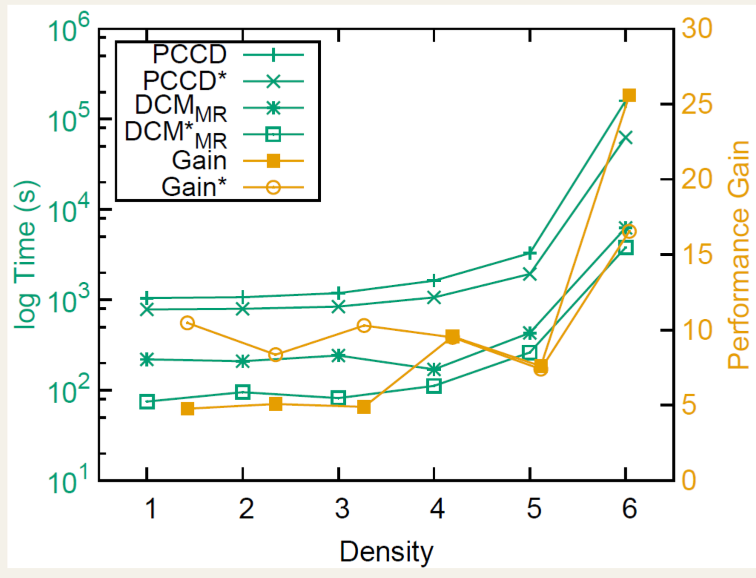


Experimental Results

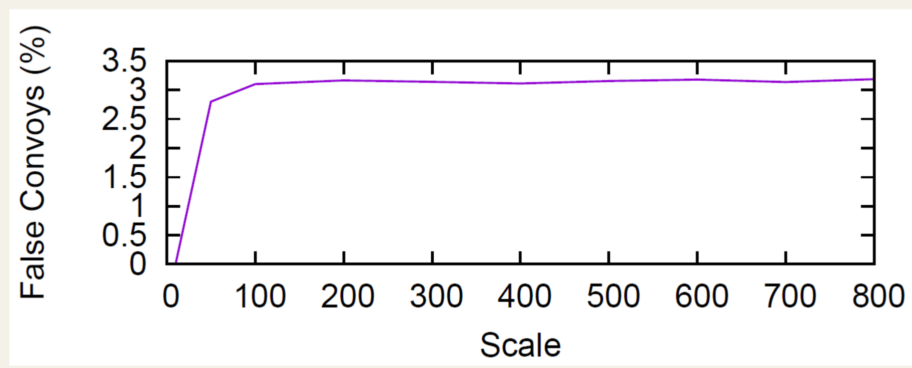
Node Scalability



Density Scalability



False Convoys



Conclution

Convoy pattern mining is computationally expensive and existing algorithms do not scale up to the huge amounts of movement data. In this paper we propose a generic, framework independent and highly scalable distributed convoy pattern mining algorithm DCM which outperforms existing algorithms by a high margin. We analyze spatial and temporal data partitioning strategies and choose temporal partitioning over spatial partitioning because of its merits for convoy pattern mining. We implement our generic algorithm in the MapReduce framework, with and without using KD-Tree index. We test our algorithm extensively on datasets with sizes varying from a couple of thousand records up to 293 million records and number of objects varying from a couple of hundreds to a quarter of million . The results validate the high performance and scalability of our algorithm w.r.t both the cluster and dataset sizes.

Future Work

1. Convoy Pattern Mining over streams
2. Distributed convoy pattern mining over streams
3. Efficient distributed indexing for NN-search
4. Collaboration with other PhD researchers

References

[1] H. Jeung, H. T. Shen, and X. Zhou. Convoy queries in spatio-temporal databases. In Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, pages 1457{1459. IEEE, 2008.

[11] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. Proceedings of the VLDB Endowment, 1(1):1068{1080, 2008.

Contact Information

Faisal Orakzai

ofaisal@ulb.ac.be

Tel: 066 - 7128671