Schönberger, J.; Kopfer. H.: Transport system responsiveness improvement. In: International Journal of Physical Distribution and Logistics Management, 39(2009)1, pp. 63-79, This paper was nominated for the best paper award of the NOFOMA 2008 conference

Transport system responsiveness improvement

Jörn Schönberger and Herbert Kopfer
University of Bremen, Bremen, Germany

Abstract

Purpose – The coping of demand oscillation is an important challenge in dynamic transport planning. A reliable request fulfillment must be provided even if the number of incoming requests temporarily climbs over the expected demand and resource scarceness appears. The aim of this paper is to propose an innovative planning approach that enables a transportation fleet to maintain a sufficiently high percentage of timely-fulfilled customer requests even in demand peak situations.

Design/methodology/approach – The effectiveness of the new approach is verified in computational simulation experiments. Quantifications for the system's responsiveness are proposed. Then, the quantified knowledge about the intermediate responsiveness is exploited to adjust the decision model representing the next schedule update task in a rolling horizon re-planning.

Findings – The observed simulation results suggest the suitability of the proposed approach. An adjustment of the plan update model supports the maintenance of a high percentage of timely completed requests during and after the demand peak.

Research limitations/implications – The generic approach presented and evaluated here motivates an adaptation to other more practical problem settings, in order to show its general applicability.

Practical implications – The proposed methodology contributes to the current demand for computational support for increasing the responsiveness of logistic systems.

Originality/value – The original contribution of this paper is the autonomous feedback-controlled adjustment of decision preferences which enables a rolling horizon re-planning framework to maintain a stable output performance even if the input oscillates significantly over time.

Keywords Transportation, Journey planning, Flexibility, Response flexibility

Paper type Research paper

1. Introduction

Flexibility subsumes the abilities of a value creating or service providing system to cope with the abruptly varying needs (requests) appearing from its uncontrollable environment (Schneeweiß and Schneider, 1999). Flexibility issues are of interest for both the system deployment level (operational planning) as well as with respect to the system configuration level (tactical planning).

Systemflexibility is a property of the investigated system that describes the general ability of the system to meet the requirements of additional requests released over a (representative) period of time. The determinations of provided capacities as well as the selection of strategies and decision preferences for the deployment are exploited to achieve a sufficiently high degree of systemflexibility. Investing into the increase of the degree of systemflexibility during the system configuration is a kind of hedging the

This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Center 637 "Autonomous Cooperating Logistic Processes" (Subproject B7).

International Journal of Physical Distribution & Logistics Management
Vol. 39 No. 1, 2009
pp. 63-79

0960-0035
DOI 10.1108/09600030910929192

system against uncertainty related to future requests. It enables the usage of the system in a larger variety of environments.

Planflexibility is a property of a plan executed in the system at a specific time. In the deployment of resources, a flexible plan is able to be updated so that the requirements of additionally appearing requests are met (responsiveness). However, it depends upon the configuration of the system which additional requests can be integrated in a plan. A sufficiently high capacity must be available and the preparation of effective and efficient decision preferences for integrating additional operations into the existing plan is necessary.

Since the additional demand cannot be forecasted, the configuration of the system requires a temporal reconfiguration if the problem difficulty changes significantly, e.g. if the available capacities are exhausted. In this paper, the exploitation of the information about the current planflexibility in the adjustment of the system configuration by varying the decision preferences used to integrate the additional demand into existing processes is proposed. The verification of the following research hypothesis is targeted:

> The long term responsiveness of a (transport) system is able to be increased if the applied decision preferences for the demand integration are adjusted to the intermediately observed responsiveness. The better the information about the intermediate responsiveness is the higher is the achieved degree of long term responsiveness.

The discussion of system and planflexibility starts with definitions of terms and measurements (Section 2). Then the investigated transport system and the planning challenge are described (Section 3). Next, a self-adjusting planning system is presented (Section 4). In Section 5, the executed computational experiments and the observed results are reported.

2. Flexibility and logistic operations

2.1 Literature

Flexibility-related investigations explore the range of different exogenous demand under which a given system performs well or "as intended" or "as required." Furthermore, bottlenecks that limit the system's ability to adjust to the changed exogenous demand situation are searched for. The goal is to analyze and widen or bypass bottlenecks, so that the responsiveness of the system increases with respect to the need originating from the systems environment.

de Groote (1994) proposes to distinguish:

(1) the investigated system (technology);

(2) the system's environment; and

(3) one or several evaluation criteria to evaluate the behavior of a system in its environment.

Most of the work related to (1) and (2) is linked to applications from production or inventory planning (Beach *et al.*, 2000; Pibernik, 2002; de Toni and Tonchia, 1998; Sethi and Sethi, 1990; Gupta and Goyal, 1989). Here, uncertainty about environmental needs is connected to the inability to forecast the demand time as well as the demand volume. Morlok and Chang (2004) also consider the uncertainty of the locations of demand for rail freight transport and thus cover the third dimension of uncertainty, which is the spatial uncertainty of demand. Different transport network types are compared by Feitelson and Salomon (2000) with respect to different non-numeric flexibility criteria. Flexibility is measured by counting the variety of a system's possible reactions to a particular environmental need, by technical indicators or by the costs to enable the system to fulfill the requirements of a new environmental demand (Schneeweiß and Schneider, 1999). The consideration of adjustment costs leads to system robustness issues (Scholl, 2001). The robustness discussion is beyond the scope of this paper, so that adjustment costs are ignored in the investigation reported in this paper.

Since a system evolves and changes its configuration over time and the flexibility property is given only for specific configurations and/or a specific time it becomes necessary to add specific information to a flexibility statement: the time (representing the system state in which the flexibility property was observed), the description of the disturbance for which the adaptability has been investigated as well as the rules for the integration of additional requests (integration logic and decision preferences).

2.2 Flexible plans and flexible systems

Let $S = (C, R)$ be a system consisting of a set C of system components and a set R of possible relations between the components. The components out of a subset of C are connected by means of activated relations taken from R at time t. This situation is referred to as state $X(S, t)$ of system S at time t.

A system control unit is able to change the state of S from $X(S, t')$ to $X(S, t^*)$ with $t' \leq t^*$. The transformation of the system during the period from t' to t^* is described by the transformation plan $P[X(S, t'), X(S, t^*)]$. The intermediate state of S during the transformation according to this plan at time t is $P[X(S, t'), X(S, t^*)](t)$.

An environmental need (request) is an exogenous demand whose appearance (in time, spatial variability and demand intensity) cannot be controlled by the system's control unit. Each request claims the fulfillment of specific requirements, that are fulfilled by the current system state or not. If and only if a request e appears at time t and if $X(S, t)$ fulfills the requirements of e then $X(S, t)$ is compatible with e at time t. A reconfiguration (transformation of the state) of S is not necessary. The compatibility of a system state comprises any requirement like time constraints (time windows), sufficient volume or the meeting of logical dependencies. If a request e is not compatible with $X(S, t)$ then it is called a disturbance of S at time t.

2.2.1 Planflexibility. The basic idea for defining the flexibility of a plan with respect to an additional request e appearing at time t is to check, whether the so far unexecuted part of this plan can be replaced by another transformation plan so that after the completion of the updated plan the requirements of e are fulfilled. If such an update of the performed transformation plan exists then the plan is denoted as e-planflexible.

Let $P[X(S, t'), X(S, t^*)]$ be the currently processed transformation plan for the reconfiguration of S from $X(S, t')$ to $X(S, t^*)$. The request e appearing at time t is a disturbance of S at time t, because it is not compatible with the intended final configuration $X(S, t^*)$ of S so that the reconfiguration of S into the state $X(S, t^*)$ is void now. A transformation of S from the current configuration $P[X(S, t'), X(S, t^*)](t)$ into a new target configuration which is also compatible with e is necessary.

If and only if there is a transformation plan Q converting the system configuration $P[X(S, t'), X(S, t^*)](t)$ into a configuration $Z_e(S, t(e))$ that is compatible also with e then the so far followed transformation plan is adaptable at time t with respect to the disturbance e. If such a transformation plan Q exists then the plan $P[X(S,t'),X(S,t^{*})]$ is e-planflexible at time t.

By means of the situation shown in Figure 1, the meaning of planflexibility is explained. The system S comprises a vehicle (server) that travels in the plane and fulfills requests by visiting a customer site associated with the request. Its travel speed is 1 length unit per time unit.

The server waits at position O at time 0. A request has been currently released and this request requires a server's visit at position D within the time window [5; 6]. The current position of the vehicle and the request are incompatible: for this reason, the vehicle starts to move to the location D following the bold arc representing the transformation plan P. According to its schedule, the server will reach D timely at time 6. At time 1, a second request appears which requires the visit of the customer site e within the time window [3; 4]. If the customer site associated with e is situated in the intersection of the ellipsis and the circle shown in Figure 2, then both the site belonging

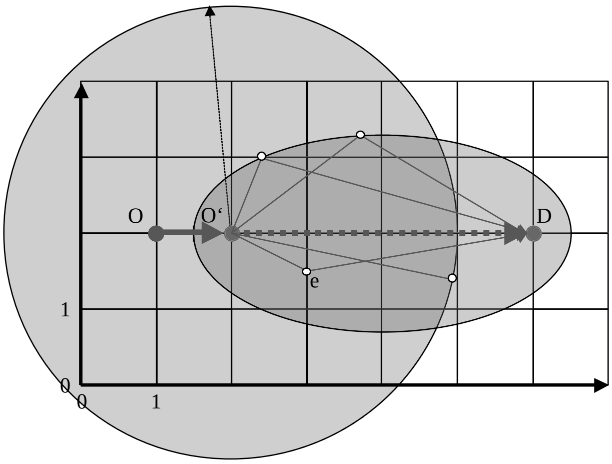

Figure 1.
An example scenario from transportation planning

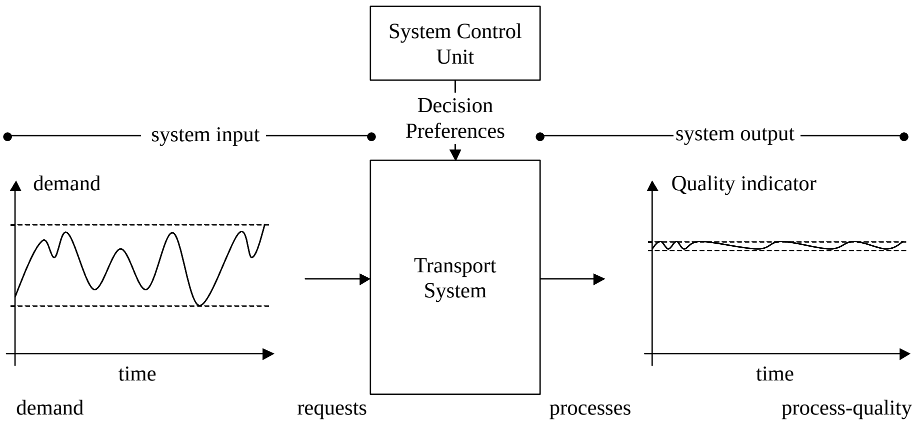

Figure 2.
Simplified representation of the investigated scenario

to e and D can be reached within the agreed time windows. In all other cases, it is impossible to visit both customer sites without a delay.

The set E is formed by all additional customer requests with time window equal to [3; 4] and a location belonging to the previously mentioned intersection. If and only if, P is e-planflexible at time 1 for all requests belonging to E, then the plan P is denoted as E-planflexible at time 1.

In the given example, the additional customer site request is inserted into the existing plan so that a minimal overall travel distance must be bridged. Therefore, the integration of the additional request follows the decision preference "least sum of travel distances". An application of another preference or integration logic like "first-come/first-serve" cannot guarantee the E-planflexibility of P at time 1. It is therefore necessary to add information about the used decision preferences V to the planflexibility statement. The refined planflexibility definition is now.

If and only if, there is a transformation plan Q which has been generated according to a given set of decision preferences V and that converts the system configuration $P[X(S,t'),X(S,t^*)](t)$ into a configuration $Z_e(S, t(e))$ compatible also with e then the so far performed transformation plan is adaptable at time t with respect to the disturbance e by utilizing the decision preferences V. If such a transformation plan Q exists then $P[X(S,t'),X(S,t^*)]$ is called e-planflexible at time t with respect to V.

2.2.2 Systemflexibility. Planflexibility is not able to inform about the ability of the decision preferences V to successfully integrate any additional request within a given time period into the so far followed transformation plan. To describe such general adaptation ability for the system S, the concept of systemflexibility is introduced. Systemflexibility describes the system's ability to react to any environmental need in a given period without having information about the explicit followed plan if the application of decision preference V is obligatory.

The system S is called e-systemflexible at time t under the decision preferences V, if and only if there is a state $X(S, t^*)$ with $t^* \geq t$, so that $X(S, t^*)$ and e are compatible. This definition does not exploit any information about the structure of S. It is not required that S is compatible with e at the appearance time t of e but it is necessary that S can be transformed into a state that is compatible with the request e. Systemflexibility describes the system's ability to change its state in order to meet the requirements of an environmental demand ("action volume" according to Schneeweiß and Schneider, 1999). The "reactivity" of S determines the speed of this change. In case that the adjustment time of S after the appearance of e is limited, e.g. if for every constant $K > 0$ the requirement $t \leq t^* \leq t + K$ is kept, then the system S is called real-time e-systemflexible at time t under the decision preferences V.

This definition of systemflexibility generalizes the "system capacity flexibility" of Morlok and Chang (2004) who state that "system capacity flexibility is the ability of a transport system to accommodate variations or changes in traffic demand while maintaining a satisfactory level of performance."

At a particular time, e-systemflexibility might be observed for different requests e. In order to enable a compact description of the systemflexibility-property at a given time t for a variety of requests, the set U formed by all these requests is introduced. The system S is called U-systemflexible at time t under the decision preferences V, if and only if S is e-systemflexible at time t under the decision preferences V for all $e \in U$. In order to enable a more general description of the systems ability to meet the demand of environmental needs, the restriction that the environmental need must appear at a particular time t is dropped in the following. Instead it is assumed that it appears in a period T. If S is U-systemflexible at time t under the decision preferences V at any time $t \in T$, then the system S is called U-systemflexible during T under the decision preferences V.

2.3 Quantification of flexibility

The consideration of responsiveness issues during the deployment preparation as well as the configuration of a system enables the creation of planflexible plans and schedules and systemflexible systems. However, it is necessary to quantify planflexibility as well as systemflexibility. The definition of a scalar or of a vector of scalars to represent the degree of flexibility is a prerequisite for a comparison of several deployment or configuration alternatives with respect to their responsiveness. Quantification requires knowledge about the totality of possible requests and disturbances and a feedback about those requests for which the associated requirements can be met. No assumptions about internal structures of the system S are made. Thus, only the comparison of the total demand the system is faced with and the demand satisfied as request by S ("output flexibility" according to Grubbström and Olhager, 1997) can be exploited.

A plan or a system might be able to appropriately incorporate no (additional) requests (0 percent), all additional requests (100 percent) or a subset of future requests. Therefore, Schneeweiß and Schneider (1999) and Barad and Even Sapir (2003) as well as Corsten and Gössinger (2006) propose to quantify flexibility by the degree of satisfaction of the requirements of additional requests which is equivalent to the relative frequency that an occurred request can be integrated appropriately.

2.3.1 Measures for planflexibility. The planflexibility of a plan P refers to the ability of the decision preference V to integrate one or more additional requests into P at a particular time t by replacing the so far unexecuted part of P.

Let N^{α} be the overall number of all requests that could appear at time t jeopardizing the current plan $P[X, Y]$, n^{α} be the number of these requests e for which $P[X, Y]$ is e-planflexible at t and E contains all requests that could appear at time t. Then the α-degree for E-planflexibility of $P[X, Y]$ at time t under V is defined as:

$$F^{\alpha}(V, E, P[X, Y], t) := \frac{n^{\alpha}}{N^{\alpha}}. \tag{1}$$

The integration of additional requests into the plan P affects the already included operations as well as the previously made assignments of tasks to resources and therefore makes the re-scheduling of execution times necessary. The α-planflexibility does not consider these crowding-out effects, since it does not consider whether the requirements of previously integrated environmental needs are still met. Let N^{β} be the overall number of all requests contained in P and let n^{β} be the number of those requests whose requirements are all satisfied at time t. The β-degree of E-planflexibility of $P[X, Y]$ at time t in case that the decision preference(s) V are applied is defined as:

$$F^{\beta}(V, E, P[X, Y], t) := \frac{n^{\beta}}{N^{\beta}}. \tag{2}$$

Neither the α-degree nor the β-degree for planflexibility considers already completed requests within the calculation of the relative frequency (probability) for the satisfaction of the requirements associated with a request.

The inclusion of recently fulfilled requests in the calculation of the flexibility degree prevents an overweighting of temporal demand peaks. On the other hand, if the fulfillment time of a request is scheduled far from now, then information about the satisfaction of the associated requirements is of reduced worth. It is reasonable to consider only those requests, whose completion is scheduled for the near future. In the γ-degree for E-planflexibility of P at t the satisfied demand is compared with the total demand that falls into the time window $[t - \Delta t; t + \Delta t]$. Let N^{γ} be the number of all requests completed within the period $[t - \Delta t; t]$ or scheduled to be completed during the period $[t; t + \Delta t]$ and let n^{γ} be the number of those requests for which all requirements have been satisfied or are expected to be satisfied. The γ-degree for E-planflexibility of P at time t in case that the decision preferences V are applied is defined by:

$$F^{\gamma}(V, E, P[X, Y], t) := \frac{n^{\gamma}}{N^{\gamma}}. \tag{3}$$

2.3.2 Systemflexibility quantification. Barad and Even Sapir (2003) define the logistics dependability of a system as the probability that all requirements of the released requests are fulfilled by using the system S during an observation period. This idea is seized and the systemflexibility-degree for U-systemflexibility during T of S is similarly defined. Let M denote the number of all possible requests belonging to U and appearing during the period T. If system S can handle m requests appearing during T as requested applying the decision preferences V to integrate the requests then the systemflexibility-degree for U-systemflexibility of S during T, denoted as $F^{\text{system}}(V, U, T)$, is defined by:

$$F^{\text{system}}(V, U, T) := \frac{m}{M}. \tag{4}$$

Using the introduced vocabulary from the concepts of planflexibility as well as systemflexibility and recognizing the decision preferences as part of the systems configuration the initially-stated research hypothesis is refined:

RH. The U-systemflexibility of a (transport) system during a period T is able to be increased if the decision preferences V of this system are adapted to the intermediately calculated degree of planflexibility. The γ-degree-application leads to a higher F^{system}-value than the β-degree-application and the β-degree-application leads to a higher F^{system}-value than the α-degree-application.

It is self-evident that this hypothesis cannot be verified for any transport system but the general correctness of the idea to exploit planflexibility knowledge to adjust the process planning system can be demonstrated by investigating a specific transport system, which is introduced in the next section. In Section 4, a planning system is proposed that is able to exploit the planflexibility knowledge in the variation of decision preferences.

3. Transport system deployment in a volatile environment

3.1 Scenario description

A control problem in the area of transportation and service providing logistics is investigated. Transport resources (vehicles) operating in a restricted area are waiting for their assignment to customer requests. Such requests represent the demand to visit a customer location and the simultaneous satisfaction of a time window-side requirement.

Today's supply chains are temporal coalitions of independent partners, each having own disposition competency and responsibility. The independent partners are coordinated by a centrally acting coordinator, which interacts with the partners' subsystem control units in order to match the cross-partner flow of goods and materials. The partners committed themselves to rules for enabling the interactions in the supply chain coalition.

A transport system is a part of a supply chain coalition that realizes the material flow through the different value creation stages. Therefore, the synchronization of the transport processes with the procurement, production and sales activities performed by the other coalition partners is obviously necessary. Here, the matching of the subprocesses is realized by the specification of time windows which are determined in order to harmonize the transport of goods with upstream and downstream activities. To ensure that the value creation is not interrupted by transport services of low quality, it is necessary that at least the portion p^{target} of all transport requests is served without delay: customer sites must be visited during the specified time windows. Since hedging all imponderables leads to exploding costs, the target planflexibility degree is fixed to a value less than 100 percent. In the remainder of this paper, it is assumed that p^{target} is set to 80 percent.

Figure 2 shows the situation of the transport system as a part of the supply chain. Customers announce their demand for the visit of a vehicle belonging to the transport system by giving a specific request to the coordinator who forwards the request to the transport system (the system input). The system control unit (the dispatcher) consolidates different requests into processes according to the currently used decision preferences (the deployment rules). The processes describe how the transport resources in the transport system fulfill the requests. Process quality indicators observe the process quality (the system output) and therefore the fulfillment of the customer demand and the specified side requirements.

The sum of demand directed to the transport system varies significantly over time. In the repair and maintenance context (Bertsimas and van Ryzin, 1989), severe and wide spread damages caused by bad weather conditions or electric supply failures let the demand temporarily increase. A forecast of the demand peak is impossible. In the freight transport context, the placed demand of a bulk purchaser requires immediate supply. Although the demand of the customer(s) varies significantly over time (left curve in Figure 2) the customer(s) expect(s) a stable supply quality which means that punctuality and reliability are expected to vary only slightly (right curve in Figure 2) because the transport of their goods must be compatible with the scheduled downstream activities of the other partners in the supply chain.

3.2 The coordinator's decision model adjustment problem

Owing to the varying demand the transport system's workload oscillates over time. In peak situations with very high workload the capacity of the own equipment is exhausted. A temporal adaptation of the own transport capacity to the workload is impossible.

In production and inventory applications, a quick capacity adjustment is achieved by shifting manpower or machine capacities between departments of a factory (production) or producing to and replenishing from stock (inventory). In transport systems, such an internal temporal capacity increase is impossible. Instead, subcontractors are hired (Krajewska, 2008). If a request is subcontracted then another transport service company is paid for the reliable fulfillment of the request. The incorporation of a subcontractor is motivated by the guaranteed timely request fulfillment (the competitive advantage of courier and express services). However, the costs for the usage of this mode are higher than the costs for using an own vehicle to fulfill the request.

The rolling horizon re-planning and re-scheduling approach is used to cope with the uncertainty about future requests. An incoming request instead of reaching a pre-specified re-planning time initiates the next re-planning. Then, a new deployment decision task has to be solved by the system control unit. During the re-planning, the additional requests are integrated into the existing routes of the own vehicles (make) or they are subcontracted (buy). In the first case, the existing routes of the own vehicles are updated but in the second case, an external partner is hired. A once subcontracted request cannot be re-integrated into the route of an own vehicle. The decision for subcontraction is irreversible.

Before the necessary re-planning decisions are made, it is checked whether the current planflexibility degree of the request fulfillment is still sufficiently high (lower arc in Figure 3). In case that it is lower than the threshold value or runs into danger to fall below the threshold, the coordinator re-configures the transport system by submitting a signal to the transport system control unit that more requests must be subcontracted (upper arc in Figure 3). The intensification of the subcontraction usage avoids further increase of the workload of the own vehicle fleet so that a reduction of the future workload of the transport system is achieved. Now, the own vehicles can finish their already started operations without additional detours. If a re-increase of the planflexibility degree is detected (compared to the last re-planning time) then a signal is send from the coordinator to the transport system control unit that it is not necessary anymore to subcontract as much requests as before.

In previous research reports, two rules for implementing the required intensification or relaxation of the subcontraction usage have been proposed. Both methods adapt the intensification to the current planflexibility by manipulating the maintained formal

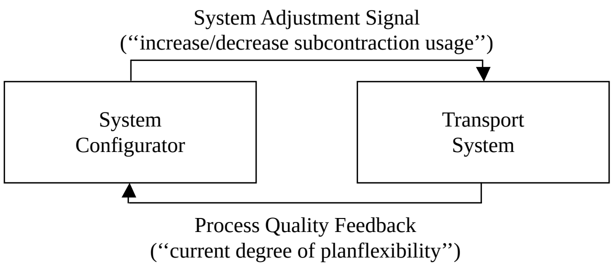

Figure 3.
System configuration cycle

mathematical optimization model used for the deployment decisions in the next upcoming re-planning task.

3.2.1 Adaptation of the constraint set (constraint set adaptation). The constraint representing the predetermined subcontraction decision is sharpened (affects more requests) if the planflexibility degree f_i decreases or if it runs into danger to fall below p^{target}. In case that the planflexibility re-increases the number of requests for which a subcontraction mode is pre-selected is reduced (Schönberger and Kopfer, 2007a, b). The set $R(t_i, f_i)$ containing all requests with a predetermined subcontraction decision becomes smaller.

3.2.2 Adaptation of the objective function (search direction adaptation). The coefficient ω_i of the self-fulfillment costs and the cost coefficient μ_i of the subcontraction in the used objective function F_{t_i} are re-calculated before the next instance of the decision model is solved: if the planflexibility degree f_i is low then the cost coefficient ω_i of the self-fulfillment mode is increased relatively to the cost coefficient μ_i of the subcontraction mode. If a re-increase of the planflexibility degree is observed then the self-fulfillment cost coefficient is lowered step-by-step (Schönberger and Kopfer, 2008).

If the rules between the coordinator and the transport system do not describe how to handle delayed arrivals at customer sites, then the least cost-oriented decision behavior of the transport system control unit damages the complete supply chain because reliability issues are completely ignored. Furthermore, it assigns the complete quality responsibility to the coordinator.

On the other hand, the transport system shareholders will not accept a rule that assigns the complete risk of a demand peak to the transport system alone, because such a rule is equivalent with the assignment of the complete risk and responsibility to the transport service provider. For this reason, it is impossible to predict the transport system to fulfill the portion p^{target} of all requests timely under all circumstances ignoring the absolute demand input to the transport system.

Constraint set adaptation (CSAD) and search direction adaptation (SDAD) represent two design alternatives for the cooperation rules agreed between the coordinator and the transport system control unit. They enable the risk sharing between the two partners (transport system and coordinator) because the transport system's profitability need is considered as well as the coordinator need for reliability of the coupled processes that determine the material flow through the supply chain.

3.3 The deployment problem

A transport system hosts 25-owned vehicles. Customer demands are released every 100 time units. The average number of incoming requests is 50 but a temporal demand peak of again 100 additional requests during time 1,500 until 1,700 interrupts the balanced incoming demand. The system control unit is able to hire a subcontractor for each request. A time window is associated with each request. If an owned vehicle or one of a subcontractor arrives before the time window opens, it has to wait. An arrival after the closure of the time window causes a penalty payment. Vehicles of subcontractors never arrive late.

The dynamic disposition task of the transport system control unit is mapped into an online optimization model (Krumke, 2001) consisting of a sequence of optimization problem instances $P_0, P_1, \ldots$ which are solved consecutively at the dispatching times

t_0, t_1, … Each instance P_i is complete in the sense that it considers all problem data known at time t_i. A once generated solution TP_i (set of processes) is executed until additional requests arrive at time t_{i+1}. At this time, the transport system control unit interrupts the execution of the so far executed processes, sets up a new optimization model instance $M(t_{i+1})$ applying CSAD or SDAD, solves this instance and replaces the not yet executed process parts from TP_i by the recently generated processes collected in TP_{i+1}.

After the data of the recent decision task has been collected at time t_i, a mathematical decision model (optimization model) $M(t_i)$ is stated as shown in the models (5)-(10). The requests in the set $R(t_i)$ are known but not yet completed at time t_i and must be distributed among the own fleet W and the incorporated subcontractor(s). All paths p assignable to a vehicle from the own fleet W are stored in the set $P(t_i)$ and all paths executable by a given vehicle $v \in W$ are collected in the set $P_v(t_i)$. Such a path starts at the current position of vehicle v and ends in the central depot. The binary parameter a_{rp} is 1 if and only if request r is served by path p. All requests, for which the SC-mode has already been selected in TP_{i-1} are collected in $R^E(t_i)$. The set $R^S(t_i)$ contains all those requests whose on-site operations have already been started but not yet finished at the current time t_i. If r has already been known at time t_{i-1} and if this request has not yet been subcontracted then $v(r)$ refers to the vehicle which was selected to visit the associated customer site according to TP_{i-1}.

The current value of the binary decision variable y_r is "1" if and only if request r is forwarded to a subcontractor. Furthermore, a (possible empty) path p has to be assigned to a vehicle v and p is assigned to v if and only if the binary decision variable x_{pv} is 1. To evaluate the decisions made, the travel costs $C^1(p)$ of the path p assigned to an own vehicles of fleet W, the penalty payments $C^2(p)$ for late arrivals associated with path p and the subcontraction costs $C^3(r)$ of request r are calculated. All decisions are made so, that the sum of costs calculated by using the currently applied objective function F_{t_i} is minimized equation (5):

$$F_{t_i} : \omega_i \cdot \sum_{p \in P(t_i)} \sum_{v \in W} (C^1(p) + C^2(p)) \cdot x_{pv} + \mu_i \sum_{r \in R(t_i)} C^3(r) \cdot y_r \rightarrow \min, \quad (5)$$

$$\sum_{p \in P_v(t_i)} x_{pv} = 1 \quad \forall v \in W, \quad (6a)$$

$$x_{pv} = 0 \quad \forall p \notin P_v(t_i), \ v \in W, \quad (6b)$$

$$y_r + \sum_{p \in P(t_i)} \sum_{v \in W} a_{rp} x_{pv} = 1 \quad \forall r \in R(t_i), \quad (7)$$

$$y_r = 1 \quad \forall r \in R^E(t_i) \cup R(t_i, f_i), \quad (8)$$

$$\sum_{p \in P_{v(r)}(t_i)} a_{rp} x_{pv(r)} = 1 \quad \forall r \in R^S(t_i), \quad (9)$$

$$y_r, x_{pv} \in \{0,1\} \quad \forall r \in R(t_i),\ p \in P(t_i),\ v \in W. \tag{10}$$

Each vehicle serves exactly one path equation (6a) and it is not allowed to assign a path p to a vehicle that cannot serve p equation (6b). Every request is either subcontracted or served by a vehicle from fleet W equation (7). However, previously subcontracted requests remain subcontracted equation (8). A request whose on-site fulfillment has already been started but not finished at time t_i cannot be re-assigned to another vehicle equation (9).

The objective function (5) and the constraint (8) enable the adaptation of the decision model to the current planflexibility degree f_i. Thereby, the knowledge acquired during the online-model processing is automatically fed back into the formulation of the next decision task(s) model. Such an approach extends the online decision-making framework which typically does not exploit feed-back information.

4. Planflexibility-exploiting planning system

4.1 Planning procedure

The algorithmic framework used to cope with the online optimization model introduced in Section 3 is shown in Figure 4.

Initially, the iteration counter i is set to 0 (a) and the first planning time is fetched (b). Next, an initial solution is generated (c) and broadcasted to the vehicles and to the subcontractors (d). Now, the procedure is idle and waits until the current solution has been completely executed or additional requests are received (e). In the first case, the procedure stops (f) and is re-started as soon as additional requests become known.

If the process execution is still in progress, then the iteration counter is increased by 1 (g) and the current system time is fetched (h). All requests just released at time t_i are collected in the set $R^+(t_i)$ (i). Next, it is checked whether the consideration of the additional requests compromises the current processes (j). The procedure falls back into an idle state if no process corruption occurs. Otherwise, the current planflexibility is calculated (k) and the intervention intensity (describing the severeness of the manipulation) is determined (l). If CSAD is applied then the requests which are prematurely directed into the subcontraction fulfillment mode are selected. Otherwise, if SDAD is used then the coefficients defining the next applied objection function are calculated (m). Afterwards, the new decision model is defined (n) and a high-quality solution of this model is derived (o) to replace the so far followed solution. The new solution is broadcasted to inform the field teams and the subcontractors (p). Again, the procedure falls back into the idle (waiting) state (q).

For the derivation of the new solution the memetic algorithm (MA) developed in Schönberger (2005) is started. The model (5)-(10) is a generalization of the vehicle routing problem with time windows which is NP-hard to solve so that is cannot be expected that an exact algorithm is able to derive an optimal solution in an acceptable computation time. However, the MA derives solutions of sufficiently high quality in an acceptable computation time.

The procedure GET_INTERVENTION_INTENSITY(f_i) evaluates the current planflexibility and returns the parameter s_{t_i} normalized between 0 and 1. If s_{t_i} is zero then no model adaptations are necessary, if s_{t_i} equals 1 then all possible adaptations are activated. The control value s_{t_i} is zero if the current planflexibility degree is higher than 85 percent but it is 1 if the current planflexibility has fallen below 75 percent. If the

```
PROCEDURE process_management();
(a)  i:=0;
(b)  t_i:= GET_CURRENT_TIME();
(c)  CurrentSolution :=GENERATE_INITIAL_SOLUTION();
(d)  BROADCAST(CurrentSolution);
(e)  wait until (CurrentSolution is completed) or (additional requests are re-
     leased);
(f)  if (CurrentSolution is completed) then goto (r);
(g)  i:=i+1;
(h)  t_i:= GET_CURRENT_TIME();
(i)  R+(t_i):= GET_RELEASED_REQUEST(t_i);
(j)  if not (SOLUTION_CORRUPTED(CurrentSolution)) then goto (e);
(k)  f_i:= GET_CURRENT_PLANFLEXIBILITY(t_i);
(l)  s_ti:= GET_INTERVENTION_INTENSITY(f_i);
(m)  R(t_i,f_i) := SPECIFY_INTERVENTION(s_ti; R+(t_i)); (only CSAD)
     (ω_i,μ_i) := SPECIFY_COEFFICIENTS(s_ti); (only SDAD)
(n)  M(t_i):= DEFINE_MODEL(t_i, CurrentSolution, R(t_i,f_i)); (only CSAD)
     M(t_i) := DEFINE_MODEL(t_i, CurrentSolution, (ω_i,μ_i)); (only SDAD)
(o)  CurrentSolution :=SOLVE_MODEL(M(t_i));
(p)  BROADCAST(CurrentSolution);
(q)  goto (e);
(r)  stop();
```

Figure 4.
Pseudo code of the algorithmic framework for the process management

planflexibility increases (decreases) between 0.75 and 0.85 then s_{t_i} is proportionally lowered (enlarged). The control value s_{t_i} is used now to setup the necessary decision model adaptations.

4.2 Implementation of the optimization model variation

If CSAD is used to adapt the decision model (5)-(10), then the ordered pair of the coefficients of the objective function is fixed to (1, 1) but the set $R(t_i, f_i)$ must be compiled for every re-planning task. At first the number $n(t_i)$ of requests to be inserted in $R(t_i, f_i)$ is set to the smallest number larger than $s_{t_i}|R^+(t_i)|$. Next, $n(t_i)$ requests are randomly selected from $R^+(t_i)$ and inserted into $R(t_i, f_i)$.

In case that SDAD is incorporated for the decision model adjustment the two coefficients ω_i and μ_i are determined before the model is solved. The coefficient μ_i is fixed to 1 and ω_i is adjusted according to the current value of the control signal s_{t_i}. If L denotes the average quotient between the fulfillment costs in the subcontraction mode and the fulfillment costs in the self-fulfillment mode then ω_i is defined as $\omega_i := (1 + s_{t_i}) \cdot L$. If the current planflexibility is quite high then the control signal s_{t_i} is 0 and $\omega_i = 1$. If the current planflexibility is low then the control signal is close to 1 and

$\omega_i \approx 1 + L$, which means that the objective function (5) recognizes the subcontraction as the cheaper fulfillment mode.

5. Computational experiments

5.1 Setup

A scenario is determined by a stream of incoming requests. Four different request streams have been used. Their generation is described in detail in Schönberger and Kopfer (2007a, b). The subcontraction rate is three times larger than the sum of penalty payments and travel costs ($L = 3$) so that the transport system control unit tends to refrain from the subcontraction mode selection.

Each scenario for the transport system has been simulated over a period of 5,000 time units. After a startup phase of 1,000 time units the planflexibility degree has been recorded throughout the next 4,000 time units while every 100 time units additional requests have been released. In the investigated artificial scenario U is defined to be the set of all possible additional requests so that this information can be dropped in the planflexibility statements. Each simulation run has been executed several times with different seeding because the memetic dispatching algorithm is a randomized procedure. Here, it is reported about the average results observed at particular times so that also the information about the particularly updated plan in the planflexibility statement is dropped. The degrees $F^{\alpha}(V,t), F^{\beta}(V,t)$ and $F^{\gamma}(V,t)$ are calculated for $t = 1{,}000, 1{,}100, 1{,}200, \ldots, 5{,}000$ in separate experiments applying $V \in \{\text{CSAD}, \text{SDAD}\}$. Furthermore, each scenario has been simulated without any decision-logic adaptation (NONE). Overall, 73 experiments were executed.

Let max(CSAD, α) denote the maximal value observed for $F^{\alpha}(\text{CSAD}, t)$ during the period [1,000; 5,000] and let min(CSAD, α) denote the minimal value observed for $F^{\alpha}(\text{CSAD}, t)$ within this time period. Furthermore, var(CSAD, α) is defined as the difference between max(CSAD, α) and min(CSAD, α). Finally, bel(CSAD, $\cdot$) gives the percentage of [1,000; 5,000] in which $F^{\alpha}(\text{CSAD}, t)$ has fallen below $p^{\text{target}} = 0.8$. The same values have been calculated for the SDAD experiments and for the β-degree as well as for the γ-degree of planflexibility.

After a scenario simulation is over, the quotient between the number of requests served within the assigned time window and the overall number of requests released during the observation period [1,000; 5,000] has been calculated, which is the observed systemflexibility F^{system} during this period.

5.2 Results

The observed key indicator values for the planflexibility degrees from the experiments are presented in Table I. The β-planflexibility degree demonstrates the highest

Table I. Observed degrees of planflexibility using CSAD and SDAD

	α-degree (percent)		β-degree (percent)		γ-degree (percent)	
	CSAD	SDAD	CSAD	SDAD	CSAD	SDAD
max($\cdot,\cdot$)	95,5	95,8	91,3	90,6	87,4	84,6
min($\cdot,\cdot$)	81,3	82,6	67,0	73,3	75,3	75,4
var($\cdot,\cdot$)	14,2	13,2	24,3	17,3	12,1	9,2
bel($\cdot,\cdot$)	0	0	9,7	4,8	17,1	17,0

variation (between 67.0 and 91.3 percent), followed by the α-planflexibility degree and the γ-planflexibility degree. The range of the α-planflexibility degree lies completely above 80 percent, so that a decrease of F^{α} influences the adaptation of the decision model only slightly. In contrast, the two remaining measures also fall below the threshold value of 80 percent. However, F^{β} falls down to a very low level of less than 75 percent but F^{γ} oscillates uniformly around the threshold value.

It is concluded that the α-planflexibility is not able to control CSAD or SDAD because it is not able to detect significant planflexibility variations. The β-planflexibility shows a more reasonable behavior (in 9.7 percent, respectively, 4.8 percent of all measurements, it falls below the threshold and detects that the decision preferences must be adjusted) but the variation amplitudes of this measure are severe. Among the three planflexibility measures, the γ-planflexibility degree seems to be the best method to map the variation of the punctuality into the portfolio of decision preference modifications. In more than 17 percent, it indicates that the threshold value is not achieved.

Finally, the observed systemflexibility values are summarized in Table II. If neither CSAD nor SDAD is applied then the systemflexibility is 74.1 percent (NONE). This value is interpreted as follows. If a request is given to the transport system during the period [1,000; 5,000] then it will be fulfilled within its associated time window with a probability of 74.1 percent.

Independently from the application of CSAD or SDAD an increase in the systemflexibility has been observed if it is switched from the α-planflexibility measure to the β-planflexibility measure or from the β-planflexibility to the γ-planflexibility measure. In the last case, a systemflexibility value of 83.8 percent (CSAD), respectively, 83.0 percent (SDAD) has been observed. This observation verifies the stated research hypothesis: it has been shown that the systemflexibility is increased if the decision preferences, which are a part of the system configuration, are adapted continuously to the intermediate degree of planflexibility. Furthermore, it has been demonstrated that the systemflexibility increases if the planflexibility measurements become more sophisticated.

6. Conclusions

The responsiveness of a transport system operating in a volatile environment can be increased. Flexibility degrees have been defined from the short-term perspective (planflexibility) and from the long-term perspective (systemflexibility). The research hypothesis that the adaptation of the transport system's configuration (particularly the applied deployment decision preferences) to the particular planflexibility degrees leads to an increase of the systemflexibility has been verified.

V	α-degree (percent)	β-degree (percent)	γ-degree (percent)
NONE	74,1	74,1	74,1
CSAD	79,2	82,3	83,8
SDAD	79,9	82,0	83,3

Table II. Observed degree of systemflexibility using different planflexibility measures

Future research activities will include the integration of the two rules CSAD and SDAD as well as the transfer of the presented planning system framework to other logistic systems which have to cope with uncertain planning data.

References

Barad, M. and Even Sapir, D. (2003), "Flexibility in logistic systems – modeling and performance evaluation", *International Journal of Production Economics*, No. 85, pp. 155-70.

Beach, R., Muhlemann, A.P., Price, D.H.R., Paterson, A. and Sharp, J.A. (2000), "A review of manufacturing flexibility", *European Journal of Operational Research*, Vol. 122, pp. 41-57.

Bertsimas, D. and van Ryzin, G. (1989), "The dynamic traveling repairman problem", Working Paper, No. 3036-89, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

Corsten, H. and Gössinger, R. (2006), "Output flexibility of service enterprises – analysis based on production theory", *International Journal of Production Economics*, No. 104, pp. 296-307.

de Groote, X. (1994), "The flexibility of production processes: a general framework", *Management Science*, Vol. 40 No. 7, pp. 933-45.

de Toni, A. and Tonchia, S. (1998), "Manufacturing flexibility: a literature review", *International Journal of Production Research*, Vol. 36 No. 6, pp. 1587-617.

Feitelson, E. and Salomon, I. (2000), "The implications of differential network flexibility for spatial structures", *Transportation Research Part A*, Vol. 34, pp. 449-79.

Grubbström, R.W. and Olhager, J. (1997), "Productivity and flexibility: fundamental relations between two major properties and performance measures of the production system", *International Journal of Production Economics*, No. 52, pp. 73-82.

Gupta, Y.P. and Goyal, S. (1989), "Flexibility of manufacturing systems: concepts and measurements", *European Journal of Operational Research*, Vol. 43, pp. 119-35.

Krajewska, M. (2008), *Potentials for Efficiency Increase in Modern Freight Forwarding*, Gabler, Wiesbaden.

Krumke, S.O. (2001), "Online optimisation – competitive analysis and beyond", Habilitation thesis, Technical University of Berlin, Berlin.

Morlok, E.K. and Chang, D.J. (2004), "Measuring capacity flexibility of a transportation system", *Transportation Research Part A*, Vol. 38, pp. 405-20.

Pibernik, R. (2002), *Flexibilitätsplanung in Wertschöpfungsnetzwerken*, DUV, Wiesbaden.

Schneeweiß, C. and Schneider, H. (1999), "Measuring and designing flexibility as a generalized service degree", *European Journal of Operational Research*, No. 112, pp. 98-116.

Scholl, A. (2001), *Robuste Planung und Optimierung*, Physica-Verlag, Heidelberg.

Schönberger, J. (2005), *Operational Freight Carrier Planning*, Springer, Heidelberg.

Schönberger, J. and Kopfer, H. (2007a), "Flexible transport process planning in volatile environments and the adaptation of a cost-based objective function", in Otto, A. and Obermaier, R. (Eds), *Logistimanagement*, DUV, Wiesbaden, pp. 263-83.

Schönberger, J. and Kopfer, H. (2007b), "On decision model adaptation in online optimization of a transport system", in Günther, H-O., Mattfeld, D.C. and Suhl, L. (Eds), *Management logistischer Netzwerke*, Physica-Verlag, Heidelberg, pp. 361-81.

Schönberger, J. and Kopfer, H. (2008), "Manipulating the decision behavior of a subordinate service centre by a closed-loop cost accounting scheme control", in Suhl, L., Mattfeld, D.C., Günther, H-O. and Voß, S. (Eds), *Informations- und Kommunikationssysteme in Supply Chain Management, Logistik und Transport*, DSOR Contributions to Information Systems, Paderborn, pp. 3-19.

Sethi, A.K. and Sethi, S.P. (1990), "Flexibility in manufacturing: a survey", *International Journal of Flexible Manufacturing Systems*, Vol. 2, pp. 289-328.

About the authors

Jörn Schönberger is a post-doc Researcher at the Chair of Logistics in the University of Bremen associated with the research project "Autonomous adaptation of vehicle schedules," which is part of the Collaborative Research Centre 637 "Autonomous cooperating logistic processes – a paradigm shift and its limitations" funded by the DFG and supervised by Herbert Kopfer. His research has been published in international journals such as *European Journal of Operations Research* or *Computers & Operations Research*. He received the Dissertation-Award of the German Operations Research Society for his PhD thesis "Operational freight carrier planning – investigations on basic concepts, optimization models and advanced memetic algorithms" in 2004. Jörn Schönberger is the corresponding author and can be contacted at: jsb@uni-bremen.de

Herbert Kopfer is a Full Professor of Logistics at the Faculty of Business Studies and Economics at the University of Bremen. He supervised several research as well as industrial projects. Currently, he is the project leader of two research projects in the Collaborative Research Centre 637 "Autonomous cooperating logistic processes – a paradigm shift and its limitations" funded by the DFG. His research has been published is several international journals like *European Journal of Operations Research*, *Computers & Operations Research*, *Journal of the Operational Research Society*, *OR Spectrum*, *Transportation Research – Part A*. In addition, he has edited several books and journal special issues.