



HAL
open science

CARV Special Issue: Real-time, cooperative enterprises for customized mass production

Laszlo Monostori, Tamás Kis, Josef Vancza, Botond Kádár, Gábor Erdős

► **To cite this version:**

Laszlo Monostori, Tamás Kis, Josef Vancza, Botond Kádár, Gábor Erdős. CARV Special Issue: Real-time, cooperative enterprises for customized mass production. *International Journal of Computer Integrated Manufacturing*, 2009, 22 (01), pp.55-68. 10.1080/09511920802369324 . hal-00513407

HAL Id: hal-00513407

<https://hal.science/hal-00513407>

Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CARV Special Issue: Real-time, cooperative enterprises for customized mass production

Journal:	<i>International Journal of Computer Integrated Manufacturing</i>
Manuscript ID:	TCIM-2007-IJCIM-0127.R1
Manuscript Type:	Special Issue Paper
Date Submitted by the Author:	10-Jun-2008
Complete List of Authors:	Monostori, Laszlo; SZTAKI, Engineering and Management Intelligence Kis, Tamás; SZTAKI, Engineering and Management Intelligence Vancza, Josef; MTA SZTAKI, Engineering and Management Intelligence Lab Kádár, Botond; MTA SZTAKI, Engineering and Management Intelligence Lab Erdős, Gábor; MTA SZTAKI, Engineering and Management Intelligence Lab
Keywords:	PRODUCTION MANAGEMENT, SCHEDULING, SUPPLY CHAIN MANAGEMENT
Keywords (user):	coordination, monitoring & control



CARV Special Issue

Real-time, cooperative enterprises for customized mass productionL. Monostori^{1,2}, T. Kis¹, J. Váncza^{1,2}, B. Kádár¹, G. Erdős¹

The paper discusses the main requirements that are posed by customized mass production towards managing production networks. Special emphasis is put on real-time, cooperative behaviour. We present a large-scale national industry-academia R&D project aimed at improving the performance of a production network that produces consumer goods in large quantities and variability. An integrated approach is outlined for planning and scheduling the behaviour of the system at network-, factory- and plant levels, as well as for adapting the various plans and schedules to real execution conditions. Novel, integrated solutions are described for rolling horizon production scheduling of a factory with more than 100 production lines, for real-time control of daily production supported by information fusion and simulation, as well as for cooperative component supply. The industrial deployment of the integrated systems is also presented together with some lessons of their routine application.

Keywords: production network, coordination, scheduling, monitoring and control

1. Introduction

Markets are typically served by competing *production networks* that consist of *autonomous* enterprises. Hence, as competition and cooperation go on hand in hand, the whole operation of the economy becomes more complex. Nodes and structures of networks are changing, thus making company relations more dynamic and less predictable. Today's greatest pressures are *time compression*, *customization* and *cost reduction*: the market increasingly demands products that are customized, yet available with shorter delivery times. While any network as a whole is driven by the overall objectives to meet customer demand at the possible minimal production and logistic costs, the efficiency of operations and the economical use of material, energy and resources hinges on the local decisions of the autonomous partners. Taking high *service level* as their main priority, manufacturers can hedge against uncertainties only by capacity and/or material buffers. This however, incurs extra equipment, labor, inventory and organizational costs, as well as – especially under dynamic market conditions – the risk of producing obsolete inventory. Enterprises are independent entities, with their own resources, performance objectives and internal decision mechanisms. They have to find their specific trade-offs between service level and cost that are acceptable for their partners. Hence, an overall solution can only emerge from the interaction of local and asynchronous decisions (Monostori et al., 2006). The key decisions are related at least to three distinct levels:

- There is an inevitable need to design both the architecture and behaviour of such network organizations that are able to perceive and respond to market demand by sustaining *coordination* and, if possible, *co-operation* among their members.
- Essential *production planning and scheduling* problems must be solved locally, since manufacturers have to exercise control over future events in their own premises. This is a key also to predictable behaviour.
- *Execution* of production schedules should be supported by real-time monitoring and control that is able to adapt execution to changing conditions, with a minimal ramification of changes.

1.1 Scope: customized mass production

¹ Computer and Automation Research Institute, Hungarian Academy of Sciences

² Budapest University of Technology and Economics

2 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

We are interested in tackling the above issues in the specific context of *customized mass production*. In this setting, production is aimed at satisfying volatile demand on markets of *mass products* where demand appears for a complex and ever changing variety of goods, both for small and large quantities, in hardly predictable temporal patterns. The distinction between customized mass production and the more commonly referred mass customized production (Tseng and Piller, 2003) is conscious; namely, we want to emphasize *customization* in *mass* production. Demand must be fulfilled with mass production efficiency, but in very short times: acceptable delivery times are usually only fractions of the production lead times. The products are typically consumer goods like low-tech electronics, mobile phones, light sources, cosmetics, etc. Customer demand is anticipated and satisfied directly by a manufacturer of end-products that works in the *focal point* of the network, while other members supply the manufacturer with necessary components including packaging materials.

1.2 Industrial motivation

The motivation of this work comes from a large-scale national industry-academia R&D project “*Real-time Co-operative Enterprises*” (VITAL) aimed at improving the performance of a production network that produces consumer goods in large quantities and variability (Monostori *et al.*, 2006). We were interested in *planning and controlling the behaviour* of the network, on different aggregation levels and time horizons, but on each horizon in a *real-time* manner. Since the focal manufacturer gives the heartbeat of the network, we put special emphasis on *scheduling and controlling* its operations.

However, the VITAL project was aimed at developing *generic* information technology (IT) solutions for networked enterprises producing customized mass products under continuously changing, uncertain conditions. Hence, beyond providing integrated solutions from the level of production networks through single enterprises down to the production lines, we had also to tackle the following general problems:

- globalisation, increasing competition, frequently changing, uncertain environment (Wiendahl *et al.*, 2002),
- growing complexity of production processes, manufacturing systems and enterprise structures (Wiendahl and Scholtissek, 1994), and
- autonomous, partly competing, partly cooperating production structures (Monostori *et al.*, 2006).

The challenges identified at the beginning of the project and the envisaged answers are summarised in table 1.

Challenges	VITAL's answers
Complexity	Autonomy & cooperation
Changes and disturbances	Intelligence
Finding the possible best technical-economic solutions	Integration & optimisation
Quick reaction ability	Real-time behaviour

Table 1. Main challenges to be answered in VITAL.

The primary user of the results is a factory – considered one of the largest of its kind in the world – situated in Hungary. The factory works with its more than 100 production lines in three shifts, and produces altogether several million units/week from a mix of thousands of products. Fulfilling orders in time is of primary importance. The customers do not necessarily realize that they face a conglomerate of enterprises, i.e., a production network. Consider figure 1 where the importance of *time* is emphasized by the various watches which represent the different levels (network, enterprise, production line) of production. External and internal changes and disturbances (indicated by thunderbolts) have to be responded with a *reaction time* that depends on the level in question. Hence, real-time operation has a different realization level by level. Further on, in order to account for unexpected future changes and to root decisions, as far as possible, in facts known for certain about the execution environment, communication and decision-making have to be performed on each level with a *rolling horizon*.

Insert Figure 1 about here

Summing up, the particular problems to be solved were as follows (referring to the notations of figure 1):

- detailed, short-term finite-capacity production scheduling (B),
- real-time production monitoring and control (C), and
- coordination of component supply (A).

The logic behind the above task sequence – which will be followed also in the subsequent sections of the paper – is the following: detailed behaviour of the factory is determined by production scheduling that takes all the known temporal, resource, material availability and technological constraints into account. Real-time production monitoring and control should ensure the execution of the schedule, while component supply should guarantee the availability of necessary materials and components.

2. Rolling Horizon Production Scheduling

The role of daily scheduling over a rolling horizon is twofold. On the one hand, it has to *schedule new production orders*, and on the other hand, it has to *guarantee feasibility of the next few days*. When scheduling new production orders (POs), first processing alternatives have to be selected from appropriate sets of alternatives, and then the POs have to be scheduled on the machines. It is a common requirement that the machine assignments and the sequence of the already scheduled tasks should be modified only if there is no other way to improve on the cost function. While scheduling algorithms mainly focus on temporal feasibility, in customized mass production where the same material can be built into several production orders, the material stock and expected shipment must cover the demand of all scheduled items in order to ensure a smooth execution. However, sufficient material supply is required in the next few days only, while on the longer term the scheduler computes only the material demand and contrasts it with the known stock levels and expected shipments. This provides information as whether additional material must be called off from the suppliers.

In the sequel we commence by defining the scheduling problems to be solved, and then briefly review the related literature. After explaining our solution methodologies, we provide some details of the implementation and examples of the possible functionalities of our system.

2.1 The Scheduling Problems

When scheduling over a rolling horizon, new production orders are inserted into the actual schedule after a *frozen period*, which consists of the next few days. There is a set of *production orders* PO_1, \dots, PO_n , where PO_i requests the production of a specific end-product or intermediate product in a given quantity q_i . Each PO_i has a *release date* r_i and a *due date* d_i . There is a *time horizon*, two weeks, say, and the r_i falls on the beginning of some day, while the d_i is the end of some day within the horizon. Each PO_i is divided a-priori into a finite number of *jobs*, where job $j \in PO_i$ contains q_j items from PO_i and $\sum\{q_j \mid j \in PO_i\} = q_i$. The production process consists of a sequence of *main production steps* and each PO requires a subsequence of this. Each PO_i has a few *routing alternatives* $R_i^1, \dots, R_i^{a(i)}$, where each R_i^l is a sequence of stages, each *stage* being a subsequence of steps. The stages of each routing alternative must be disjoint and their union must be the set of steps required by PO_i . With each stage s , there is associated a set of machines M_i^s . The *processing time* of job $j \in PO_i$ on some machine M_k is defined as $p_{j,k} = q_j / v_i^k$, where v_i^k is the *yield* of M_k (items/time unit) when processing any job of PO_i . Each machine M_k has a calendar specifying those time periods when the machine is available for processing. There are *sequence dependent setup times* between the jobs of different POs scheduled on the same machine. It is assumed that the setup times satisfy the triangle inequality. Since the new production orders have to be inserted into the *actual schedule*, it is also part of the input.

A solution to an instance of this scheduling problem selects a routing alternative for each new PO, assigns a machine to each stage of each job and specifies an order of job-stages (called *operations*) on each machine. The quality of the solution can be measured by e.g. total PO tardiness, that is, the sum of the tardiness of each PO, where the tardiness of PO_i is the time passed after d_i until the completion of its last job (this can be 0 when all jobs of PO_i get finished not later than d_i). The solution incorporates the input schedule possibly with some modifications: it is permitted to change the order of already scheduled operations on some machines or even to assign some operations to new machines proved this leads to improvement of the objective function. This is how scheduling over a rolling horizon progresses: the schedule obtained today becomes part of the input tomorrow, etc.

In order to ensure that the schedule of the next few days can be executed, a different problem has to be solved. Namely, given a solution to the above problem, and in addition material requirements for each stage of each job along with initial stock levels and expected shipment days and quantities for the different materials, a

4 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

subset of operations has to be removed from the schedule and inserted back after this period, in order to ensure that all scheduled jobs on the next t days have sufficient materials. To fill in the gaps, other jobs from the future have to be reinserted within the next t days while maintaining temporal and material feasibility in this period.

2.2 Review of Related Literature

We are not aware of any papers that present a solution approach for a problem with all those features like ours. Nevertheless, there are several approaches for job shop scheduling with the *mean tardiness objective*. The recent approaches can be categorized as local search algorithms, genetic algorithms, and ad-hoc methods. A *local search* algorithm maintains a candidate solution which is modified in small steps with the aim of improving it over several iterations. In (Kreipl, 2000) a simulated annealing procedure is presented, while in (De Bontridder, 2005) a new neighbourhood embedded in tabu search is described for job shop scheduling with the weighted tardiness objective. A genetic algorithm for the same problem is evaluated in (Mattfeld and Bierwirth, 2004). An interesting approach is taken in (Raman and Talbot, 1992), where the machines are processed one-by-one in several rounds until no improvements are found. When (re)scheduling a machine, for each job on the machine, task due dates are set in a systematic way and each setting is evaluated by a dispatching rule. This approach is also appropriate for scheduling on a rolling horizon. However, reassigning jobs to new machines, machine calendars and setup times are not considered in this approach either.

2.3 Scheduling of New Production Orders

This scheduling problem is solved in three phases. Firstly, for each new PO a routing alternative is selected and a fractional assignment of the work represented by the POs to the machines is chosen by solving a relaxation of the problem. In the relaxed problem we neglect the jobs of the production orders. Instead, we allocate the production capacity of the machines to the POs. Namely, for each PO_i we introduce a set of binary variables to select exactly one routing alternative, another set of binary variables to measure tardiness, and several continuous variables to represent the amount of working time (in minutes) allocated to the PO on each day and on every machine that can perform some of its stages. The constraints of the relaxation ensure that (i) precisely one routing alternative is selected, (ii) only those machines are allocated to a PO that perform the stages of the selected routing alternative, and (iii) subsequent stages feed one-another over the days (see figure 2).

Insert Figure 2 about here

We solve this mixed-integer linear program (MILP) by means of a standard solver. The solver is based on branch-and-bound in which the lower bound in every node is obtained by solving the linear relaxation of the MILP. Therefore, even if we do not solve the problem to optimality (this can be too time consuming in some cases), we can stop any time once a feasible solution is found, and as a by product of this method, we obtain a lower bound of the optimal solution which is useful for evaluating the quality of the final solution. In the second phase, we build an *initial schedule* where the new jobs are assigned to the machines proportionally to the amount of work allocated to the machines by the feasible solution of the relaxed problem. Finally, in the third phase, the initial schedule is improved by tabu search (see e.g., Glover 1989), the objective being to minimize the total PO tardiness as a primary objective, and to minimize the total job tardiness as a secondary objective.

Insert Figure 3 about here

In fact, without the secondary objective it would be hard to decrease total PO tardiness, because the search process would have no clue how to improve on the schedule. To improve the schedule, we have extended the notion of critical operations to take the effect of machine non-availability periods into account (see figure 3). In each iteration of the tabu search algorithm, we find a critical forest of the critical jobs, and select a critical tree in this forest: the algorithm tries to improve the schedule by systematically testing the critical jobs in all feasible position on all eligible machines. At this stage it is permitted to modify the actual schedule, i.e., change the machine assignment or the order of the jobs which were already scheduled, provided it leads to an improvement of the objective function.

Rescheduling to Ensure Material and Temporal Feasibility

The rescheduling of jobs of the next t days is performed in two steps: firstly, a subset of tasks is selected for displacement into the future such that the remaining tasks have sufficient material supply. This subset is chosen

by solving one multidimensional knapsack problem using the well-known MIP formulation for each of the t days. The second step consists of moving forward some jobs to fill in gaps. This task is performed by a simple heuristic procedure, which, nevertheless, ensures that no material shortage is created.

2.4 Implementation, Testing and Deployment

We have designed a scheduling system with modular structure and implemented the various functionalities by combining these modules. The main modules are the “Assigner”, “Temporal Scheduler”, and “Material Scheduler”. Roughly speaking, the Assigner selects a *processing alternative* for each production order, the Temporal Scheduler builds an *initial schedule* and then *improves* it by local search, while the Material Scheduler *reschedules* the next few days to ensure temporal and material feasibility. All modules have been implemented in the programming language C++ and we also use a commercial mathematical programming package for solving the mathematical problems.

As for possible uses of these modules, we mention (1) Scheduling of new POs: first apply the Assigner, then the Temporal Scheduler, (2) Update the schedule with finished jobs: apply the Temporal Scheduler, (3) Ensure that there is no material shortage on the next two days: invoke the Material Scheduler.

We have thoroughly tested the scheduling of new POs both on computer generated and real-world problem instances. In each case we also computed a lower bound (cf. Section 2.3). We found that the final solution of the temporal scheduler is 1.2 to 1.5 times off from the lower bound on the total PO tardiness on the hardest instances (in those cases when the lower bound was non-zero). The test instances consisted of about 1,000 POs and 5,000 jobs to be scheduled on more than 100 machines. Each PO had 2 routing alternative on average. Nevertheless, a time-feasible solution of the above quality could be obtained in about 15 minutes on a PC with 3 GHz CPU.

The scheduling system has been deployed at the focal manufacturer and is applied for solving its short term scheduling problems. The new production orders are scheduled after a frozen period.

The managements reported that the performance of the system in terms of solution quality meets the users' requirements. The system also makes possible to shorten the cycle time which was not possible previously.

3. Real-time Control of Daily Production

Real-time control of the daily production is an important prerequisite of customer responsiveness. Its main function is to adapt the scheduled operations to the changing environment, while preserving efficiency with respect to cost, time and quality requirements.

Problems of such kind are extensively investigated world-wide. A real-time schedule monitoring and filtering approach based on statistical throughput control is described in (Wilhelm, et al., 2000), for recognizing and evaluating the impact of disturbances. The schedule repair algorithm is activated only in case of severe disturbances in order to decrease system nervousness. A deadlock-free rescheduling algorithm is introduced in (El-Maraghy, ElMekkawy, 2002). Intelligent techniques for recognizing changes and disturbances and to adapt the production to current internal and external circumstances are enumerated in (Monostori, 2003).

In the solution described here, the reference of real-time production control is the optimized, daily schedule described in the previous section. The information about the overall factory is collected in the MES (Manufacturing Execution System) Cockpit (figure 4), which has a database common with the Production Monitoring system and the Scheduler. This main database is instantly synchronized in real-time according to changes on the shop-floor and the real-time synchronization mechanism is also responsible for the update process in the ERP system of the factory.

Insert Figure 4 about here

By default, the MES Cockpit itself provides an overall view of the factory; however, the status of separate plants, cells or specific machines can also be checked. The platform also notifies the users about deviations from the production schedules together with the option to find the cause of the deviation (e.g. raw material unavailability, machine breakdown, lack of operator, etc. see figure 4 left side.). Parallel to the useful functionalities of the MES Cockpit that enables the monitoring of the production, in our system a new Factory simulator was developed enabling the analysis, testing and validation of the previously calculated schedules (figure 3 right hand side.). The following section details the structure and the functionalities of the Factory simulator module and presents the first results achieved with the system.

6 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

3.1 Simulation Support for Production Control

The main functions (or operation modes) of the simulator in the proposed architecture (depicted in figure 5) are as follows (Monostori et al., 2007):

- Off-line validation, sensitivity analysis of the schedules. Evaluation of the robustness of daily schedules *prior to the execution* against uncertainties, such as machine unavailability or job slipping. In this way, it can point out those resources which can endanger the realization of the daily schedule.
- On-line, anticipatory recognition of deviations from the planned schedule by running the simulation parallel to the plant activities. By using a look ahead function (supposing of keeping the sequences as planned), support of situation recognition (*proactive operation mode*, figure 5).
- On-line analysis of the possible actions and minimization of the losses after a disturbance already occurred (*reactive operation mode*, figure 5).

Insert Figure 5 about here

The model structure in the simulator is the same for the three operation modes, however, the granulation (level of modelling detail), time horizon, applied failure models and considered outputs depend on the purpose of the assignments.

In the on-line modes the simulation models represent various virtual mirrors of the plants and run parallel to the real manufacturing environment, simulating also the future processes for a predefined short period. The performances of the predicted and the so far executed schedule are compared (highlighted as 'Performance measure of interest' in figure 5).

The off-line operation mode refers to either the factory or individual plants, while in the on-line modes the work of a plant-level Decision-maker is supported (figure 5). The main goal of the Decision-maker is to ensure the completion of the jobs assigned by the scheduler to the given plant, and if it is not possible, to minimize the tardiness of jobs. In case of intervention, a rescheduling action has to be performed with limited scope (in space and time) in correspondence to the sphere of authority of the Decision-maker. The control action made in this rescheduling point incorporates the selection of the appropriate rescheduling policy and method.

3.2 Main steps for building the simulation model

As highlighted in figure 4, the Factory simulator and the Scheduler have the same production database. Resources, products, process plans, resource calendars, production information, i.e., directly and indirectly usable data are transformed exactly to the same form for all system components. The on-line database connection has been developed by using the built-in database interface component of discrete event simulation software, and standard SQL queries are applied to download the necessary data into the simulation module. Note that simulation relevant data are stored locally in the simulation model.

Data-processing is carried out before the overall simulation (figure 6, phase a) and the redundant data storage in the simulation model is compensated by the advantage of the shorter response time. Modelling real production systems frequently brings up the problem of handling hundreds of resources in a simulation model. Having the modelling objects in hand, which were created on the base of the conceptual model, in our architecture the simulation model is created automatically based on the pre-processed data (phase b) in figure 6). Note that the process phases a) and b), can be considered the key elements for the successful application of simulation. Hereby, the complexity of integrating the simulation module into the system is significantly reduced. None-the-less, the common data tables ensure data integrity during the creation of the simulation; moreover, the data-model serves as a basis for the more detailed shop-floor model.

Insert Figure 6 about here

The automatic generation of the model is followed by the initialization phase (phase c) in figure 6). In this phase, besides classical parameter settings, the procedure involves the generation of input-parameter-specific model components (entities such as products, operators). Contrary to the previous phase, this one is carried out for each replication.

The simulation runs are repeated until the required number of replications is obtained (phase d) in figure 6). Each replication is a terminating, non-transient simulation run, having the same initial parameters and settings, but different parameters for uncertain simulation times and events generated on the base of random numbers.

In the last phase the schedule is evaluated by using the evaluation criteria and the results of the evaluation process are interpreted by shop-floor managers (indicated as Decision-maker in figure 6) who are responsible for taking necessary actions (phase e) in figure 6).

3.3 Experimental results

The main focus of the prospective simulation experiments was the analysis of the human resource (HR) availability on schedule execution in uncertain environment. Each day, simulations are initiated in proactive operation mode for a one-week (7 days, 3 shifts) time period by the Decision-maker. Disturbances generated during the model runs, influencing the schedule execution are generally:

- Machine breakdowns,
- Stochastic processing times,
- Sequence changes.

However, material shortage of certain raw materials, which is one of the most recent causes of delays, is not considered directly as an input factor for the simulation. But the production orders, which cannot be processed due to this reason, are reallocated in the schedule, i.e., usually a sequence change is made to resolve the problem. This is important, while the MRP database is not accessible for the simulation module, but the Scheduler “transforms” this information into the production database.

Results of a single proactive simulation trial are presented in figure 7. Tests have been performed for the two different human resource types in consideration (highlighted in blue (Res1) and in red (Res2)), representing the two main service groups required for the production lines, namely, operators and packaging personnel.

Dashed lines illustrate the calculated maximum human resource (HR) demand, i.e., threshold values changing in time, while continuous lines represent the simulated (anticipated) HR demand for the one-week period. Here, the classification of deviations means that, e.g., in the current experiment the simulated HR demand exceeds the calculated capacity (highlighted with the ellipse in figure 7). Similarly, at the end of the time period (approx. from the middle of day 4) further analysis is advisable, considering the very high utilization for a longer period.

As it was previously stated, in case the proactive simulator warns the Decision-maker on a deviation, a reactive simulation can be initiated, and thus, the possible actions and minimization of the losses after a disturbance already occurred can be analyzed.

Insert Figure 7 about here

4. Coordinating Production and Supply Logistics

The main problem of coordinating supply in a production network is how to achieve and maintain the right overall behaviour of the whole network if the autonomous business partners decide locally, based on asymmetric and partially incomplete and inconsistent information. This problem can be stated as follows: the common goals of each partner are (1) to provide *high service level* towards its customer, while, at the same time, (2) to keep the overall expected production and logistics *costs at a minimum*. Note that these requirements are conflicting:

- Due to uncertain market conditions, inventories (of components, packaging materials, products) are inevitable to provide service at the required level.
- In mass production technology, low costs can be achieved only with large lot sizes, which involve, again, higher product and component inventories as well as increased work-in-process.
- Markets of customized mass products are volatile. If the demand unexpectedly ceases for a product, then accumulated inventories become obsolete and cause significant losses.

Further challenge is that all network partners are autonomous, and those in the supplier's role take typically part in other network relations as well.

We have taken a three-phased approach to tackle the supply coordination problem:

- Establishment of a media for sharing information about the actual and expected situations, as well as of the future intentions (i.e., plans, schedules) of autonomous network partners.
- Elaboration of novel, cooperative inventory planning methods.
- Design and set up of incentive mechanisms that facilitate truthful information exchange, the sharing of risks and benefits, and cooperative behaviour.

8 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

4.1 Logistics Platform for Information Exchange

Though there exist a number of enterprise resource planning (ERP) and supply chain management (SCM) systems that provide technology for information storing, retrieval and sharing within and between the nodes of a production network, these systems are mainly transactional: they do not really support coordinated decision making (Stadler 2005; Li and Wang, 2007). Hence, we developed a so-called *logistics platform* (LP) for coordinating the partners' decisions along individual supply channels (Egri and Váncza, 2006). The key design idea is to detach the two main conflicting objectives and thus make both of them manageable: while service level is tackled on the short-term, where detailed production schedules provide reliable information about the close future, cost-efficient production is concerned on the medium-term. Hence, the platform consists of two levels (see also figure 8):

1. On the *planning level*, the supplier receives medium-term demand forecasts of components from the manufacturer, together with some information about the reliability of forecasts, decides on production quantities and manages the inventory. On this level, decisions are made in a weekly cycle.
2. On the *scheduling level*, the supplier meets the exact, short-term component demand of the manufacturer. This demand is generated from the actual daily production schedule of the manufacturer in form of call-offs and has to be satisfied by direct, just-in-time delivery from an inventory. Decisions are made on a daily basis, with a horizon of 1 to 2 weeks. With this short look-ahead, demand uncertainty is hedged by safety stocks.

Insert Figure 8 about here

The system provides an interface between planning functions of the manufacturer and the supplier. Decisions have to be made locally, but various checks (whether supply meets demand, inventory is sufficient) are performed within the platform. In particular, the LP is organized around the concept of *supply channel*. Each channel is defined by a customer, a supplier and a particular component that is delivered through the channel. For each channel, there is a complex inventory composed of the in-transit, consignment, as well as of the on-hand inventories at the supplier and the customer. The LP keeps track of the inventory items on a daily basis. However, as figure 8 shows, inventory is a passive element whose level is influenced by the local decisions of the supplier (who builds up the inventory) and the customer (who consumes the inventory). The LP collects, presents, analyses and aggregates all relevant information concerning the future, present and past of the channels. Hence, each channel has various dynamic future related information, like forecasted demand, open orders, scheduled demand (generated by the customer), production plan and delivery schedule (generated by the supplier). Departing from the actual inventories, projected inventories are calculated both in the medium and short term, and inventory statuses are evaluated from both partners' point of view.

4.2 Cooperative Inventory Planning

For supporting *inventory planning* that links the planning and scheduling functions of autonomous partners along a supply channel, we developed a portfolio of novel lot sizing methods that take the total production and logistics costs into account, regarding also the uncertainty of demand. Namely, component demand may stop whenever the market demand for the end-product(s) built of the component ceases for any reason. In this situation called *run-out* the component inventory becomes obsolete.

Decisions that *coordinate a channel* can be made on the basis of information coming partly from the manufacturer (component demand forecast and its uncertainty) and partly from the supplier (setup, production and inventory holding costs). Two different situations have been modelled:

1. Run-out of a product can occur with a certain probability any time in the future, but no further details are known
2. At the *end of a product's life-cycle*, the fact of the run-out and its date are known, but the amount of demand is still uncertain.

For the case of the unknown run-out date, we introduced *heuristic policies* and an extended version of the classical *Wagner-Whitin* (WW) method (Hopp and Spearman, 1996). These methods depart from a medium-term multi-period demand forecast and the probability of run-out, as well as from the cost factors of production, setup and inventory holding and determine the *optimal lot size*. As a novelty, the expected cost of obsolete inventory is also considered. The heuristic methods disregard the less trusted remote forecasts and minimize the expected average costs (Váncza and Egri, 2006), while the modified WW (WWr for Wagner-Whitin with run-out) plans the whole horizon and determines the setups (Egri and Váncza, 2007). The WWr plans the whole horizon by minimizing the expected total cost; hence it can also approximate the number of setups.

Figure 9 shows a particular, comparative example of the results of the three different multi-period methods, together with the so-called economic order quantity (EOQ) applied in the everyday practice. As for the WWR method, the number of setups is also shown. In the figure one can follow how the optimal lot size depends on the probability of run-out. Under stable conditions, the actual practice based on EOQ is too risk-averse, while if the run-out of the product has some however minor chance it would be better to produce less quantities.

Insert Figure 9 about here

For handling the case of the end of a product's life-cycle (which is fairly frequent with customized products) we have extended the so-called *newsvendor* method (Cachon, 2003). The standard one-periodic model describes the situation when stocks cannot be carried from one period to another. In our case, demand is given as a random variable with a particular distribution. Since service level is of primary importance, all market demand must be satisfied by the manufacturer. This may necessitate an additional ordering from the supplier and paying another setup. All in all, the expected total cost consists of the expected cost of setup, purchase, obsolete surplus and emergency setup costs. The lot sizing decision is the responsibility of the supplier. As we have shown elsewhere, under specific conditions the problem has a unique solution (Egri and Váncza, 2006).

Insert Figure 10 about here

All the methods have been extensively tested on historic industrial data and figure 10 shows a summary of the methods and a decision scheme for applying them in various situations. We can but note here, together with (Chikán, 2007), that the traditional inventory planning methods with some appropriate extensions got new significance in the context of managing complete supply networks.

The proposed methods are based on information sharing, but assume *truthfulness* of the partners: channel coordination hinges on whether the manufacturer reveals its real demand forecasts and/or product related uncertainties. However, the manufacturer has an incentive to inflate its forecast or to underestimate its uncertainty. In any case, the effect will be more optimistic: it yields larger lot sizes and inventories. The manufacturer can be on the safe side, but the network will operate with higher inventories than necessary and the chance of producing obsolete inventories will also increase. Hence, selfish (rational) *distortion of information* will necessarily lead to additional costs. Currently, we have designed such *mechanisms* that drive the partners towards disclosing and using unbiased information when trying to coordinate the channel. Accordingly, the supplier provides a *service* to the manufacturer by committing itself to meet all of its short-term demand. In return, the manufacturer pays for (1) the components delivered, (2) the flexibility of the supplier, and (3) the forecast deviation. We have defined both for the multi- and the single-period inventory planning methods such a *protocol* and *payment scheme* that drive the manufacturer to communicate its real forecast, and the supplier to make network-wide optimal decisions on the quantities to be produced.

4.3 Implementation and deployment

The logistics platform has been deployed at the manufacturer and integrated with its main planning functions, as shown in figure 11. It follows the focal structure of the supply network where it is being applied. The system is a Java Enterprise Edition (EE) web application built on the manufacturer's proprietary web application framework. This framework manages the database connection pool, the request dispatching and corporate Single Sign On (SSO) authentication. The application can be accessed from the manufacturer's intranet as well as from the external suppliers through the Virtual Private Network (VPN) of the manufacturer.

Insert Figure 11 about here

While the system reads only planning level data from the production planner system of the manufacturer, it is synchronized with the scheduler (see Section 3). In fact, the suppliers are responsible for avoiding material shortage on the manufacturer's side by feeding the factory with components, having respect to the short term scheduled demand. Primary criterion is that scheduled tasks should not remain without their required components. Hence, the suppliers read the scheduled demand for components from the actual production schedule. In return, they provide their *delivery schedules*. This information is fed back to the production scheduler system that regards delivery schedules firm promises when checking the material feasibility. In this setting monitoring the performance of suppliers is essential. We follow the principle of forward coverage: the manufacturer's on-hand inventory together with the suppliers' scheduled delivery must always cover demand of the next few days.

10 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

Any case where this coverage does not hold, calls for immediate actions; either from the supplier (urgent delivery) or from the focal manufacturer (re-scheduling), or from both.

The logistics platform is now being used by the planners of the manufacturer and of the suppliers. To start with, channels for thousands of components (including packaging materials) have been set up. The main benefit of using the application is that the supply and demand planners can see, compare and analyze dynamic information coming from different and heterogeneous sources. This way, they can anticipate future conflict situations. Our final goal is to make efficient local planning and scheduling, as well as reliable real-time execution each partner's primary interest. This is the key for inspiring cooperation in a production network whose global behaviour emerges from the decisions of its autonomous members.

5. Conclusions

One of the most "vital" features of enterprises is their ability to cooperate as well as to give quick responses to changes and disturbances. In this paper, we presented cooperative supply planning, as well as production scheduling and execution monitoring methods that were developed to improve the overall logistic and production performance of a supply network that has to operate under uncertain market and technical conditions. Since the methods come from various, novel areas of informatics, operational research and knowledge-based systems, their integration is expected to balance the aspects of optimization, autonomy, and cooperation.

A special challenge of the VITAL project was that the R&D activities, the necessary reconstruction of the IT infrastructure and the business process reengineering had to be accomplished hand-in-hand, without endangering the continuity of the production. Therefore, special emphasis was laid on digital enterprise technologies, supporting the smooth introduction of the solutions developed. The systems described in the paper have been successfully deployed in the industry and are in daily use.

The scheduling system is being further developed into a scalable, modular system. The effectiveness of the real-time production control will be further improved by the introduction of active identifiers. An agent-based production control structure is also envisaged. The functionalities of the logistic platform is being enhanced (Váncza et al., 2008) and expected to be introduced in a global supply network with more than 100 nodes in the second half of 2008. Some of the solutions will be offered for small- and medium-sized enterprises as web services.

Acknowledgements

The work was supported by the NKFP Grant No. 2/010/2004 (VITAL), the OTKA Grants No. T049481, T73376, and by the EU-projects AC/DC IST 031520 and Collplexity 12781(NEST). Tamás Kis gratefully acknowledges the support of János Bolyai research grant, no. BO/00380/05.

6. References

- Cachon, G.P., Supply chain coordination with contracts. in *Supply chain management: Design, coordination, cooperation*, Edited by A.G. de Kok, S.C. Graves, pp. 229-339, 2003 (Elsevier)
- Chikán, A., The new role of inventories in business: Real world changes and research consequences, *International Journal of Production Economics*, 2007, **108**, 54-62.
- De Bontridder, K.M.J., Minimizing total weighted tardiness in a generalized job shop, *Journal of Scheduling*, 2005, **8**, 479–496.
- Drótos, M., Erdős, G. and Kis, T., Computing lower and upper bounds for a large-scale industrial job-shop scheduling problem, *European Journal of Operational Research*, under review
- Egri, P. and Váncza, J., A logistics framework for coordinating supply chains on unstable markets. in *Digital Enterprises Technology – Perspectives and Future Challenges*, Edited by Cunha P.F., Maropoulos, P.G., pp. 59-66, 2007, (Springer)
- Egri, P. and Váncza, J., Incentives for cooperative planning in focal supply networks, in *6th International Workshop on Emergent Synthesis, IWES'06*, Kashiwa, Japan, 2006, 17-24
- ElMaraghy, H.A. and ElMekawy, T.Y., Deadlock-free rescheduling in flexible manufacturing systems, *CIRP Annals – Manufacturing Technology*, 2002, **51**(1), 371-374.
- Glover, F., Tabu search - Part I. *ORSA Journal of Computing*, 1989, **1**, 190-206
- Hopp, W.J. and Spearman, M.L., *Factory Physics – Foundations of Manufacturing Management*. 1996 (Irwin)
- Kreipl, S., A large step random walk for minimizing total weighted tardiness in a job shop, *Journal of Scheduling*, 2000, **3**, 125–138.
- Li, X. and Wang, Q., Coordination mechanisms of supply chain systems. *European Journal of Operational Research*, 2007, **179**, 1-16.

- Mattfeld, D.C. and Bierwirth, C., An efficient genetic algorithm for job shop scheduling with tardiness objectives, *European Journal of Operational Research*, 2004, **155**, 616–630.
- Monostori, L., AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. *Engineering Applications of Artificial Intelligence*, 2003, **16**(4), 277-291.
- Monostori, L., Kádár, B., Pfeiffer, A. and Karnok, D., Solution approaches to real-time control of customized mass production, *CIRP Annals – Manufacturing Technology*, 2007, **56**(1), 431-434.
- Monostori, L., Váncza, J. and Kumara, S.R.T., Agent-based systems for manufacturing, *CIRP Annals – Manufacturing Technology*, 2006, **55**(2), 697-720.
- Monostori, L., Váncza, J., Kis, T., Kádár, B. and Viharos, Zs.J., Real-time, cooperative enterprises: requirements and solution approaches, in *Information Control Problems in Manufacturing*, Vol. I, 561-566, 2006 (Elsevier)
- Raman, N. and Talbot, F.B., The job shop tardiness problem: decomposition approach, *European Journal of Operational Research*, 1993, **69**, 187-199.
- Stadtler, H., Supply chain management and advanced planning – Basics, overview and challenges. *European Journal of Operational Research*, 2005, **163**, 575–588.
- Tseng, M.M., Lei, M. and Su, C., A collaborative control system for mass customization manufacturing. *CIRP Annals – Manufacturing Technology*, 1997, **46**(1), 373-376.
- Tseng, M.M. and Piller, F.T., (eds.), *The Customer Centric Enterprise*, 2003 (Springer).
- Váncza, J. and Egri, P., Coordinating supply networks in customized mass production – A contract-based approach. *CIRP Annals – Manufacturing Technology*, 2006, **55**(1), 489-492.
- Váncza, J., Egri, P. and Monostori, L., A coordination mechanism for rolling horizon planning in supply networks. *CIRP Annals – Manufacturing Technology*, 2008, **57**(1) in print
- Wagner, H.M. and Whitin, T.M., Dynamic Version of the Economic Lot Size Model. *Management Science*, 1958, **5**, 89-96.
- Wiendahl, H.-P. and Scholtissek, P., Management and control of complexity in manufacturing. *CIRP Annals – Manufacturing Technology*, 1994, **43**(2), 533-540.
- Wiendahl, H.-P. and Lutz, S., Production in Networks. *CIRP Annals – Manufacturing Technology*, 2002, **51**(2), 1-14.
- Wilhelm R., Chandrashekar, R., Sun, R., Hegedus, M., Chu, B., Tolone, W.J. and Long, J., Finite schedule monitoring and filtering in a computer integrated manufacturing environment. *CIRP Annals – Manufacturing Technology*, 2000, **49**(1), 335-338.

12 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

BIOGRAPHY OF THE AUTHORS



Prof. László Monostori has been with the Computer and Automation Institute of the Hungarian Academy of Sciences (SZTAKI) since 1977, now he serves as Deputy Director Research and Head of the Engineering and Management Intelligence Laboratory. He is also full time professor at the Department of Manufacturing Science and Technology, Budapest University of Technology and Economics. He is a Fellow of the International Academy for Production Engineering (CIRP); Full Member of the European Academy of Industrial Management (AIM); Chairman of the Coordinating (CC) on Manufacturing and Logistic Systems, International Federation of Automatic Control (IFAC), and Chairman of the TC on Technical Diagnostics, International Measurement Confederation (IMEKO). He is Chairman of the Editorial Committee of CIRP Annals; Editor-in-Chief of the CIRP Journal of Manufacturing Science and Technology; Associate Editor of Computers in Industry, and the CIRP Journal of Manufacturing Systems, and member of the editorial boards of other international scientific periodicals. For his research achievements published in more than 330 publications resulted in about 1300 independent citations and for his development activities – among others – the Dennis Gabor Prize was given to him in 2004.



Dr. Tamás Kis has been with the Computer and Automation Research Institute since 1994. Now he is a Senior Research Associate at the Engineering and Management Intelligence Laboratory. He stayed at the Swiss Federale Institute of Technology at Lausanne, Switzerland, for five years, where he received his PhD degree in Mathematics. Since then, he has worked on various research and R&D projects, mostly related to production planning and scheduling problems. His main research interest is scheduling theory and applications and integer programming. He has over 40 publications with more than a 100 independent citations.



Dr. József Váncza has been affiliated with the Computer and Automation Research Institute of the Hungarian Academy of Sciences for more than twenty years. Now he holds the position of Senior Research Associate and Deputy Chair of the Laboratory of Engineering and Management Intelligence. His current interest is in computer-aided process planning, production planning and scheduling as well as in coordinating supply networks. Dr. Váncza has been lecturing in the above topics for a decade or so at the Budapest University of Technology and Economics where he was appointed as associate professor at the Faculty of Mechanical Engineering in 2003. He is associate member of CIRP.



Dr. Botond Kádár is a senior researcher at the Laboratory on Engineering and Management Intelligence of the Computer and Automation Institute of the Hungarian Academy of Sciences (SZTAKI). He obtained his M.Eng. and Ph.D. degrees at the Budapest University of Technology and Economics, Hungary in 1993 and 2002, respectively. His current interest includes production control, simulation and multi-agent approaches for production engineering and manufacturing systems and he is involved in several research and development projects from these fields. Dr. Botond Kádár is author or co-author of 70 publications with over 120 citations.



Dr. Gábor Erdős is a Senior Research Associate at the Computer and Automation Research Institute. He received his MS and PhD degrees in mechanical engineering from Budapest University of Technology and Economics. He also holds an MS degree in mechanical and aerospace engineering from SUNY Buffalo. Dr Gábor Erdős was a post-doc assistant for five years at Institute of Production and Robotics at EPF Lausanne. His main research interests includes computer aided mechanism modeling and digital manufacturing.

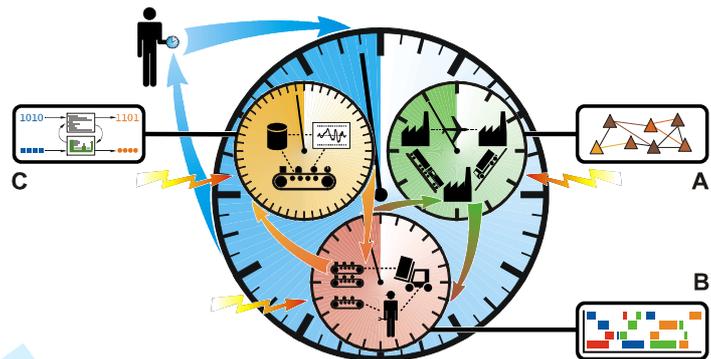


Figure 1. General concept of the VITAL project.

2 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

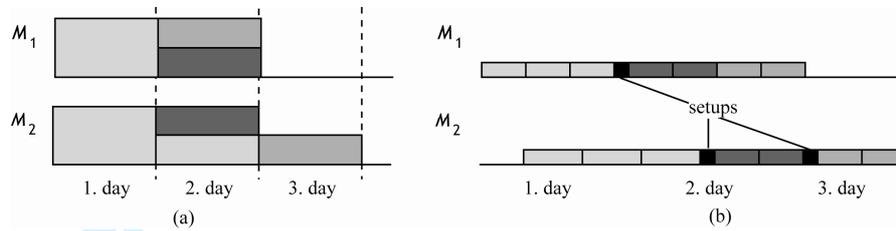


Figure 2. (a) Relaxed problem (only POs); (b) expansion to feasible schedule with jobs.

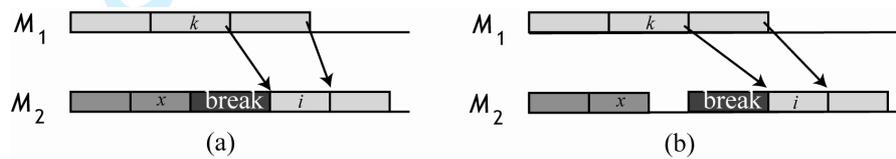


Figure 3. Suppose k and i are successive operations of the same job. (a) the critical predecessor of i is x , (b) the critical predecessor of i is k , since there is idle time between x and the break before i .

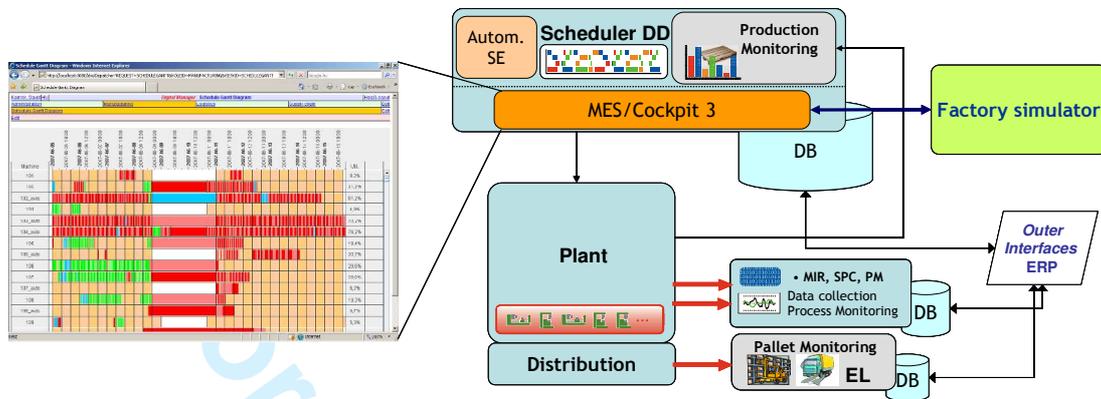


Figure 4. Structure of the realized system and the graphical user interface (GUI) of the MES Cockpit.

4 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

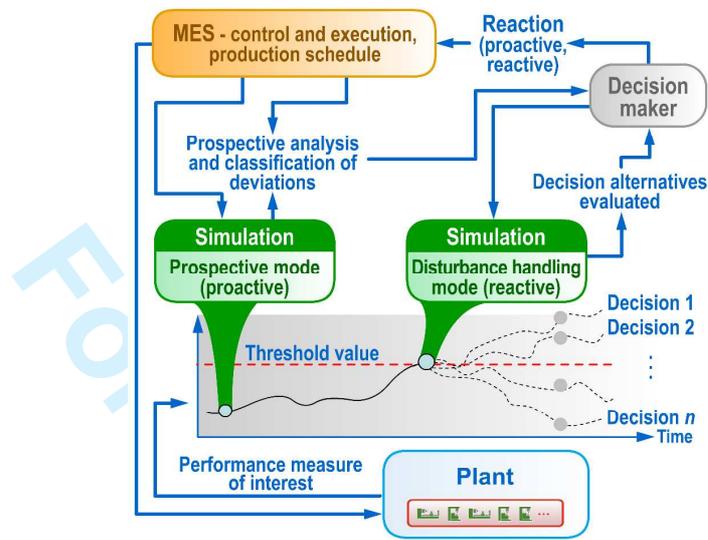


Figure 5. Plant-level active disturbance handling by using reactive/proactive operation modes of simulation.

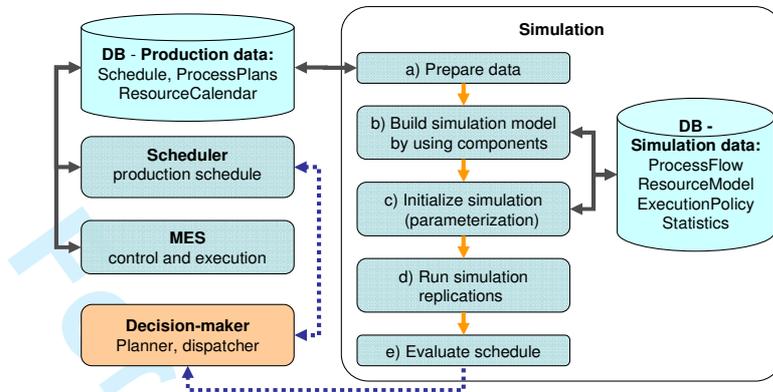


Figure 6. Architecture and the main process flow in the simulation module.

6 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

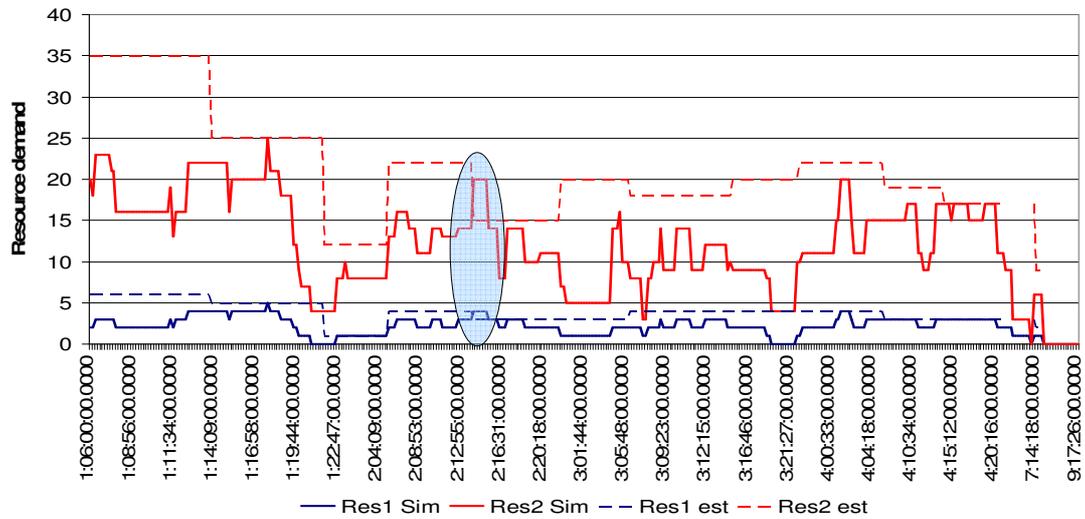


Figure 7. Simulation results, obtained for the packaging section of the plant (1-week period). Considered disturbances are simulated machine breakdowns (day 2, 3x2hours) and stochastic processing times.

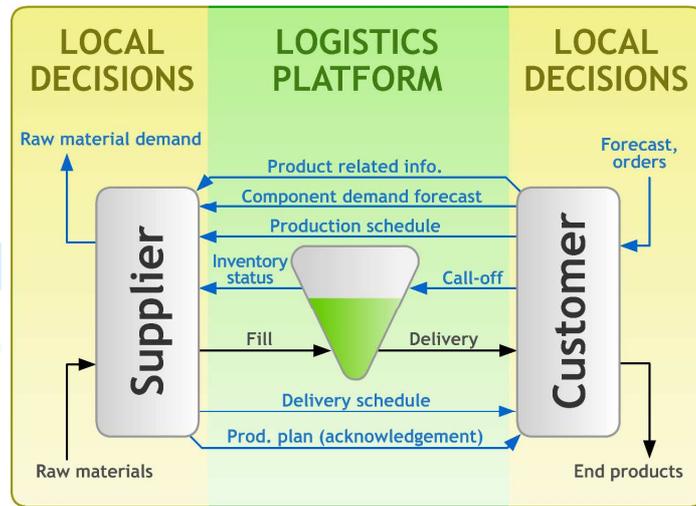


Figure 8. Information flow through the logistics platform.

8 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

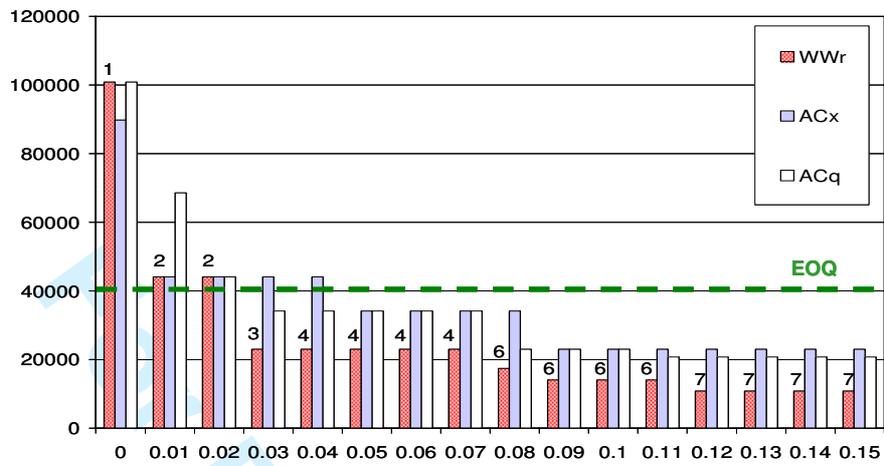


Figure 9. Suggestions for the first lot size of various multi-period inventory handling methods as a function of run-out probability.

Peer Review Only

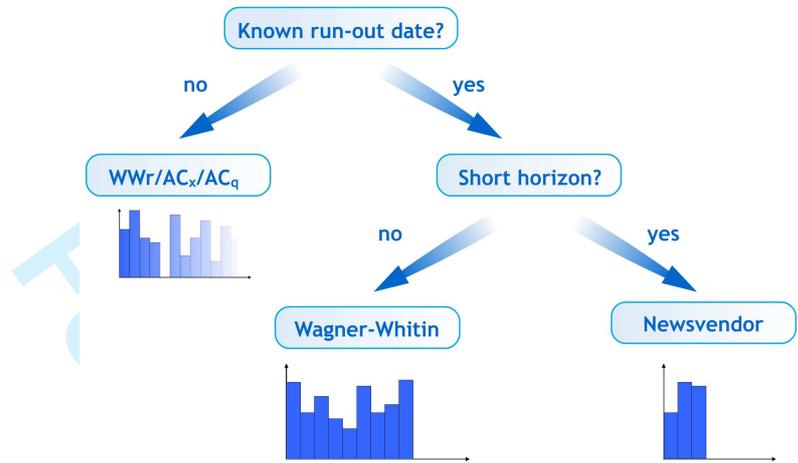


Figure 10. Decision scheme for applying the inventory planning methods.

10 L. Monostori, T. Kis, J. Váncza, B. Kádár, G. Erdős

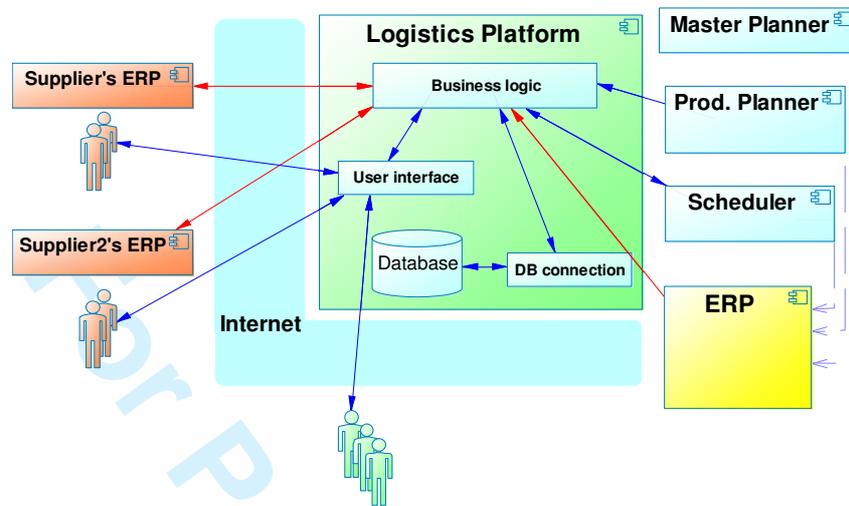


Figure 11. The logistics platform and its place in the IT architecture of the focal manufacturer.