Controlling Multi-Echelon Inventory Systems with Waiting Time Fill Rate Constraints
Jacob Agardh & Ludwig Brieditis
Industrial Engineering and Management, Lund University

For many manufacturers of industrial equipment and machines it is becoming more and more common to offer your products as a service rather than a single physical product. A key part of the service offering is guaranteeing a high uptime of machinery through availability of critical spare parts. Ensuring high global availability of spare parts often leads to the deployment of complex distribution networks consisting of multiple levels of warehouses and dealers as well as millions of dollars tied up in spare parts across the globe. What if a tradeoff could be made with customers to lower the amount of needed inventory?

A possible trade-off dimension is lead time. More specifically, the contract between the manufacturer and customer could state that the agreed upon availability should be maintained within a couple of days from announcing their order, a so called allowable waiting time. As an example, a target could be that 95% of demand should be satisfied within 2 days. In most cases demand could still be satisfied immediately, but it would incentivise customers to announce their orders ahead of time when possible. Early announcement of orders can make sense in situations when the part will be used for planned maintenance or when the customer can perform other maintenance while replacing the part.

However, in order to actually use a service measure that includes an allowable waiting time a couple of problems must first be solved: (1) How can the resulting waiting time service of the manufacturer’s network be estimated? (2) How can (near-)optimal reorder points be found efficiently for the system such that a target waiting time service is fulfilled? (3) In what cases can the optimization model be expected to reach the target?

It was found that the waiting time service can be estimated using a result called “lead time shift”. This means that an allowable waiting time can be included in an analytical model by simply modifying the lead time by subtracting the allowable waiting time.

The lead time shift is a quite general result meaning that it can be used in basically any optimization algorithm. For this project the chosen model was the induced backorder cost optimization model by Berling and Marklund of LTH.

The performance of the optimization model was analyzed through simulation of a number of test series based on data from Syncron’s customers. The results showed that dealers’ customer arrival rate and parameters that affect the variability of the central warehouse’s performance had a high impact on the model’s ability to reach the target service. In cases where the variability was caused by negative warehouse reorder points, requiring the warehouse’s reorder point to be equal to or greater than -1 greatly improved performance. The inclusion of an allowable waiting time slightly reduced the model’s performance. This negative effect was slightly larger for dealers with larger customer arrival rates.

For a series of different products the inclusion of an allowable waiting time of 5 days with the same availability target lowered the achieved availability by 1.1 percentages, but simultaneously lowered the required inventory by 11%. 