Adaptive Checkpointing for
Non-stationary Graph Algorithms
on Distributed Dataflow

Broadly speaking, graph algorithms can be divided into two categories (based on their communication characteristics across supersteps): stationary and non-stationary [1]. An algorithm is stationary if its active vertices send and receive the same distribution of messages across supersteps. A graph algorithm is non-stationary if the destination or size of its outgoing messages changes across supersteps. The variations in non-stationary algorithms can create workload imbalances across supersteps. That is, the execution time of each superstep might keep changing. Hence, a static interval to checkpoint the vertices is oblivious to such variation of superstep execution. We aim to propose an adaptive checkpointing strategy to consider this variation during checkpointing.

The goal of this project
is to evaluate the execution of typical non-stationary graph algorithms, e.g., Distributed Minimal Spanning Tree (DMST) and Simulating Advertisements on Social Networks, on dataflow engines like Flink and Spark and propose an adaptive strategy for checkpointing.

Reference


For a detailed introduction to topic please get in contact via email.

Advisor and Contact:
Chen Xu <chen.xu@tu-berlin.de>