A common practice to minimize contention within databases is the application of caching. One of the caching mechanisms - "write behind" - allows request results to be stored on the cache and afterwards transferred at once to the database storage. Therefore, the database management system has to synchronize fewer times. In such a way, the overall performance can be highly increased.

The current research is modeling and optimizing a queueing network with features inspired by the "write behind" mechanism. More precisely, we consider a system in which it is possible to postpone the service of a request by storing it in a finite size buffer. Therefore, the server is able to perform two different types of work. The first one is pre-processing and putting requests into the buffer, one at a time. The second one is processing jobs, possibly multiple at once. Once processed, the jobs leave the system. In such a way, jobs can be grouped to an optimal level and served all together as a batch. We take the service time of a batch to consist of an initialization period followed by a processing time that depends on its size. The goal is to find the optimal dynamic level strategy that minimizes the average waiting time in the queue.