A Coordination Mechanism for Rolling Horizon Planning in Supply Networks

J. Váncza[1,2] (2), P. Egri[1], L. Monostori[1,2] (1)

[1]Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary

[2] Dept. Manuf. Science and Technology, Budapest University of Technology and Economics, Hungary

Abstract

We model planning in a supply network as a distributed effort for matching future demand with supply on a rolling horizon, by relying on asymmetric and in part uncertain information. For achieving high service level and low overall costs throughout the network there is a need of managing the intentions and interactions of the partners. We present a novel coordination mechanism where sharing information truthfully and planning local production optimally serve both system-wide and individual objectives. The work is nested in practice: application examples are taken from the production of customized mass products.

Keywords:

Production, Co-operative, Planning

1 INTRODUCTION

Planning production in a supply network that consists of autonomous enterprises is a distributed and recurring effort to match future demand with supply by relying on asymmetric and in part uncertain information. While the network as a whole is driven by the overall objectives to meet market demand at the possible minimal production and logistic cost, the efficiency of operations and the economical use of material, energy and human resources hinges on the local decisions of the individual partners. Clearly, they can never be completely aware of each other's goals and intended courses of actions. Information asymmetry and local autonomy lead together time and again to inefficiencies like acute shortage situations or excess inventories. We deem this phenomenon known for a long time as **double marginalization** a symptom of the prisoners' dilemma in supply networks.

This dilemma can be resolved by information sharing and **cooperative planning** that aligns individual interests with system wide objectives [1]. In fact, manufacturing system theory suggested, from its formation days until recently, network organizations where production is carried out by partners who are able to coordinate their actions and cooperate with each other [2,3]. With the maturation of agent technology, we also have the proper modeling, system building and simulation tools at hand [4]. Surprisingly, when making a broad review, we found quite a few deployed applications of agent technologies in supply network management. Such systems exhibit complex adaptive behavior that emerges from the relatively simple local relationships of the partners [5]. However, industry needs both guarantees for the emergence of some useful properties (like high service levels) and safeguards against unwanted behavior. Our research goal was to develop a planning mechanism where sharing information truthfully and planning local production optimally serve both system-wide and individual interests.

This work has also been motivated and nested in **practice**. In particular, it was aimed at improving the performance of a production network that makes consumer goods for markets of low tech electronics [6,7]. The network is woven around a focal manufacturer by autonomous suppliers of various components. Since the weekly production rate amounts to several million units, mass production technology is applied to exploit economies of scale. At the same time, customization and responsiveness are competitive necessities. While coordinating the planning functions of related enterprises, key issues of **mass customization** [8] had to be resolved by exploiting the cooperative attitude of the partners.

2 PROBLEM STATEMENT

In general, a supply network is **coordinated** if (1) the service through its channels can be guaranteed at a reasonably high level, and (2) the expected total production and logistic costs along the supply channels are minimal [1]. Whenever order lead times acceptable by the market are shorter than the actual production lead times, then the above criteria are in conflict: due to uncertain market conditions, production should be driven by **forecasts**, and **inventories** (of components, packaging materials, products) are inevitable to provide service at the required level. Low production costs can be achieved only with larger production quantities, which involve, again, higher product and component inventories as well as work-in-process. However, if demand unexpectedly ceases for a product, then accumulated inventories become **obsolete**.

In order to make the above conflicting objectives manageable, we suggested detaching them: while service level should be tackled on the short term (where information concerning demand is almost certain), cost-efficient production should be the main concern of medium term planning [6,9]. We proposed also coupling the two levels by the suitable management of safety stocks.

This paper discusses the problem of coordinating the **planning level** decisions of autonomous partners. We depart from a two-echelon, single product chain, where the manufacturer of end-products is in the role of the **customer** who needs component delivery from the **supplier**. The customer, being closer to the market of end-products prepares forecast for market demand, plans its production and derives the dependent demand for components. This forecast for component demand is passed to the supplier, who, in turn, has the right information for making appropriate decisions about the production of components. In order to account for unexpected future demand, communication and decision-making is performed on a **rolling horizon**, as suggested in [10].

While market demand is an external source of **uncertainty**, additional, internal uncertainties have to be dealt with that ensue from the rational behavior of partners. Given the risk of component shortage, the customer may tend to forward inflated forecasts to the supplier. Alternatively, if planners at the customer's side are rewarded for over-performing their plans, then they deliberately underestimate the demand and share too pessimistic forecasts with the suppliers. The supplier, however, will try to outwit the customer and tinker with the forecast. All this typically leads to corrupted service, increased nervousness, obsolete inventories and additional costs.

We are interested in eliminating this internal, and in fact, unnecessary uncertainty by coordinated planning. The main requirement towards the coordination mechanism is that it should be **incentive compatible** [11]: if the partners act locally in a rational way (minimize their costs and payments, respectively), then, as a result, the channel will be coordinated. Incentives, in turn, should be based on new, appropriate performance measures [12].

Our **mechanism design** problem, as suggested in [1], was resolved in three stages: First, supposing a central agency, planning methods for uncertain demand were developed. Next, a decentralized mechanism was elaborated that drives the partners to achieve the theoretically first-best, central solution. Finally, channels for information exchange were implemented to make the ideas operational. The subsequent sections follow this logic.

3 CENTRAL CHANNEL COORDINATION

3.1 Model for multi-period planning

Integral part of supply chain management is to decide whether and how much to produce from particular components, given uncertain demand. Here, we model this problem first without any regard to information asymmetry. Demand forecast of a component is provided by the customer for a multi-period, finite-length planning horizon. However, this demand is uncertain: it may stop whenever the 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. Further on, there are no inventory and production capacity limits. Consequently, the supplier has **no speculative motive**: it is always preferable to produce at a later period than producing earlier and holding stock. The lead-time of components (manufacturing plus shipment) fits into a single time period. Finally, the supplier has to satisfy all demand; i.e., backlogs are not allowed. The objective is to minimize the **expected total cost** that includes the costs of setups, inventory holding and eventual obsolete inventories.

Our model has a planning horizon of n periods where demand forecast for any period i ($1 \le i \le n$) is given by a non-negative f_i. The random variable $\eta \in \{1,\ldots\}$ represents the time period when run-out happens (see Figure 1). The relevant cost parameters are the inventory holding cost per piece and per time period (h), the setup cost (c_s), and the production cost per piece (c_p). We measure also the unit cost of obsolete inventory by c_p. The basic question is how much to produce in each period; hence, the decision variables are the production quantities $x_0,\ldots,x_n$. Finally, the auxiliary variables $I_0,\ldots,I_n$ represent inventory levels at the end of the periods.

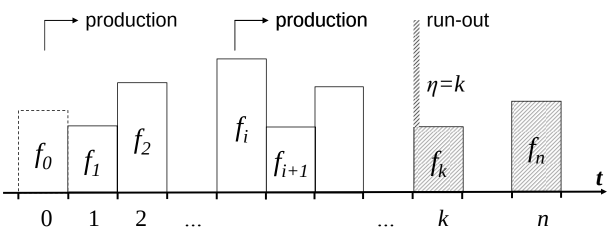

Figure 1: Planning on a multi-period horizon.

3.2 Solution alternatives

Earlier, we have worked out two kinds of solutions to the problem formulated above. Disregarding the less trusted remote forecasts, **heuristic policies** minimize the expected average cost – either per time period or per piece – on the initial segment of the horizon [6]. These average cost methods (*ACx*, *ACq*, respectively) determine only the first production quantity, x_o.

Recently, so as to cover the whole horizon, we have extended the classical **Wagner-Whitin method** with the probability of run-out, and its consequence, obsolete inventory [7]. This method (called *WWr*) has a specific property that can be derived from the lack of speculative motive: in any period, it is optimal not to produce, unless the inventory would become empty otherwise. The goal is to construct such a $x_0,\ldots,x_n$ production plan that minimizes the expected total cost which consists of (1) the setup cost in the initial period, (2) the sum of the expected setup plus inventory holding costs assuming no run-out happens, and (3) the expected cost of obsolete inventory in case of run-out. Formally, the expected total cost, $E[TC]$ equals

$$c_s + \sum_{i=1}^{n}\left(\Pr(\eta > i)\left(c_s\delta(x_i) + h\left(I_{i-1} - f_i/2\right)\right) + \Pr(\eta = i)c_p I_{i-1}\right)$$

where $\delta(x) = 0$ if $x = 0$, and $\delta(x) = 1$ if $x > 0$. The **stochastic programming** formulation of the problem minimizes $E[TC]$ subject to inventory balance and integrity constraints. The search space is restricted to solutions that exhibit the Wagner-Whitin property: $I_n = x_n = 0$, and for each period $i \in \{1,\ldots,n\}$, $(I_{i-1} - f_i)\delta(x_i) = 0$. For details of *WWr* and a solution algorithm, see [7].

As reported earlier, extensive experiments with the *AC* and *WWr* methods on industrial datasets confirmed their practical applicability. In these test, run-out was assumed to have a geometric distribution, i.e., it had the same probability p in the next period if run-out did not happen before. As for an example, see Table 1 that presents different production plans generated for a particular forecast considering various run-out probabilities. Taking realistic cost parameters, h = 0.02, c_p=3, and c_s = 10000. With the increase of run-out probability more and more cautious production quantities are planned, even in the face of relatively high setup costs.

i		0	1	2	3	4
fi		N/A	9990	10310	23091	22734
$p=0$	*ACx*	66125	N/A	N/A	N/A	N/A
	ACq	66125	N/A	N/A	N/A	N/A
	WWr	66125	0	0	0	0
$p=0.05$	*ACx*	20300	N/A	N/A	N/A	N/A
	ACq	43391	N/A	N/A	N/A	N/A
	WWr	20300	0	45825	0	0
$p=0.15$	*ACx*	20300	N/A	N/A	N/A	N/A
	ACq	20300	N/A	N/A	N/A	N/A
	WWr	20300	0	23091	22734	0

Table 1: Solutions for a multi-period planning problem.

The above formulation is **invariant** to the actual distribution of run-out probability. In fact, η can have any discrete distribution, though those with one parameter (like geometric, uniform or Poisson) are preferred in order to facilitate the easy exchange of this uncertainty related information between the partners (see Section 4.4).

4 DECENTRALIZED COORDINATION MECHANISM

4.1 Principles of the mechanism

In a real supply network, **no central planner** exists with all the required information as assumed above. Instead, available information is asymmetric and responsibilities are distributed. In a two-echelon chain, the supplier is familiar with the production and setup cost (c_p, c_s) for the components, while the customer can provide the f_i demand forecasts, together with information about run-out probabilities. The supplier is able (and interested) to control also its own costs. Clearly, the decision right on when and how much to produce should be assigned to the supplier. Would the supplier be confident that it gets unbiased information about the demand, then its optimal decision would coordinate the channel. However, in this setting there are two basic issues with regard to the nature and quality of the communicated information:

- When planning on a rolling horizon, the customer generates in each time period a new forecast. Forecasts relating to the same period typically differ from each other, just like from the realized demand.
- The customer may have various motivations to distort the demand related data (see Section 2). Knowing this, the supplier may try to outguess the customer. All this leads to deteriorated performance.

As a resolution, we present a **mechanism** that drives the partners towards disclosing and using unbiased information when trying to coordinate the channel. Accordingly, the supplier provides a **service** to the customer by committing itself to meet all the actual demand. Preparations to this service are made by relying on the series of forecasts and the related uncertainty. In return, the customer pays for (1) the components delivered, (2) the difference between the forecasted and realized demand, and (3) the uncertainty of the demand. Hence, the customer takes responsibility for the quality of its forecasts. Since the proposed coordination mechanism is incentive compatible, cooperation between autonomous partners will be an emergent, but at the same time, guaranteed property of the system.

4.2 Measuring forecast imprecision

The customer's main criterion is the difference between its planned and actual behavior, which is measured by the **imprecision** of the forecasts. On a rolling horizon, new forecasts are generated period by period. Let $f_{i,j}$ denote the forecast for period i generated on period j ($j+1 \le i \le j+n$), while u_i the realized usage in period i.

	Production period										
	-7	-6	-5	-4	-3	-2	-1		*d*	*c1 d*	*RC*
-11	13893								N/A	N/A	N/A
-10	8219	8121							N/A	N/A	N/A
-9	15364	13497	7861						N/A	N/A	N/A
-8	8944	22460	11235	9357					1842	N/A	N/A
-7	5334	31938	10874	9628	11742				3405	N/A	72
-6		23884	9990	10310	23091	22734			3801	N/A	111
-5			11946	11946	30146	21077	16117		3651	N/A	-64
-4				18200	27384	19742	15203		N/A	37	152
-3					23772	27383	16439		N/A	68	-95 *
-2						27413	23928		N/A	76	46 *
-1							24506		N/A	73	32 *
e	6271	8907	1956	7890	5674	4679	6584	5827	3175	63	
$c_1 e$	125	178	39	158	113	94	132	117			

Table 2: Rolling horizon forecasts, their imprecision measures and compensations ($c_1=0.02$).

For an example, see Table 2 that contains actual data from our application (see Section 5): the weekly demand for some packaging material. The index of the current week is 0, so the table contains historic data. Each row presents a forecast generated on a particular week for the next n=4 weeks. Forecast used in our previous example (see Table 1) was generated six weeks ago, on week -6. The u_i values of usage are in blue boxes.

For measuring imprecision, one usual option is to take the **error** for period i as the average of the absolute differences between the realized usage and the various forecasts referring to that particular period (in the example, one of the bottom lines shows these errors).

$$e_i = \frac{1}{n} \sum_{j=i-n}^{i-1} \left| f_{i,j} - u_i \right| .$$

Alternatively, we introduce the **deviation** of the forecast generated in period j as the absolute average difference between the **total** forecasted and realized demands throughout the horizon. Formally,

$$d_j = \frac{1}{n} \left| \sum_{i=j+1}^{j+n} (f_{i,j} - u_i) \right| .$$

For example, see column d on Table 2. When comparing errors with deviations one can see that errors are usually larger, because they contain double penalty for demand shifting between periods. E.g., on Table 2 the average error is 5827, while average deviation is 3175. In general, deviation is a more appropriate measure if demand is satisfied from the inventory and there is no need to penalize demand shifts between periods on a relatively short horizon.

4.3 Payment scheme with rolling compensation

The key issue in supply coordination is to make the customer motivated in sharing its demand related information truthfully. Hence, the customer should take responsibility for the quality of the forecasts. First, we discuss the case when run-out may not happen. Suppose the supplier satisfied in period k the actual demand u_k. In return to this service, the customer should pay (1) for the realized demand, and the (2) the imprecision of the forecast.

If we measure imprecision in terms of the forecast error, then this **payment** in any period k is $P_k = c_0 u_k + c_1 e_k$ (here c_0 is the unit price for delivered items and c_1 is the compensation factor). The expected payment of the customer is minimal if it always shares its true forecasts.

The case is more complicated if the relevant measure is deviation, because the complete payment can be calculated only at the end of the horizon, after n periods. With the payment function $P_k = c_0 u_k + c_1 d_k$, sharing the true forecast is an **optimal** but not unique decision for the customer; any re-distribution of the demand that produces the same sum is also optimal. Both of the above issues can be resolved if we introduce the idea of **rolling compensation**. Let us define the payment for the service in period k distributed throughout the periods $k, \ldots, k+l$ as follows:

$$P_{k,0} = c_0 u_k,$$
$$P_{k,l} = c_0 u_k + \frac{c_1}{l} \left| \sum_{i=k+1}^{k+l} (f_{i,k} - u_i) \right|, \qquad (l = 1, \ldots, n).$$

By definition, $P_i = P_{i,n}$. With rolling compensation, the customer should pay

- $P_{k,0}$ in period k for the immediate delivery, and
- $P_{k,l} - P_{k,l-1}$ in the periods $k+l$ ($l=1,\ldots,n$), respectively.

The partners estimate expected payment for period k in the periods of the subsequent interval [k, $k+n$], and as times advances the customer revises the compensation it paid earlier. In the end, the total payment will be just P_k. Payment in some of the periods may also be negative, meaning that the compensation was overestimated previously. E.g., see the RC column on Table 2 (where * denotes that compensations for weeks -3, …, -1 having incomplete horizons are excluded). Since any re-distribution would result in an early over-compensation that would be paid back only later, sharing the true forecast is the unique optimum for the customer.

4.4 Compensation with considering run-out

In Section 3 we have presented a model for medium-term planning that includes also the probability of run-out in an arbitrary time period. When coordinating a channel, also this information should be shared with the supplier. In what follows, the type of distribution is considered **common knowledge**, while the actual value of the parameter is **private information** of the customer. Eventual obsolete inventory what is the consequence of run-out will belong to the supplier. We propose the following extension of the rolling compensation scheme:

- The original payment scheme should be applied until the time period when run-out eventually happens.
- The customer should also pay for the additional uncertainty. The smaller the communicated uncertainty, the smaller is this compensation item.
- In case of run-out the customer also compensates, but smaller uncertainty communicated results in higher compensation.

These requirements are met by the scheme below:

$P_k = c_0 u_k + c_1 d_k + c_2 q(\eta_k)$ if no run-out happens,

$P_k = P_{k,l-1} + c_2 r(\eta_k)$ if run-out happens in period l.

This P_k payment function has three components: (1) Payment for the delivered items, which is independent from the decision variables. (2) Compensation for forecast deviation, which is independent from the probability of run-out. Therefore, the customer has the same incentive to share the right forecast as in the deterministic case. (3) Compensation for run-out uncertainty which is independent from $f_{i,j}$. Hence, in an incentive compatible mechanism $q(\eta_k)$ and $r(\eta_k)$ should be constructed so that the customer be interested in sharing the right parameter of the run-out probability. Due to lack of space, here we can but present (and not prove) that the functions $q(\eta_k) = -\ln(\Pr(\eta > n))$ and $r(\eta_k) = -\ln(1 - \Pr(\eta > n))$ have this property. All in all, the expected payment of the customer is minimal if it shares any time unbiased forecasts and uncertainty information with the supplier.

5 INDUSTRIAL APPLICATION

As part of the work reported here, we have implemented and deployed a so-called Logistics Platform (LP) in order to facilitate coordinated planning in the supply network presented earlier. The architecture of the web application corresponds to the focal structure of the network (see Figure 2). While master data is taken from the transactional enterprise resource planning (ERP) system, the series of $f_{i,j}$ dependent component demand forecasts are generated by the production planner of the focal manufacturer. High service level is assured via the exchange of short term scheduling related information, (an aspect of coordination not discussed in this paper). Short term scheduled demand is generated for each component day by day, which are, in turn, compared to the projected delivery schedules of the suppliers. Together with the forecast imprecision measures presented above, we have implemented also novel performance criteria for the service of suppliers (for details, see [9]).

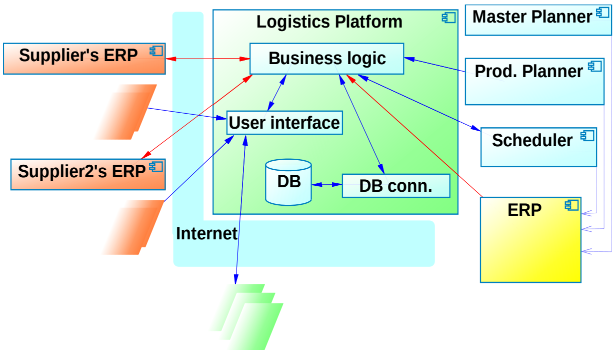

Figure 2: The Logistics Platform and its connections.

The LP which is in everyday use handles the data of cc. ten thousand components. By now, the performance of partners improved: the supply process became more transparent, future conflict situations can be anticipated, and, thanks to the new performance measures, the operation of each partner is continuously evaluated.

6 CONCLUSIONS

We have presented a novel coordination mechanism for supply planning that, by its incentives, rules out the opportunistic use of private information. The outcomes of local decisions are coordinated channels on the medium-term. The required input information except run-out probability is directly available in standard enterprise information systems, and the proposed protocol for rolling horizon planning fits the managerial practice. In order to make the coordination mechanism fully operational, the partner enterprises have to agree in novel **business models** and contracts that control the flow of materials, information and financial assets alike. This would not only increase the responsiveness of a network operating under uncertain market conditions, but also make its members committed to take common responsibility for the economical use of resources. Such a commitment is a prerequisite of sustainable development.

7 ACKNOWLEDGMENTS

This work was supported by the NKFP Grant No. 2/010/2004 (VITAL), as well as the FP6 projects AC/DC No. 031520 and COLL-PLEXITY No. 012781.

8 REFERENCES

[1] Li, X., Wang, Q., 2007, Coordination Mechanisms of Supply Chain Systems, European Journal of Operational Research, 179/1:1-16.

[2] Hatvany, J., 1985, Intelligence and Cooperation in Heterarchic Manufacturing Systems, Robotics and Computer-Integrated Manufacturing, 2/2:101-104.

[3] Wiendahl, H-P., Lutz, S., 2002, Production in Networks, CIRP Annals Manufacturing Technology 51/2:573-586.

[4] Monostori, L., Kumara, S., Váncza J., 2006, Agent-Based Systems for Manufacturing, CIRP Annals Manufacturing Technology, 55/2:697-720.

[5] Ueda, K., Márkus, A., Monostori, L., Kals, H.J.J., Arai, T., 2001, Emergent Synthesis Methodologies for Manufacturing, CIRP Annals Manufacturing Technology, 50/2:535-551.

[6] Váncza, J., Egri P., 2006, Coordinating Supply Networks in Customized Mass Production – A Contract-Based Approach, CIRP Annals Manufacturing Technology, 55/1:489-492.

[7] Egri P., Váncza J., 2007, Cooperative Production Networks – Multiagent Modeling and Planning, Acta Cybernetica, 18:223-238.

[8] Tseng, M.M, Piller, F.T., (eds.), 2003, The Customer Centric Enterprise, Springer.

[9] Váncza J., Egri P., Karnok D., 2007, Planning in Concert: A Logistics Platform for Production Networks, Proc. of 4th Int. CIRP Conference in Digital Enterprise Technology, 461-470.

[10] Tolio, T., Urgo, M., 2007, A Rolling Horizon Approach to Plan Outsourcing in Manufacturing-to-Order Environments Affected by Uncertainty, CIRP Annals Manufacturing Technology, 56/1:487-490.

[11] Laffont, J.J., Martimort, D., 2002, The Theory of Incentives, Princeton University Press.

[12] Hon, K.K.B., 2005, Performance and Evaluation of Manufacturing Systems, CIRP Annals Manufacturing Technology, 52/2:675-690.