Bounding the Makespan of Transaction Schedules

Tim Baccaert

Software Languages Lab, Vrije Universiteit Brussel, Belgium
tim.baccaert@vub.be

The performance of transactional database systems is typically evaluated by measuring the amount of transactions they can commit to the database per second, better known as the transaction throughput. However, fairly measuring this for the same workload on different systems is not trivial, as each system has different characteristics. There is hardly any work that analyzes schedule efficiency directly, and the space of all possible efficient schedules is largely unknown. Hence we cannot identify whether there is any room for improvement in the algorithms used for concurrency control. Prior transaction theory largely centers on decision problems relating to safety, such as the serializability, robustness, and allocation problems. Most pertinently, these problems take already scheduled transactions as input, and do not directly consider the efficiency of those schedules.

In this presentation, I want to show our current progress at mapping out this space of efficient schedules. This includes a formalization of the transaction scheduling problem, a proof of NP-completeness, and various upper and lower bounds on schedule efficiency for specific classes of transaction sets.

Furthermore, I want to briefly show some of the experiments we are currently doing. I have encoded the transaction scheduling problem as a mixed integer linear program, and use this to compute the optimal schedules for various common benchmark workloads such as Smallbank or TPC-C. This then allows us to compare the real execution (of say a MVCC or 2PL-based protocol) to the optimal possible execution. This part is work in progress, and I welcome all the feedback I can get on it.