CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

The Design of Supply Networks under Maximum Customer Order Lead Times

Frank Meisel
Walter Rei
Michel Gendreau
Christian Bierwirth

September 2011

CIRRELT-2011-57

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Université Laval · UQÀM Université du Québec à Montréal · HEC Montréal · École Polytechnique Montréal · Université de Montréal

The Design of Supply Networks under Maximum Customer Order Lead Times

Frank Meisel[1,*], Walter Rei[2,3], Michel Gendreau[2,4], Christian Bierwirth[5]

[1] Department of Logistics Mangement, University of Bremen, 28359 Bremen, Germany

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[3] Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

[4] Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, P.O. Box 6079, Station Centre-ville, Montréal, Canada H3C 3A7

[5] School of Economics and Business, Martin-Luther-University Halle-Wittenberg, 06108 Halle, Germany

Abstract. We consider the problem of designing facility networks for producing and distributing goods under restricted customer order lead times. Companies apply various instruments for fulfilling orders within preset lead times, like locating facilities close to markets, producing products to stock, choosing fast modes of transportation, or delivering products directly from plants to customers without the use of distribution centers. We provide two alternative mathematical models that consider these options when designing a multi-layer multi-product facility network under restricted order lead times. A computational study compares the solvability of both models and investigates the impact of lead time restrictions on network structure and cost. From these results, it is identified how the different instruments can be combined for meeting lead times requirements within production and distribution networks.

Keywords. Production and distribution network, facility location, bill of materials, make-to-order, make-to-stock, customer order lead time.

Acknowledgements. This work was supported by a fellowship within the Postdoc-Program D/09/49358 of the German Academic Exchange Service (DAAD). The authors would also like to gratefully acknowledge the ongoing support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC).

---

* Corresponding author: frank.meisel@uni-bremen.de

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2011

1 Introduction

Today's competitive markets force companies to spend considerable efforts and resources on designing efficient production and distribution networks. In such multi-layer networks, suppliers, manufacturers, retailers and carriers collaborate in order to transform raw materials into intermediate and final products which are then distributed to customers. Such networks often spread out globally in order to access raw material sources in countries where labor costs are low and then, through intermodal transport, supply customer markets. Overall, a company gains an important competitive edge if it is able to design and then use a supply network that satisfies the needs of its customers by providing high quality products at low prices and in short order lead times.

The design and configuration of a production and distribution network is a complex problem, which is comprised of both strategic and tactical decisions. Such decisions may include: the selection of suppliers, the opening of plants and distribution centers, the allocation of products and production processes to the plants, the selection of transportation modes, the allocation of customers to distribution centers, and the determination of commodity flows within the network. Such decisions are usually taken while considering the objective of minimizing the costs of both designing and using the network. However, focusing solely on cost minimization may cause an unacceptable deterioration in service quality. The introduction of service level constraints in the context of network design has therefore become the strategy of choice whenever service quality requirements for customers are paramount. Traditionally, service quality constraints relate to demand volume. In the context of deterministic models, it is typically required that the complete demand of customers in a given period be fulfilled, see for example Cordeau et al (2006). As for stochastic models, in the case of random customer demands, a β-service level is typically used (value β defining here the minimum portion of a customer's demand that is fulfilled in a given period), see for example Lin et al (2006).

When considering service quality measures to be applied in the context of network design, the order lead time offered to customers is a much less investigated approach. The lead time for an order is typically defined as the time span separating the moment when the order is accepted and the moment when it is fulfilled. If customers are willing to wait for a certain amount of time until a product is delivered, network design and operations must be aligned carefully in such a way that the execution of production and logistics operations does not exceed this tolerance, Holweg (2005). When trying to meet the lead time expectations for their customers, companies employ various strategies. Such strategies may include the use of both raw material suppliers and producing plants that are closer to customer markets. By doing so, a company reduces the transportation distances traveled by orders which, in turn, may translate into significant reductions in transportation times. By decoupling the upstream parts of its network that produce in a make-to-stock (MTS) fashion from downstream parts that produce in a make-to-order (MTO) fashion, a company can also reduce the overall delivery times to its customers. In such a case, the processing times of upstream operations no longer affect the order lead times. Another possible strategy to shorten delivery times is to bypass distribution centers altogether through direct deliveries from plants to customers. Finally, if different modes of transportation are available for shipping goods, a company can select the mode that balances cost and transportation time to the desired level.

In this paper, we propose to use instruments for controlling lead times in the design of combined production and distribution networks. We consider multi-stage facility networks, complex product structures, and multiple modes of transportation. Two optimization models are proposed for the design of networks that guarantee a fulfillment of customer orders within a prescribed maximum time span. The first model is based on an extension of the one provided by Cordeau et al (2006), where we incorporate lead time drivers and lead time restrictions together with complex multi-stage production processes. We provide a second model defined on an alternative problem formulation, which is based on a time-space network. This second model is shown to be significantly more efficient when considering the computational effort needed to solve it. In both cases, valid inequalities are provided to strengthen the formulations obtained. Optimal network designs are produced using ILOG Cplex and a thorough computational study is conducted to explore the impact of lead time restrictions on the network structures obtained.

2 Literature

The design and configuration of supply networks has been approached in different ways, e.g., from the perspective of facility location, supplier selection, production and distribution management, information management, technology management, see the reviews provided by Erengüç et al (1999), Kreipl and Pinedo (2004), and Daskin et al (2005). Various case studies illustrate the application of the proposed approaches within diverse industries like the chemical industry (You and Grossmann, 2008) and computer manufacturing (Graves and Willems, 2005).

In the case of facility location in network design, the current state of research is surveyed by Klose and Drexl (2005), ReVelle et al (2008), and Melo et al (2009). Beginning with seminal studies such as Balinski (1961) and Geoffrion and Graves (1974), a continuous trend has been observed in the literature towards enriching network design models. Such enrichments have usually entailed the compounding of different decisions that have to be made in such planning contexts in order to develop more integrated models. Production related decisions concerning the capacity of opened facilities, the technologies applied there, and the quantities to produce are explicitly considered in Verter and Dasci (2002), and Melo et al (2006). In Minner (2003), inventory management is incorporated in the case where customer demands are stochastic. Moreover, when designing networks for globally acting companies, extending the models to include such concepts as duties and exchange rates becomes critical, see, e.g., Goetschalckx et al (2002) and Meixell and Gargeya (2005). When considering the objectives used in network design problems, the recent survey of Melo et al (2009) shows that most studies investigate models that pursue cost objectives such as the minimization of the fixed costs for opening facilities and the variable costs for the production and transportation of goods, see, for example, Cordeau et al (2006).

Lead times are an issue that is either ignored in most studies on network design, or lead times define unchangeable problem parameters that are not affected by the decisions made. When lead times are given in the planning context, they are either used to model the time lags of material flows in multi-period models (see Robinson and Bookbinder, 2007) or for determining the parameters of an inventory policy (as in Sabri and Beamon, 2000, and Sourirajan et al, 2007). Only a very limited number of studies explicitly consider the impact of the decisions made in network design on the achieved customer order lead times. The early approach of Arntzen et al (1995) considers minimizing the total time needed by all production and transportation processes that take place anywhere in a network when deciding on facility locations. However, minimizing the total time components does not necessarily lead to acceptable order lead times for all customers. Recent models tend to measure customer order lead times individually and then bound the values using constraints. Graves and Willems (2005) propose to configure a given network of production stages for a single product by solving a selection problem where, through the choice of suppliers, parts, processes, and transportation modes, the order lead times of customers are fixed. In Cheong et al (2005), distribution networks are designed where the lead time of orders is defined through the selection of servicing the demand either using a central warehouse, or, through local warehouses. In the considered model, the achieved lead times, in turn, impact the customer demands. Hence, minimizing lost sales caused by long lead times is one of the objectives considered. Paquet et al (2006) design a network of production plants by determining facility locations, process allocations, and the positioning of decoupling points. A decoupling point separates a supply network into an upstream part, which operates in an MTS fashion, and a downstream part, which operates in an MTO fashion. In this way, a better tradeoff can be achieved between keeping the inventory holding costs under control and increasing the responsiveness of the network to customer demands. Kohler (2008) investigates the design of global supply chains that also involves local-content requirements, duty constraints, transportation modes, and maximum order lead times. In this case, the networks obtained are completely order driven, i.e., reducing lead times by making goods to stock is not a considered option. You and Grossmann (2008) measure the lead time needed throughout the production stages of a network to replenish local retailers. This lead time is then multiplied by the stock-out probability associated with retailers to determine the expected order lead time of customers. Also in this study, decoupling of MTS and MTO production operations for controlling lead times is not considered.

3 Characterization of the Production-Distribution System

We consider the problem of designing a facility network that produces and distributes multiple types of products for customers with known demand under restricted order lead times. An overview of the parameters that define the considered problem is provided in Table 1. The following assumptions are made in this study:

- The total customer demands for final products are deterministic and given. Customers order the final products in separate orders of identical size (i.e., $\overline{q}$ product units per order).
- Production processing times and transportation times are deterministic and given. The effects of congestion, which may influence the production processing times at the facilities, are not considered.
- Facilities are either restricted to produce a commodity exclusively under MTS or MTO fashions, or they are allowed to use both production fashions. Companies that adopt the second policy gain more flexibility. We therefore develop models for both policies.
- Inventory levels are derived at a tactical planning level according to the flows of commodities that result from the network design. The determination of inventory levels is therefore excluded from the strategic network design problem investigated here. We assume that estimated cost rates are given for MTO and MTS production, which appropriately reflect the cost effects of both strategies.

In the following, we describe the general network structure that is considered for the problem as well as the formulation used to define the order lead times in this paper.

3.1 Products, Facilities, and Transportation

The production system considered is capable of producing different types of final products by transforming raw materials and intermediate products according to specified processes. We denote by R, I, and F the sets of raw materials, intermediate products, and final products, respectively. For convenience, we define $K = R \cup I \cup F$ as the set of all commodities considered in the problem. Deterministic production coefficients $a^{k,k'}$ describe the amount of commodity $k \in K$ used for producing one unit of commodity $k' \in K$. The composition of the final products is described through a multi-product bill of material (MPBOM), which is represented by a directed graph as illustrated in the example in Fig. 1a. It should be noted that the network design problem considered here allows the use of arbitrary acyclic MPBOMs. Let I^k and F^k denote respectively the sets of intermediate products and final products that directly require commodity $k \in K$. In Fig. 1a, we observe $I^{\text{R1}} = \{\text{I1}, \text{I2}\}$ and $F^{\text{R1}} = \{\text{F1}\}$. To indicate the different roles that a commodity can take within the production process, we make use of a tree graph representation of the MPBOM as shown in Fig. 1b. In this graph, the label attached to a vertex defines the amount of the commodity that is required within the corresponding branch of the tree to produce one unit of the final product. For instance, to make one unit of F2, two units of I2 are required, which requires six units of I1, creating a material demand of 30 units of R1 at the lowest production stage.

Table 1: Parameters of the network design problem.

Commodities:	
R	Set of raw materials
I	Set of intermediate products, $I^k \subseteq I$ denotes the set of intermediate products that directly require commodity k
F	Set of finished products, $F^k \subseteq F$ denotes the set of finished products that directly require commodity k
K	Set of commodities, $K = R \cup I \cup F$
$a^{k,k'}$	Amount of commodity k contained in one unit of commodity k'
Facilities and customers:	
S	Set of suppliers, $S^r \subseteq S$ denotes the set of potential suppliers providing raw material r
P	Set of plant locations, $P^k \subseteq P$ denotes the set of potential plants for making commodity k
W	Set of warehouse locations, $W^f \subseteq W$ denotes the set of potential warehouses for distributing product f
C	Set of customers, $C^f \subseteq C$ denotes the set of customers requesting final product f
O	Set of origins, $O = S \cup P \cup W$, O^k denotes the set of potential origins of commodity k
D	Set of destinations, $D = P \cup W \cup C$, D^k denotes the set of potential destinations for commodity k
K_o	Set of commodities that can be produced or offered at origin o
$\overline{q}$	Default size of orders placed by customers
d_c^f	Demand of customer c regarding product f
Capacities:	
u_o	Overall production capacity of origin o in equivalent units
u^k	Amount of production capacity required by one unit of commodity k
q_o^k	Upper limit on the amount of commodity k to be produced at origin o
q_{od}^k	Upper limit on the amount of commodity k shipped from o to d
g_{od}^m	Capacity of transportation mode $m \in M_{od}$ between origin o and destination d
g^{km}	Amount of capacity required by one unit of commodity k in mode m
Transportation and lead times:	
M_{od}	Set of transportation modes between origin o and destination d
tt_{od}^m	Transportation time needed for a shipment from o to d on transportation mode $m \in M_{od}$
pt_o^{kq}	Processing time needed to produce q units of commodity k at origin o
Q^k	Set of material demand quantities of commodity k that are of relevance for the determination of lead times
τ_c^f	Maximum lead time promised to customer c for fulfilling an order of size of $\overline{q}$ units of product f
Cost rates:	
c_o^{fix}	Fixed cost for opening origin o
c_{ok}^{fix}	Fixed cost for producing or offering commodity k at origin o
c_{kod}^{fix}	Fixed cost for shipping commodity k from o to d
c_{mod}^{fix}	Fixed cost for using transportation mode m between o to d
c_{ok}^{MTO}	MTO unit production cost of commodity k at origin o
c_{ok}^{MTS}	MTS unit production and inventory holding cost of k at o
c_{kmod}	Unit cost for shipping k from o to d on transportation mode m
M	a large positive value

Three sets of facilities are defined, namely suppliers S, plants P, and warehouses W. Suppliers provide raw materials that are transformed at the plants into intermediate and final products. The final products are then shipped to customers, represented by set C, either via a warehouse or via a direct delivery from a plant to a customer.

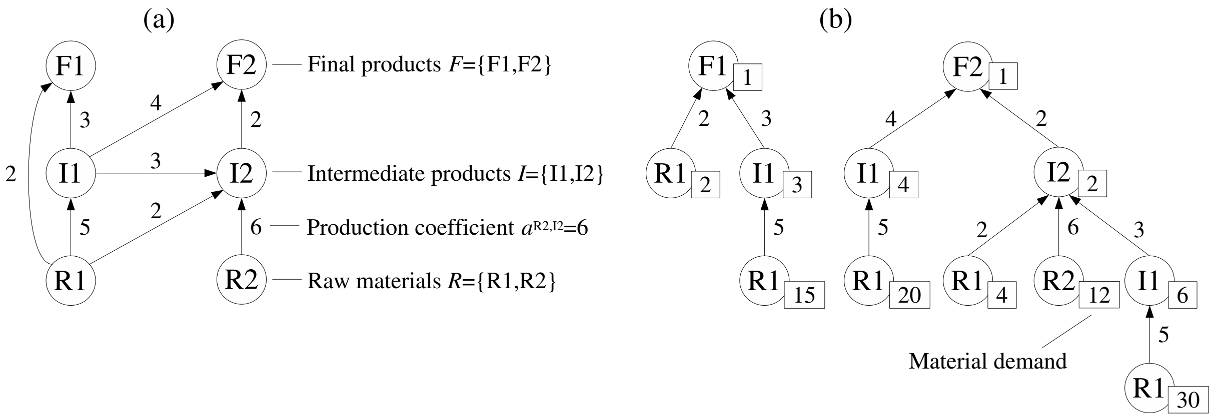

Figure 1: Graph representation of a multi-product bill of material.

Figure 2a illustrates an example based on a network which includes three suppliers, two plants, two warehouses, and five customers. Since not every facility is able to produce every commodity, let K_s, K_p, and K_w denote respectively the set of those materials that are offered by supplier $s \in S$, the commodities producible at plant $p \in P$, and the final products that can be handled at warehouse $w \in W$. These sets are shown in Fig. 2a as labels attached to the different facility vertices. Accordingly, we define $S^r = \{s \in S \mid r \in K_s\}$, $P^k = \{p \in P \mid k \in K_p\}$, and $W^f = \{w \in W \mid f \in K_w\}$ as the subset of suppliers, plants, and warehouses that can provide raw material $r \in R$, commodity $k \in K$, and final product $f \in F$, respectively. Moreover, for notational convenience, let $O = S \cup P \cup W$ denote the set of all potential origins of commodity flows in the network and let $O^k = \{o \in O \mid k \in K_o\}$ denote the potential origins of commodity $k \in K$. Correspondingly, let $D = P \cup W \cup C$ denote the set of potential destinations of commodity flows and let $D^k \subset D$ denote the subset of potential destinations for commodity $k \in K$. For raw materials and intermediate products (i.e., $k \in R \cup I$), let $D^k = \{p \in P \mid (I^k \cup F^k) \cap K_p \neq \emptyset\}$ define the set of those plants that can produce intermediate or final products that require k. For final products $f \in F$, let $C^f = \{c \in C \mid d_c^f > 0\}$ be the set of those customers $c \in C$ that show a non-zero demand d_c^f for product f. Therefore, for $f \in F$, let $D^f = W^f \cup C^f$ denote the set of warehouses distributing product f and customers demanding it.

We consider the following capacity restrictions at facilities. The overall production capacity available at origin $o \in O$ is denoted by u_o and value u^k is defined as the capacity requirement for producing one unit of commodity $k \in K$. Moreover, the producible amount of a commodity k at an origin o is bounded to a commodity-specific upper limit defined as value q_o^k. With regards to transportation activities, an upper limit on the amount of k to be shipped from the origin o to the destination $d \in D$ is given by q_{od}^k.

The commodity flow within the network takes place on transport relations that link facilities with each other and to the customers. We consider relations connecting suppliers to plants, plants to warehouses, and warehouses to customers. Furthermore, we allow transport relations between two plants for the exchange of intermediate

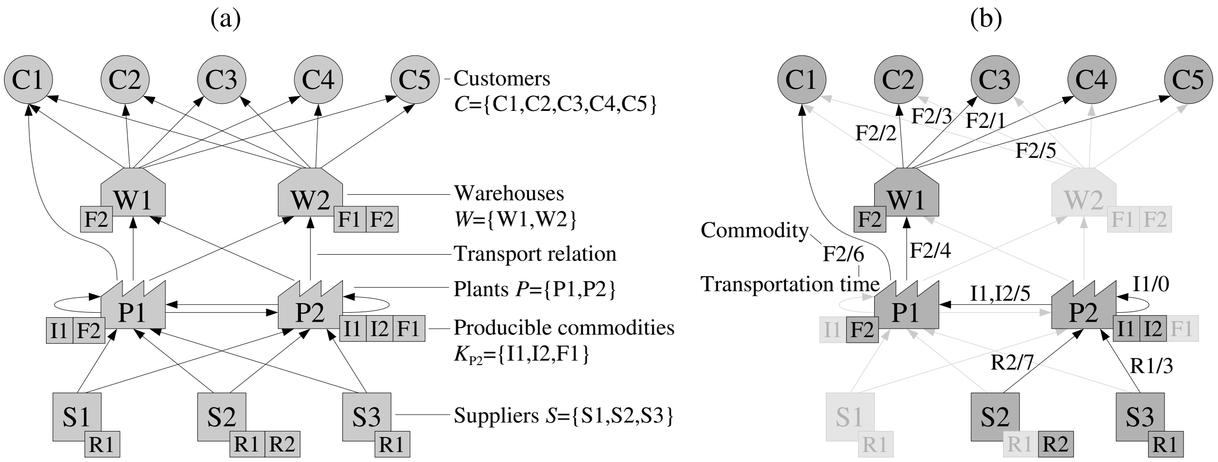

Figure 2: Supply network structure.

products, relations cyclically connecting a plant to itself for the supply of self-made intermediate products, and relations connecting a plant to a customer for direct deliveries of products, see the example of arc (P1,C1) in Fig. 2a. Figure 2b highlights a potential network structure that could be used for the production and distribution of product F2 defined in Fig. 1b. For every transport relation between an origin $o \in O$ and a destination $d \in D$, a set M_{od} of transportations modes are available. Modes $m \in M_{od}$ differ in terms of transportation times t_{od}^m, fixed costs c_{mod}^{fix} (incurred for using mode m), variable cost c_{od}^{km} (incurred for transporting one unit of k using m), and transportation capacities g_{od}^m. Finally, g^{km} defines the amount of capacity required for shipping one unit of commodity k using mode m.

3.2 Determination of Lead Times

For the determination of order lead times, we assume in this study that all customer orders are of identical size of $\overline{q}$ product units. For instance, for $\overline{q} = 1$, all customers follow a one-by-one replenishment strategy for satisfying their demand. An order for a final product $f \in F$ placed by customer $c \in C$ must be fulfilled within a maximum order lead time of τ_c^f time units. In the present paper, it is considered that the lead time of an order is determined through the production processing times at the facilities that perform the operations involved in producing product f and through the transportation time of the used transportation modes. The effects that congestion at the facilities and consolidation of multiple orders within shipments may have on lead times are not explicitly considered in the present study. We denote by p_o^{kq} the known and deterministic processing time needed for producing q units of commodity k at origin o and by t_{od}^m the transportation time for sending a shipment from origin o to destination d using transportation mode m. Therefore, the outbound lead time of a production step is determined as the sum of the maximum replenishment time over all ingoing commodities plus the step's processing time. The replenishment

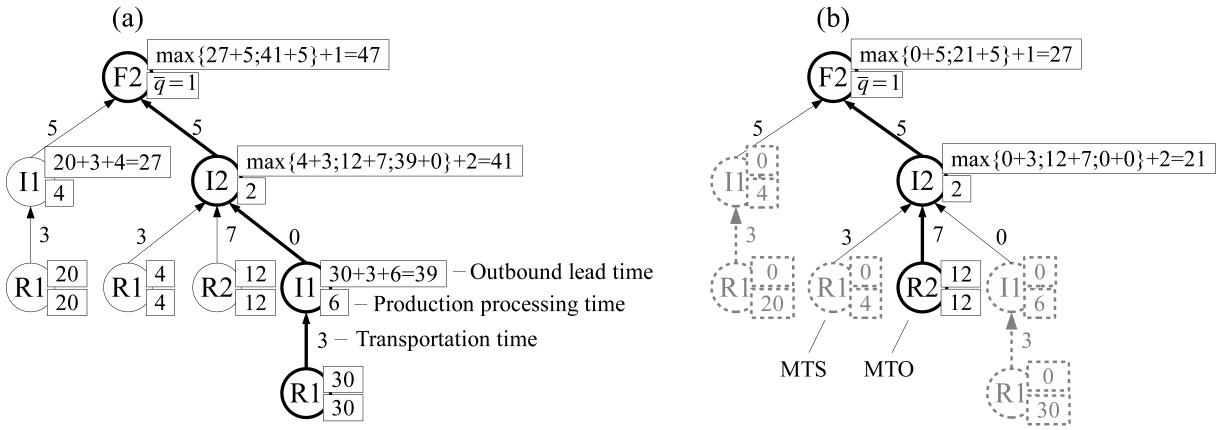

Figure 3: Calculation of lead times.

time (or inbound lead time) for a commodity at a facility is determined as the outbound lead time of the supplying facility that provides the requested commodity plus the transportation time for delivering it to the receiving facility.

Figure 3a illustrates the computation of the lead times for the following example: $\overline{q} = 1$ units of product F2 produced using the designed network highlighted in Fig. 2b. The figure reproduces the tree structure of F2, where the label attached at the bottom of a vertex now represents the production processing time at the selected facility and the arc weights represent transportation times between the selected facilities. For reasons of simplicity, we assume for this example that the processing time for producing q units of k at facility o equals the amount of k to produce, i.e we set $p_o^{kq} = q$ time units. The outbound lead times of the production steps are computed within the labels attached at the top of the vertices. The lead time for producing $\overline{q}$ units of the final product then corresponds to the longest path in terms of total production and transportation time found in this graph. This path is depicted in bold in Fig. 3a, indicating that one unit of F2 can be provided at plant P1 within 47 time units under the network design of Fig. 2b. The order lead time of a customer can then be obtained by adding the delivery time. In the case of a direct delivery, the transportation time for shipping the order from the plant to the customer is added. If an order is fulfilled via a warehouse, the delivery time is obtained by summing the time needed for shipping the product to the warehouse, the time needed to process the order at the warehouse, and finally the transportation time between the warehouse and the customer. In both cases, the resulting order lead time is not allowed to exceed the maximum lead time (i.e., value τ_c^f).

The previous example shows how the production coefficients of the MPBOM impact the order lead time through the induced material demands at the different stages of the production process. In turn, it is required to determine the lead time for every demand quantity appearing for a commodity in the tree graph. As illustrated in Fig. 3a, lead times have to be calculated for 4 and 6 units of intermediate product I1. If we additionally consider product F1 of Fig. 1b, then the lead time for 3 units of I1 is of relevance for the network design too. Therefore,

Table 2: Decision variables in the network design problem.

Variable	Description
X_{od}^{km}	Amount of commodity k shipped from o to d on transportation mode m
A_{ok}^{MTO}	Amount of commodity k produced or offered at origin o under MTO
A_{ok}^{MTS}	Amount of commodity k produced or offered at origin o under MTS
T_o^{kq}	Lead time for providing $q \in Q^k$ units of commodity k at origin o
T_c^f	Lead time for fulfilling an order of final product f for customer $c \in C^f$
U_o	= 1 if origin o is selected
V_o^k	= 1 if commodity k is assigned to origin o, i.e. o produces or distributes k
Y_{od}^k	= 1 if origin o provides commodity k to destination d
Z_{od}^m	= 1 if transportation mode m is opened between o and d
Z_{od}^{km}	= 1 if commodity k is shipped from o to d on transportation mode m
B_o^k	= 1 if origin o produces commodity k under MTO, i.e. $A_{ok}^{\text{MTO}} > 0$

for determining the relevant lead times in the network design model of the next section, we introduce Q^k as the set of those *material demand quantities* of commodity k that appear in the MPBOM of a problem. For example, the MPBOM of Fig. 1b yields $Q^{\text{I1}} = \{3,4,6\}$ and $Q^{\text{R1}} = \{2,4,15,20,30\}$. For $\overline{q} > 1$, the same process is used for calculating lead times, where material demands at the production stages are simply adapted to the specified order size.

One further point that affects lead times in a supply network is the use of MTS or MTO production fashions. MTS is used to reduce lead times by producing to stock such that commodities are immediately available when required. Deciding on the decoupling point where production switches from MTS to MTO in a network is an important issue when trying to balance the lead times of MTO processes with the higher cost of MTS production caused by inventory holding, cf. Paquet et al (2006). Therefore, we explicitly incorporate this issue into our model by selecting the production strategy for every commodity produced at a facility. As such, in the previous example, if R1 and I1 are produced under MTS, the lead time for providing one unit of F2 at plant P1 decreases from 47 to 27 time units, see Fig. 3b.

4 Modeling the Network Design Problem

4.1 Model Formulation

In the network design problem considered here, decisions to be taken define the structure of the network that is designed and they are related to how the created network is used. Decisions defining the structure of the network include: the selection of suppliers, plants and warehouses to be included in the network, and the selection of both transportation relations and transportation modes for linking facilities. As for decisions related to the use of the network, we consider: the assignments of commodities to facilities, the quantities to produce of each commodity under either MTO or MTS at each facility, and the flow of goods on the chosen transportation modes between facilities. Given that each of these decisions entail a specific cost, the general objective of the problem is to minimize the total cost of both designing and using the network while meeting customer demands for final products in such a way as to fulfill orders within promised lead times. Cost rates are clearly stated in the lower part of Table 1. As for the decision variables, they are summarized in Table 2. The Supply Network Design (SND) model can now be defined as follows:

$$
\text{[SND]} \qquad \min \to Z = \sum_{o \in O} c_o^{\text{fix}} U_o + \sum_{k \in K} \sum_{o \in O^k} \left(c_{ok}^{\text{fix}} V_o^k + c_{ok}^{\text{MTO}} A_{ok}^{\text{MTO}} + c_{ok}^{\text{MTS}} A_{ok}^{\text{MTS}} \right) + \sum_{k \in K} \sum_{o \in O^k} \sum_{d \in D^k} \left(c_{kod}^{\text{fix}} Y_{od}^k + \sum_{m \in M_{od}} c_{od}^{km} X_{od}^{km} \right) + \sum_{o \in O} \sum_{d \in D} \sum_{m \in M_{od}} c_{mod}^{\text{fix}} Z_{od}^m \tag{1}
$$

$$
\sum_{o \in O^k} \sum_{m \in M_{op}} X_{op}^{km} - \sum_{k' \in (I^k \cup F^k) \cap K_p} \sum_{d \in D^{k'}} \sum_{m \in M_{pd}} a^{k,k'} X_{pd}^{k',m} = 0 \qquad k \in K; p \in P \cap D^k \tag{2}
$$

$$
\sum_{p \in P^f} \sum_{m \in M_{pw}} X_{pw}^{fm} - \sum_{c \in C^f} \sum_{m \in M_{wc}} X_{wc}^{fm} = 0 \qquad f \in F; w \in W^f \tag{3}
$$

$$
\sum_{p \in P^f \cup W^f} \sum_{m \in M_{pc}} X_{pc}^{fm} = d_c^f \qquad f \in F; c \in C^f \tag{4}
$$

$$
\sum_{k \in K_o} \sum_{d \in D^k} \sum_{m \in M_{od}} u^k \cdot X_{od}^{km} - u_o \cdot U_o \leq 0 \qquad o \in O \tag{5}
$$

$$
\sum_{d \in D^k} \sum_{m \in M_{od}} X_{od}^{km} - q_o^k \cdot V_o^k \leq 0 \qquad k \in K; o \in O^k \tag{6}
$$

$$
\sum_{m \in M_{od}} X_{od}^{km} - q_{od}^k \cdot Y_{od}^k \leq 0 \qquad k \in K; o \in O^k; d \in D^k \tag{7}
$$

$$
\sum_{k \in K_o | d \in D^k} g^{km} \cdot X_{od}^{km} - g_{od}^m \cdot Z_{od}^m \leq 0 \qquad o \in O; d \in D; m \in M_{od} \tag{8}
$$

$$
g^{km} \cdot X_{od}^{km} - g_{od}^m \cdot Z_{od}^{km} \leq 0 \qquad k \in K; o \in O^k; d \in D^k; m \in M_{od} \tag{9}
$$

$$
B_o^k + Y_{op}^k + V_p^{k'} - B_p^{k'} \leq 2 \qquad k \in R \cup I; k' \in I^k \cup F^k; o \in O^k; p \in P^{k'} \tag{10}
$$

$$
B_p^f + Y_{pw}^f + V_w^f - B_w^f \leq 2 \qquad f \in F; p \in P^f; w \in W^f \tag{11}
$$

$$
\sum_{d \in D^k} \sum_{m \in M_{od}} X_{od}^{km} - A_{ok}^{\text{MTO}} - q_o^k \cdot (1 - B_o^k) \leq 0 \qquad k \in K; o \in O^k \tag{12}
$$

$$
\sum_{d \in D^k} \sum_{m \in M_{od}} X_{od}^{km} - A_{ok}^{\text{MTS}} - q_o^k \cdot B_o^k \leq 0 \qquad k \in K; o \in O^k \tag{13}
$$

$$
T_s^{rq} - p_s^{rq} \cdot B_s^r = 0 \qquad r \in R; s \in S^r; q \in Q^r \tag{14}
$$

$$
T_o^{k,q \cdot a^{k,k'}} + \sum_{m \in M_{op}} Z_{op}^{km} \cdot t_{op}^m + p_p^{k',q} - \mathrm{M} \cdot \left(2 - Y_{op}^k - B_p^{k'} \right) \leq T_p^{k'q} \qquad k \in R \cup I; k' \in I^k \cup F^k; o \in O^k; p \in P^{k'}; q \in Q^{k'} \tag{15}
$$

$$
T_p^{f,\overline{q}} + \sum_{m \in M_{pw}} Z_{pw}^{fm} \cdot t_{pw}^m + p_w^{f,\overline{q}} - \mathrm{M} \cdot \left(2 - Y_{pw}^f - B_w^f \right) \leq T_w^{f,\overline{q}} \qquad f \in F; p \in P^f; w \in W^f \tag{16}
$$

$$
T_o^{f,\overline{q}} + \sum_{m \in M_{oc}} Z_{oc}^{fm} \cdot t_{oc}^m - \mathrm{M} \cdot \left(1 - Y_{oc}^f \right) \leq \tau_c^f \qquad f \in F; c \in C^f; o \in O^f \tag{17}
$$

$$X_{od}^{km}, A_{ok}^{\text{MTO}}, A_{ok}^{\text{MTS}}, T_o^{kq}, T_c^f \geq 0 \qquad k \in K; o \in O^k; d \in D^k; m \in M_{od}; q \in Q^k; f \in F; c \in C^f \tag{18}$$

$$U_o, Z_{od}^m \in \{0,1\} \qquad o \in O; d \in D; m \in M_{od} \tag{19}$$

$$V_o^k, B_o^k, Y_{od}^k, Z_{od}^{km} \in \{0,1\} \qquad k \in K; o \in O^k; d \in D^k; m \in M_{od} \tag{20}$$

The objective function (1) minimizes the sum of the fixed costs incurred for opening facilities, assigning commodities, establishing supply relations and using transportation modes, as well as the variable costs for producing commodities under MTO and MTS and for shipping them from origins to destinations. Constraints (2) impose flow conservation at plants by considering every combination pair of commodity k and plant p (which can process commodity k). These constraints ensure that the inflow of commodity k at p equals the total outflow of intermediate and final products $k' \in (I^k \cup F^k) \cap K_p$ that directly require k and that can be produced at plant p. Equivalent units are obtained through the use of the production coefficients $a^{k,k'}$. The inflow and outflow of final products at warehouses are balanced in Constraints (3). Constraints (4) ensure that each customers' demand for final products is met, be it through direct delivery from a plant or through indirect delivery via warehouses.

Constraints (5) to (9) ensure a consistent use of the network with respect to the flow of goods and the available production and transportation capacities. From (5), facility o must be opened if there is an outflow of commodities and (6) assigns commodity k to origin o if o sends this commodity to some destination. Constraints (7) ensure that a supply relationship is established if origin o sends k to destination d using any transportation mode. Constraints (8) impose that a transportation mode is activated if there is flow assigned to it and (9) set those variables that indicate whether commodity k is transported on mode m between o and d.

Constraints (10) and (11) separate the upstream parts from the downstream parts of the network through the selection of MTO or MTS production strategies for commodities at facilities. If production runs under MTO at a facility, subsequent production at its downstream facilities must also run under MTO. Hence, from (10), if origin o produces k under MTO and provides it to plant p for the production of k' (i.e., $B_o^k = Y_{op}^k = V_p^{k'} = 1$), then k' must also be produced under MTO (i.e., $B_p^{k'} = 1$). Similarly, warehouses distribute products under MTO, if the supplying plant already produced it under this policy. In this situation, modeled in (11), a warehouse serves as a cross-dock distribution center (i.e., it does not keep the product on stock). Constraints (12) and (13) set the production quantities A_{ok}^{MTO} and A_{ok}^{MTS} for commodity k at origin o with respect to the selected production strategy and the amount of k projected to be sent from o to other destinations in the network.

Lead time restrictions are imposed by constraints (14) to (17). Constraints (14) determine the outbound lead time T_s^{rq} for providing q units of material r at a supplier s. If s produces r under MTO then the lead time equals the production processing time p_s^{rq}, otherwise, it is simply fixed to 0. Constraints (15) determine the lead time $T_p^{k'q}$ for providing q units of an intermediate or final product k' at plant p. These constraints consider pairs of commodities k and k' where k is required to produce k', together with potential origins o of k and plants p that can produce k'. A valid lead time is computed, if o supplies p with commodity k $\left(\text{i.e., } Y_{op}^{k} = 1\right)$ and if p produces k' under MTO (i.e., $B_p^{k'} = 1$). If both conditions hold, the lead time for providing q units of k' at p is obtained by summing: the lead time for providing $q \cdot a^{k,k'}$ units of k at the supplying facility o, the transportation time of the selected transportation mode, and the production processing time $p_p^{k'q}$ at facility p. Accordingly, (16) defines the lead time for providing final products at warehouses. These lead times are computed for the default order size of $\overline{q}$ product units requested by customers. Again, lead times are only considered in the case of MTO-operations, where $p_w^{f,\overline{q}}$ refers to the time for processing the order and handling the goods at the warehouse. Constraints (17) bound the customer order lead time for each product f and each customer c who requests the product to the maximum lead time τ_c^f. In turn, these constraints ensure that the configuration of the network and the decisions related to production and transportation operations meet the service requirements. Finally, non-negativity and integrality restrictions on all defined variables are imposed by (18) to (20).

4.2 Valid Inequalities

The SND model defined in (1)-(20) is a linear mixed integer problem. To solve this problem, we employ the general-purpose solver ILOG Cplex, which has been applied successfully in various studies on facility location planning, cf. Melo et al (2009). The performance of this simplex-based mixed integer solver can be improved through the inclusion of valid inequalities aimed at strengthening the linear relaxation of the model being solved. To strengthen the setting of binary variables, the following inequalities, originally proposed in the context of network design problems by Cordeau et al (2006), can be added to the SND model:

$$U_o \geq V_o^k \quad k \in K; o \in O^k \tag{21}$$

$$\sum_{p \in P^f \cup W^f} Y_{pc}^f \geq 1 \quad f \in F; c \in C^f \tag{22}$$

$$\sum_{p \in P^f} Y_{pw}^f \geq V_w^f \quad f \in F; w \in W^f \tag{23}$$

$$\sum_{o \in O^k} Y_{op}^k \geq V_p^{k'} \quad k \in R \cup I; k' \in I^k \cup F^k, p \in P^{k'} \tag{24}$$

$$\sum_{m \in M_{od}} Z_{od}^m \geq Y_{od}^k \quad k \in K; o \in O^k; d \in D^k \tag{25}$$

$$V_o^k \geq Y_{od}^k \quad k \in K; o \in O^k; d \in D^k \tag{26}$$

Inequalities (21) ensure that a facility is opened if a commodity is assigned to it. Constraints (22) impose that at least one supply relation is established for each customer and requested final product. Given that a warehouse distributing a final product must be linked to a plant producing this product then (23) are valid. Constraints (24) formulate the fact that plants producing a commodity k' using a commodity k must be supplied by at least one origin of k. By imposing (25), at least one transportation mode is chosen on every established transport relation. Finally, through (26), a commodity k is assigned to origin o if o delivers this commodity to some destination d.

To further strengthen the formulation, we propose a series of new inequalities, which specifically target the MTS and MTO production decisions considered in the SND model. They are formulated as follows:

$$A_{ok}^{\mathrm{MTS}} + A_{ok}^{\mathrm{MTO}} - \sum_{d \in D^k} \sum_{m \in M_{od}} X_{od}^{km} = 0 \qquad k \in K; o \in O^k \tag{27}$$

$$\sum_{s \in S^r} A_{sr}^{\mathrm{MTS}} \geq MTS_r^{\min} \qquad r \in R \tag{28}$$

$$\sum_{k \in P^k} A_{pk}^{\mathrm{MTS}} \geq MTS_k^{\min} \qquad k \in I \cup F \tag{29}$$

In (27), the quantity of a commodity k produced at origin o is set equal to the total flow of k leaving o. Inequalities (28) and (29) utilize lower bounds $MTS_k^{\min}$ on the total quantities of commodities $k \in K$ to produce inevitably under the MTS strategy in a problem instance. $MTS_k^{\min}$ is determined by considering each combination of a k-vertex in the MPBOM, the final product f corresponding to this vertex, and a customer c who demands the final product. If, in every feasible network design for a problem instance, the shortest lead time needed for transforming k into f and for delivering f to c exceeds the accepted lead time τ_c^f, k must inevitably be produced under MTS in order to deliver f to c on time. In this case, at least $MTS_k^{\min} = d_c^f \cdot q$ units of k must be produced under MTS, where d_c^f is the demand of customer c for the final product f and q is the material demand quantity of the k-vertex in the MPBOM. The determination of the shortest possible lead time is described in an example in Appendix A. The process described there is repeated for every combination of a MPBOM-vertex and a customer in a problem instance such that $MTS_k^{\min}$ finally represents the total amount of k for which MTS production is inevitable. Inequalities (28) and (29) then ensure that the MTS-production quantities of raw materials at suppliers and of (intermediate) products at plants is at least $MTS_k^{\min}$ for any $k \in K$.

5 A Time-Space Network Representation

In model SND, lead time restrictions are imposed through the inclusion of a series of explicit constraints (i.e., constraint set (14) to (17)). In this section, we show that the considered network design problem can also be formulated as a fixed-charge flow problem in a time-space network. The time-space network combines the structure of the organizational facility network and the product structure of the MPBOM while reflecting the processing time of production and transportation operations. This network is generated in such a way as to enforce lead time restrictions

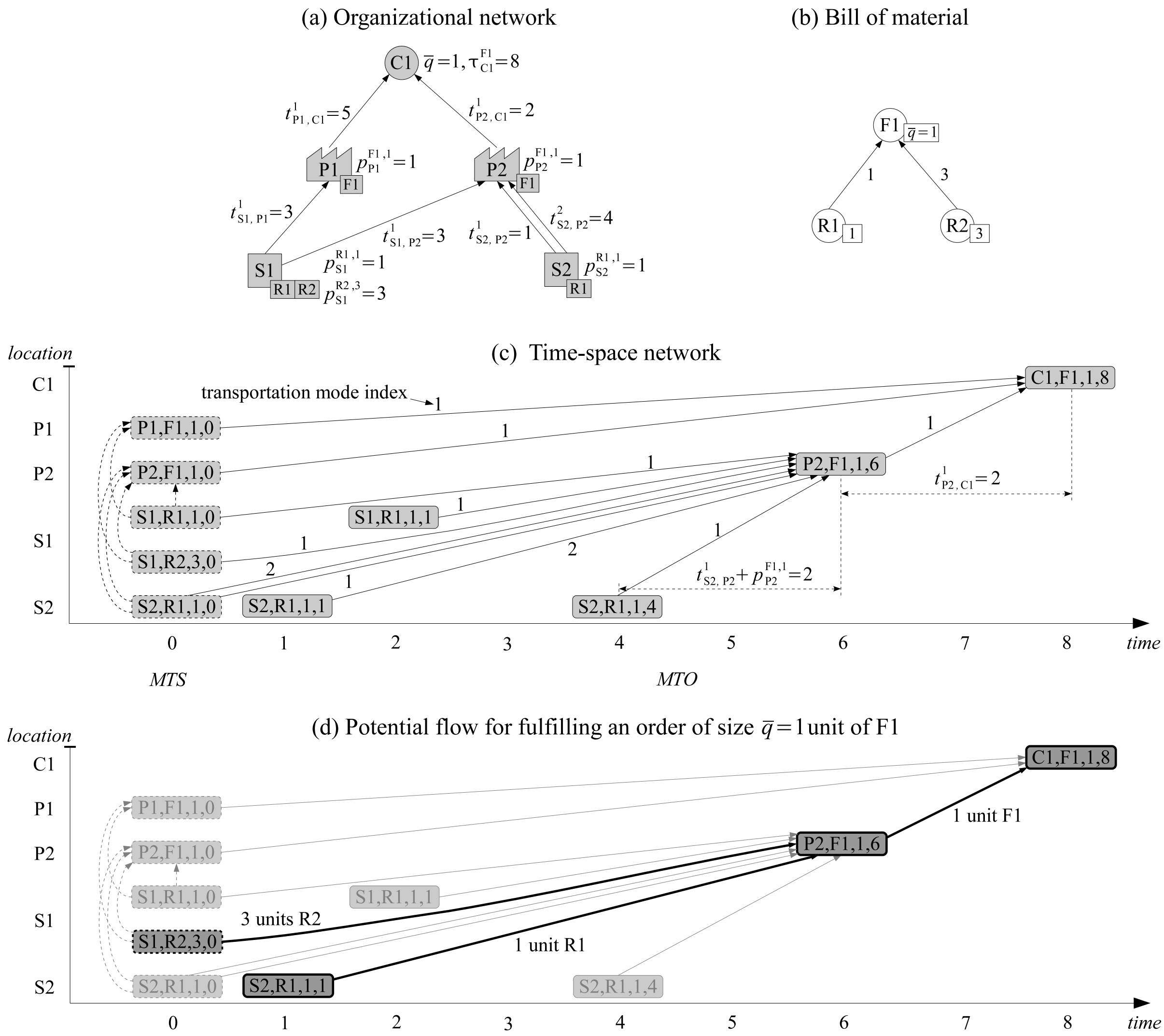

Figure 4: An example of time-space network generation.

by means of its structure. Therefore, the inclusion of explicit lead time constraints within the resulting optimization model is no longer necessary, which, in turn, is expected to ease the solution process. Figure 4 provides an example of a time-space network, which is used in the following for illustrating the generation process. For reasons of simplicity, but without loss of generality, the example does not contain warehouses or intermediate products. It considers a single final product F1 and one customer C1, who expects delivery within $\tau_{C1}^{F1} = 8$ time units.

The remaining of this section is divided as follows: in Subsection 5.1, we present the different procedures used to generate the time-space network, in Subsection 5.2, we reformulate the SND model on the time-space network, and in Subsection 5.3, we present a series of valid inequalities that can be added to the proposed model to strengthen the formulation.

5.1 Network Generation

The time-space network is defined as a directed acyclic graph $G = (V,A)$ with V being the set of vertices and $A \subseteq V \times V$ being the set of arcs. The general idea for constructing this time-space network is to first define vertices representing customers at the time coordinates of their maximum lead times. By applying a backtracking process, arcs are then included to account for those supply relationships that allow for on-time deliveries of the products. For each supply relationship thus added, a new vertex representing the supplying facility at the time coordinate corresponding to its outbound lead time is included to the vertex set. These general steps are repeated to account for all feasible supply relationships. Given the resulting time-space network, a flow problem can then be solved to select those facilities and services that will be used to perform all necessary operations to deliver products to clients within accepted lead times while minimizing overall costs.

More precisely, each vertex included in the network represents a certain quantity of a given commodity that can be provided at a particular facility within a certain time span. Hence, a vertex $i \in V$ is defined by the following quadruple: (loc_i, k_i, q_i, t_i), where loc_i denotes the facility that provides the commodity (the spatial dimension of the time-space network), $k_i \in K_{loc_i}$ denotes the type of the offered commodity, $q_i \in Q^{k_i}$ denotes the particular quantity considered, and t_i is the point in time when the q_i units of k_i are ready for shipment at loc_i (the temporal dimension of the network). For example, the vertex (P2, F1, 1, 6) in Fig. 4(c) indicates that plant P2 is able to provide one unit of final product F1 within a lead time of no more than 6 time units. Value t_i refers to the outbound lead time of vertex i, which is the latest point in time where q_i units of commodity k_i must be available for shipment at facility loc_i in order to ensure that the lead times of adjacent vertices remain feasible. Therefore, vertices with time value $t_i = 0$ have to provide k_i immediately, which means that the corresponding facility produces the commodity in MTS fashion and fulfills orders from stock. Vertices with time value $t_i > 0$ represent opportunities for MTO production. Each arc $a \in A$ in the network connects two vertices $i, j \in V$, defining a supply relationship, and is associated to a given transportation mode $m \in M_{loc_i,loc_j}$ that can be used to perform shipments between locations loc_i and loc_j. Hence, each arc is defined by a triple (i, j, m).

ADD_CUSTOMER_VERTICES()

$V \Leftarrow \emptyset; A \Leftarrow \emptyset;$
for each $f \in F, c \in C^f$ **do**
  $j \Leftarrow (c, f, \overline{q}, \tau_c^f); V \Leftarrow V \cup \{j\};$ ▷ Create customer-vertex
  for each $o \in P^f \cup W^f, m \in M_{oc}$ **do** ▷ For each potential supply relationship
    if $tt_{oc}^m \leq \tau_c^f$ **then** ADD_FACILITY_VERTEX($(o, f, \overline{q}, 0), j, m$); ▷ Take up MTS option
    if $tt_{oc}^m < \tau_c^f$ **then** ADD_FACILITY_VERTEX($(o, f, \overline{q}, \tau_c^f - tt_{oc}^m), j, m$); ▷ Test MTO option
  end for
end for

Figure 5: Procedure ADD_CUSTOMER_VERTICES.

Two procedures, ADD_CUSTOMER_VERTICES and ADD_FACILITY_VERTEX, are used to generate the vertices and arcs that are included in the time-space network, see Figs. 5 and 6. Procedure ADD_CUSTOMER_VERTICES begins by adding a vertex j, defined as $(loc_j, k_j, q_j, t_j) = (c, f, \overline{q}, \tau_c^f)$, to set V for every product $f \in F$ and every customer $c \in C^f$. Each of these vertices defines that a customer c is to receive an order for product f of standard order size $\overline{q}$ within no more than τ_c^f time units. For each of these customer-vertices, a series of warehouses and/or plants, which can deliver the products on time, are identified. Therefore, for each potential origin $o \in P^f \cup W^f$ and each transportation mode $m \in M_{oc}$ the transportation time t_{oc}^m is compared to the maximum lead time τ_c^f. If o produces f under MTS, then $t_{oc}^m \leq \tau_c^f$ is a sufficient condition for ensuring on-time delivery to c via transportation mode m. Hence, a vertex $(o, f, \overline{q}, 0)$ is added to the network and connected to the customer-vertex j via transportation mode m. This is done using procedure ADD_FACILITY_VERTEX(i, j, m), defined in Fig. 6, which adds a new vertex i to set V and connects it to an already existing vertex j through an arc that represents transportation mode $m \in M_{loc_i, loc_j}$. The procedure thus handles requests for adding vertices representing MTS-operations at either warehouses, plants or suppliers (referring to the conditions respectively tested on lines: 1, 15 and 29, within procedure ADD_FACILITY_VERTEX). If the supplying facility o is to produce f under MTO then a necessary condition to do so is that the transportation time to the customer be strictly lower than the maximum order lead time (i.e., $t_{oc}^m < \tau_c^f$), because o must additionally be replenished and perform its production operation within the available lead time. Therefore, procedure ADD_CUSTOMER_VERTICES calls ADD_FACILITY_VERTEX with the potential MTO-operation $i = (o, f, \overline{q}, \tau_c^f - t_{oc}^m)$. However, before adding the MTO-vertex i, it must be ensured that the remaining lead time $t_i = \tau_c^f - t_{oc}^m$ is actually sufficient to replenish facility loc_i with the required materials and to perform the production operation. If the remaining lead time t_i cannot be met, then the vertex is simply discarded. Let $p_i = p_{loc_i}^{k_i, q_i}$ be the processing time of the operation represented by vertex i. In order to be completed on time, operation i must begin no later than $t_i - p_i$. This means that loc_i has to receive all required materials within a replenishment lead time of at most $t_i - p_i$ time units. This condition is met if, for each required material, at least one origin o and one transportation mode $m' \in M_{o, loc_i}$ exist such that $t_{o, loc_i}^{m'} \leq t_i - p_i$. Within procedure ADD_FACILITY_VERTEX, on line 7, the previous verification is made in the case of MTO-operations at warehouses. It should be noted that the $min\{\cdot\}$ function is used here to find the fastest replenishment option for the warehouse considering plants $o \in P^{k_i}$. If the lead time defined for vertex i can be met by using this option, then the vertex is added to the network. When considering MTO-operations at plants, the previous condition must be extended given that for every commodity k' that is required for the production of k_i, a replenishment within $t_i - p_i$ must be possible (see line 21 within procedure ADD_FACILITY_VERTEX). Given that raw material suppliers do not have further upstream stages in the considered supply network, merely $p_i \leq t_i$ must hold for a feasible MTO-operation (see line 32 within procedure ADD_FACILITY_VERTEX).

ADD_FACILITY_VERTEX(vertex i, vertex j, transportation mode m)

if $loc_i \in W$ **and** $t_i = 0$ **then** ▷ Vertex represents a **MTS-warehouse operation**
  $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Add vertex and arc to network
  for $\forall o \in P^{k_i}, m' \in M_{o,loc_i}$ **do** ▷ For each potential supply relationship
    ADD_FACILITY_VERTEX$((o,k_i,q_i,0),i,m')$; ▷ MTS option
end if

if $loc_i \in W$ **and** $t_i > 0$ **then** ▷ Vertex represents a **MTO-warehouse operation**
  if $\min\{tt^{m'}_{o,loc_i} \mid o \in P^{k_i}, m' \in M_{o,loc_i}\} \leq t_i - pt_i$ **then** ▷ Test for MTO condition
    $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Add vertex and arc to network
    for $\forall o \in P^{k_i}, m' \in M_{o,loc_i}$ **do** ▷ For each potential supply relationship
      if $tt^{m'}_{o,loc_i} \leq t_i - pt_i$ **then** ADD_FACILITY_VERTEX$((o,k_i,q_i,0),i,m')$; ▷ MTS option
      if $tt^{m'}_{o,loc_i} < t_i - pt_i$ **then** ADD_FACILITY_VERTEX$((o,k_i,q_i,t_i - pt_i - tt^{m'}_{o,loc_i}),i,m')$; ▷ MTO option
    end for
  end if
end if

if $loc_i \in P$ **and** $t_i = 0$ **then** ▷ Vertex represents a **MTS-plant operation**
  $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Add vertex and arc to network
  for $\forall k' \in R \cup I, a^{k',k_i} > 0, \forall o \in O^{k'}, \forall m' \in M_{o,loc_i}$ **do** ▷ For each potential supply relationship
    ADD_FACILITY_VERTEX$((o,k',a^{k',k_i} \cdot q_i,0),i,m')$; ▷ MTS option
end if

if $loc_i \in P$ **and** $t_i > 0$ **then** ▷ Vertex represents a **MTO-plant operation**
  if $\forall k' \in R \cup I, a^{k',k_i} > 0 : \min\{tt^{m'}_{o,loc_i} \mid o \in O^{k'}, m' \in M_{o,loc_i}\} \leq t_i - pt_i$ **then** ▷ Test for MTO condition
    $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Add vertex and arc to network
    for $\forall k' \in R \cup I, a^{k',k_i} > 0, \forall o \in O^{k'}, \forall m' \in M_{o,loc_i}$ **do** ▷ For each potential supply relationship
      if $tt^{m'}_{o,loc_i} \leq t_i - pt_i$ **then** ADD_FACILITY_VERTEX$((o,k',a^{k',k_i} \cdot q_i,0),i,m')$; ▷ MTS option
      if $tt^{m'}_{o,loc_i} < t_i - pt_i$ **then** ADD_FACILITY_VERTEX$((o,k',a^{k',k_i} \cdot q_i,t_i - pt_i - tt^{m'}_{o,loc_i}),i,m')$; ▷ MTO option
    end for
  end if
end if

if $loc_i \in S$ **and** $t_i = 0$ **then** ▷ Vertex represents a **MTS-supplier operation**
  $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Add vertex and arc to network

if $loc_i \in S$ **and** $t_i > 0$ **then** ▷ Vertex represents a **MTO-supplier operation**
  if $pt_i \leq t_i$ **then** $V \Leftarrow V \cup \{i\}; A \Leftarrow A \cup \{(i,j,m)\}$; ▷ Test for MTO condition

Figure 6: Procedure ADD_FACILITY_VERTEX.

Table 3: Additional parameters and decision variables for models SND-TS.

Parameters:	
V	Set of vertices in the time-space network
V^{P}	Set of vertices representing plants, i.e. $V^{\mathrm{P}} = \{i \in V \mid loc_i \in P\}$
V^{W}	Set of vertices representing warehouses
V^{C}	Set of vertices representing customers
loc_i	Customer- or facility-location represented by vertex $i \in V$
k_i	Commodity represented by vertex i
q_i	Quantity represented by vertex i
t_i	Time coordinate of vertex i, i.e. the service outbound time of the operation
A	Set of arcs in the time-space network
c_i^{prod}	Variable cost for producing one unit of k_i using the production operation represented by vertex $i \in V$
Decision variables:	
X_{ij}^m	Flow assigned to arc (i, j, m)

For every created vertex associated with a warehouse or plant operation (i.e., $loc_i \in W$ or $loc_i \in P$), the procedure ADD_FACILITY_VERTEX further creates supply relationships for the required materials and intermediate products in a recursive fashion. If i refers to an MTS-operation at a warehouse given a product k_i, every combination of plants $o \in P^{k_i}$ and transportation modes $m \in M_{o,loc_i}$ is considered as a potential supply relationship (see lines 3-4 within procedure ADD_FACILITY_VERTEX). In this case, only those plant operations that are run under MTS are considered. If i refers to an MTO-operation at a warehouse, supply operations at plants may be run under either MTS or MTO, given that the conditions defined on lines 10-11 within procedure ADD_FACILITY_VERTEX are fulfilled. If i represents a production operation at a plant, supply relationships are established for every material and intermediate product $k' \in R \cup I$, for which $a^{k',k_i} > 0$, see lines 17 and 23. In this case, the quantity of k' delivered by the supplying facilities must be $a^{k',k_i} \cdot q_i$ units, see lines 18, 24, and 25 in Fig. 6. Using the outlined procedures, a complete time-space network can be produced for a problem instance within which, customer orders are routed timely. Appendix B illustrates this process in detail at the example of Fig. 4.

5.2 Model Formulation

Recall that the total demand d_c^f of customer c for product f is assumed to be split into individual orders of size $\overline{q}$. Since congestion effects at facilities are not considered in this study, we can route all these $d_c^f/\overline{q}$ orders at minimum cost through the time-space network, where the temporal structure of the network ensures that the maximum lead times are not violated. In other words, an optimal network design to the problem can be derived from a minimum cost flow in the time-space network where each vertex associated to a particular customer-product combination receives the amount of the final product demanded by the customer. Given that each vertex $i \in V$ in the time-space network is associated to a given quadruple (loc_i, k_i, q_i, t_i), we model the flow of commodities in the network by variable X_{ij}^m who represents the amount of commodity k_i sent from loc_i to loc_j using transportation mode m. To evaluate the cost of a solution, all flow leaving a vertex i is charged according to production cost rate c_i^{prod}, which is either set to $c_i^{\text{prod}} = c_{loc_i,k_i}^{\text{MTS}}$, for those vertices i that represent MTS-operations (i.e., $t_i = 0$), or to $c_i^{\text{prod}} = c_{loc_i,k_i}^{\text{MTO}}$, for those vertices that represent MTO-operations (i.e., $t_i > 0$). Furthermore, associated with each arc (i, j, m) included in the network is a variable transportation cost rate $c_{loc_i,loc_j}^{k_i,m}$, which is defined as before, see Table 1. The additional notation used to formulate the optimization model proposed in this section are summarized in Table 3. The supply network design problem defined on a time-space network (SND-TS), can now be defined as follows:

$$\text{[SND-TS]} \quad \min \rightarrow Z = \sum_{o \in O} c_o^{\text{fix}} U_o + \sum_{k \in K} \sum_{o \in O^k} c_{ok}^{\text{fix}} V_o^k + \sum_{k \in K} \sum_{o \in O^k} \sum_{d \in D^k} c_{kod}^{\text{fix}} Y_{od}^k + \sum_{o \in O} \sum_{d \in D} \sum_{m \in M_{od}} c_{mod}^{\text{fix}} Z_{od}^m + \sum_{(i,j,m) \in A} X_{ij}^m \left(c_i^{\text{prod}} + c_{loc_i, loc_j}^{k_i, m} \right) \tag{30}$$

$$\sum_{(j,i,m) \in A | k_j = k'} X_{ji}^m - a^{k',k_i} \sum_{(i,j,m) \in A} X_{ij}^m = 0 \qquad i \in V^{\text{P}}; k' \in I^{k_i} \cup F^{k_i} \tag{31}$$

$$\sum_{(j,i,m) \in A} X_{ji}^m - \sum_{(i,j,m) \in A} X_{ij}^m = 0 \qquad i \in V^{\text{W}} \tag{32}$$

$$\sum_{(j,i,m) \in A} X_{ji}^m = d_{loc_i}^{k_i} \qquad i \in V^{\text{C}} \tag{33}$$

$$\sum_{(i,j,m) \in A | loc_i = o} u^{k_i} \cdot X_{ij}^m - u_o \cdot U_o \leq 0 \qquad o \in O \tag{34}$$

$$\sum_{(i,j,m) \in A | loc_i = o, k_i = k} X_{ij}^m - q_o^k \cdot V_o^k \leq 0 \qquad k \in K; o \in O^k \tag{35}$$

$$\sum_{(i,j,m) \in A | loc_i = o, loc_j = d, k_i = k} X_{ij}^m - q_{od}^k \cdot Y_{od}^k \leq 0 \qquad k \in K; o \in O^k; d \in D^k \tag{36}$$

$$\sum_{(i,j,m) \in A | loc_i = o, loc_j = d} g^{k_i,m} \cdot X_{ij}^m - g_{od}^m \cdot Z_{od}^m \leq 0 \qquad o \in O; d \in D; m \in M_{od} \tag{37}$$

and $(19) + (20)$

$$X_{ij}^m \geq 0 \qquad (i,j,m) \in A \tag{38}$$

The objective function (30) is equivalent to the one defined for model SND, i.e., function (1). The variable costs of both production and transportation operations are now simply associated to the flow variables X_{ij}^m. Constraints (31) enforce flow conservation at vertices $i \in V^{\text{P}}$, which represent plant operations for transforming raw materials and intermediate products k' into commodity k_i. Here, the first sum on the left side determines the inflow of k' from vertices j with $k_j = k'$ and the second sum defines the outflow of product k_i from vertex i. Equivalent units are obtained through the use of the production coefficient a^{k',k_i}. Constraints (32) impose flow conservation for warehouse operations and (33) ensure that customer demands are met. Inequalities (34) to (37) define the linking constraints between the design and the flow variables. Constraints (34) specify that an origin o can supply commodities up to its capacity, only if it is selected (i.e., $U_o = 1$). Similarly, (35) to (37) set variables V_o^k, Y_{od}^k, and Z_{od}^m according to the flow on the relevant subsets of arcs. Finally, constraints (19), (20) and (38) define integrality requirements for the design variables and non-negativity for the flow variables used in the model. It should be noted that a facility can now produce a product under both MTO and MTS fashions, i.e., model SND-TS is a generalization of model SND.

5.3 Valid Inequalities

The LP relaxation of model SND-TS can be strengthened by adding valid inequalities (21)-(26). We can further strengthen the relaxation by adding inequalities that take advantage of the fact that vertices $i \in V$ in the time-space network represent specific production operations for q_i units of commodity k_i at facility loc_i. To do so, both the MPBOM and the customer demand for final products are used to determine the total flow for each combination of a commodity k and a specific production quantity q that has to be processed at the vertices with $k_i = k$ and $q_i = q$. This flow value is denoted by $A^{\max}_{kq}$. As an example, consider Fig. 1b and let us assume that the total customer demand for product F1 is 1000 units. For intermediate product I1, which specifies a material quantity of 3 in this tree, the total demand for I1 at this production stage is $A^{\max}_{\text{I1},3} = 3000$ units. If a particular MPBOM contains multiple vertices that represent the same combination (k,q), then $A^{\max}_{kq}$ is determined as the customer demand weighted sum of the material quantities for these vertices. With respect to the time-space network, value $A^{\max}_{kq}$ represents an upper bound on the total outflow of vertices $i \in V$, for which $k_i = k$ and $q_i = q$. In the previous example, the total flow of I1 leaving a vertex i of the time-space network, where $k_i = \text{I1}$ and $q_i = 3$, is at most $A^{\max}_{\text{I1},3} = 3000$ units. Therefore, values $A^{\max}_{kq}$ can be applied as big-M parameters in (39)-(42), thus tightening constraints (34)-(37).

$$\sum_{(i,j,m)\in A|loc_i=o,k_i=k,q_i=q} X^m_{ij} - A^{\max}_{kq} \cdot U_o \leq 0 \qquad k \in K; q \in Q^k; o \in O^k \tag{39}$$

$$\sum_{(i,j,m)\in A|loc_i=o,k_i=k,q_i=q} X^m_{ij} - A^{\max}_{kq} \cdot V^k_o \leq 0 \qquad k \in K; q \in Q^k; o \in O^k \tag{40}$$

$$\sum_{(i,j,m)\in A|loc_i=o,k_i=k,q_i=q,loc_j=d} X^m_{ij} - A^{\max}_{kq} \cdot Y^k_{od} \leq 0 \qquad k \in K; q \in Q^k; o \in O^k; d \in D^k \tag{41}$$

$$\sum_{(i,j,m)\in A|loc_i=o,k_i=k,q_i=q,loc_j=d} X^m_{ij} - A^{\max}_{kq} \cdot Z^m_{od} \leq 0 \qquad k \in K, q \in Q^k, o \in O^k; d \in D^k; m \in M_{od} \tag{42}$$

6 Computational Study

Two main objectives are pursued through the computational study that is conducted here. The first is to compare the two mathematical formulations proposed for the considered design problem of supply networks. As such, we evaluate both formulations in terms of solution time efficiency and test the effectiveness of the different valid inequalities proposed to strengthen the models. The second objective is to investigate the impact of considering explicitly the restriction of order lead times for customers. Specifically, we analyze the impact that such restrictions have on the resulting network designs and costs. We next describe how the used instances were generated, which we then follow by the presentation of the computational results.

6.1 Generation of Test Instances

We generate four sets of instances, denoted: *A*, *B*, *C*, and *D*, that differ according to the number of facilities, customers, and commodities. In Table 4, the different parameters relating to the size of the instances are summarized. For example, each instance in set *A* contains $|C| = 30$ customers, $|S| = |P| = |W| = 3$ suppliers, plants, and warehouses as well as $|R| = |I| = |F| = 3$ raw materials, intermediate products, and final products. The number of raw materials that can be provided by a supplier is shown in column $|K_s|$ of Table 4. When defining which of the raw materials can be supplied by $s \in S$, $|K_s|$ commodities are simply selected randomly from set R. Similarly, a plant $p \in P$ can provide $|K_p|$ commodities. Half of these commodities are intermediate products, selected randomly from set I, and the other half are final products, selected randomly from set F. Warehouses $w \in W$ are allowed to handle all final products, i.e., $K_w = F$. To obtain feasible problems, all instances are generated such that every raw material is supplied by at least one supplier and that every intermediate and final product can be produced by at least one plant.

The product structure of each instance is determined randomly such that the associated MPBOM represents a connected acyclic graph. The graphs are generated by a process described in Hall and Posner (2001), which delivers a precedence graph of desired density for the commodities. The density has been chosen such that for each combination of a raw material $r \in R$ and an intermediate or final product $k \in I \cup F$, a 20% probability exists that r is needed to produce k. The same probability is applied to each pair of intermediate products and to pairs of an intermediate product and a final product. The generation process ensures that every raw material is used within at least one intermediate product and every intermediate product is used within at least one final product. The resulting MPBOMs include up to 200 vertices and up to 6 production stages, when represented as a tree graph such as the one exemplified in Fig. 1b. As for the production coefficients $a^{k,k'}$, they are randomly generated using a uniform distribution $U[1,3]$. The processing time to produce or handle q units of commodity k at origin o is set to $p_o^{kq} = q$ time units, for all commodities and origins (i.e., processing times grow linearly with respect to the quantity that is produced). Customer demands are generated as follows: for each customer $c \in C$ and each final product $f \in F$, there is a 50% probability that the customer is assigned a positive demand for the product; in which case, d_c^f is randomly generated using the distribution $U[50,500]$. The default size of orders placed by customers is set to

Table 4: Parameters of test instances.

set	$\lvert C\rvert$	$\lvert S\rvert, \lvert P\rvert, \lvert W\rvert$	$\lvert R\rvert, \lvert I\rvert, \lvert F\rvert$	$\lvert K^s\rvert$	$\lvert K^p\rvert$	fixed cost multiplier
A	30	3	3	2	4	1
B	50	5	5	3	6	10
C	80	8	8	4	8	50
D	100	10	10	5	10	100

$\overline{q} = 1$ product unit. The maximum order lead times are initially set to 10 time units for all customers and products (i.e., $\tau_c^f = 10$, $\forall c \in C$ and $f \in F$). It should be noted that, in order to investigate the sensitivity of solutions with respect to variations in the maximum lead times, this parameter will later be fixed to different values.

Capacity parameters are obtained using the generation process described in Cordeau et al (2006). For each commodity $k \in K$, a per unit capacity requirement u^k is randomly generated using the distribution $U[1,10]$. Let u^{R}, u^{I}, and u^{F} define the total capacity required to produce the total amount of raw materials, intermediate products, and final products that are needed to fulfill all customer demands. These values are used to define the capacities of the different facilities within the network. The production capacity u_s of supplier s is randomly generated using $U[2 \cdot u^{\mathrm{R}}/|S|, u^{\mathrm{R}}]$. The production capacity u_p of plant p is randomly obtained from $U[2 \cdot (u^{\mathrm{I}} + u^{\mathrm{F}})/|P|, u^{\mathrm{I}} + u^{\mathrm{F}}]$. The capacity of warehouse $w \in W$ is set to $u_w = u^{\mathrm{F}}$. Upper limits q_o^k on the amount of commodity k to be provided at origin o are determined similarly. Let a^k define the total amount of commodity k that is necessary to service all demands. Values q_s^r associated to suppliers $s \in S^r$ are randomly obtained from $U[2 \cdot a^r/|S^r|, a^r]$, values q_p^k associated to commodities k at plants $p \in P^k$ are randomly generated using $U[2 \cdot a^k/|P^k|, a^k]$, and values q_w^f are set equal to a^f for all warehouses. The upper limit q_{od}^k on the amount of commodity k that can be shipped from o to d is set to $q_{od}^k = q_o^k$. The capacity requirement per unit of k on a transportation mode m is set to $g^{km} = 1$ and the total capacity g_{od}^m of mode m is set to the maximum amount of flow of commodities that can be sent from o to d. Given these settings, the capacities of both warehouses and transportation modes do not restrict the material flows in the design problem, but they serve as appropriate big-M values in the models.

The facilities and customers are located randomly on a plane of size 10×10 distance units. For every transportation link between two vertices in the network, up to two transportation modes are created. For the first mode, the transportation time t_{od}^m is set to the euclidean distance between facilities o and d. The transportation cost rate c_{od}^{km} is set to the euclidean distance divided by 10. Cost rates for direct transportation from plants to customers are multiplied by a factor of five, given that direct deliveries are considered to be more expensive due to the additional transportation resources required. A second mode of transportation is added randomly, with a probability of 50%, on all transportation links in the network. When available, this second mode requires only half of the transportation time defined for the first mode. However, the variable transportation cost is increased by a factor of 50%. The unit production costs c_{ok}^{MTO} for commodity k at origin o are randomly obtained using $U[1,2]$, $U[3,4]$, $U[5,6]$, and $U[1,2]$ for raw materials at suppliers, intermediate products at plants, final products at plants, and for handling commodities at warehouses, respectively. We set $c_{ok}^{\mathrm{MTS}} = 1.4 \cdot c_{ok}^{\mathrm{MTO}}$ to account for the additional costs related to MTS production. The fixed costs c_o^{fix} defined for using a supplier, a plant, and a warehouse are randomly generated using $U[1000,2000]$, $U[5000,10000]$, and $U[1000,10000]$, respectively. Fixed costs c_{ok}^{fix}, c_{kod}^{fix}, and c_{mod}^{fix} are all randomly obtained using $U[100,500]$. It should be noted that the variable costs of solutions will tend to increase for

Table 5: Number of variables and constraints.

	Both Models				Model SND								Model SND-TS	
set	U_o	V_o^k	Y_{od}^k	Z_{od}^m	A_{ok}^{MTO}	A_{ok}^{MTS}	B_o^k	T_o^{kq}	T_c^f	Z_{od}^{km}	X_{od}^{km}	# Constr.	X_{ij}^m	# Constr.
A	9	27	315	291	27	27	27	36	48	454	454	1555	920	752
B	15	70	1378	807	70	70	70	108	131	2007	2007	6120	4552	2673
C	24	160	5149	2169	160	160	160	313	334	7612	7612	21447	20117	8797
D	30	250	9774	3419	250	250	250	635	496	14413	14413	41111	43752	15932

the instances in sets *B* to *D*, due to the increasing number of customers, commodities, and more complex product structures. Therefore, to ensure that the impact of the fixed costs remain comparable on the larger instances, their values are multiplied by the preset parameter reported in the last column of Table 4.

6.2 Computational Results

The first numerical tests conducted are used to compare the two models proposed (i.e., SND and SND-TS). Through this comparison, we investigate the effectiveness of both the presented valid inequalities and the problem reformulation based on the time-space network. All tests are conducted using ILOG Cplex 12 on a 2.4 GHz AMD Opteron 64-bit processor. We set a maximum runtime of 10 hours per instance.

To first compare the sizes of the problems to be solved, Table 5 lists the number of variables and constraints that are defined using each of the proposed models for instances included in sets *A*, *B*, *C*, and *D*. The table distinguishes those variables that are part of both models from those that are specific to a particular formulation. As expected, both the numbers of variables and constraints increase significantly for the two models, as the size of instances increases. Considering the largest instances (set *D*), it should be noted that the number of constraints is much higher in model SND when compared to model SND-TS. However, the latter contains significantly more variables to account for the high number of alternative flows that exist in the time-space network. When comparing the type of variables specifically used in each formulation, one notices that the SND-TS model requires flow variables (i.e., X_{ij}^m), which are continuous, whereas the SND model requires a substantial number of extra binary variables Z_{od}^{km}, which are used to determine order lead times. From this observation, one may expect that the SND-TS model is easier to solve using an MIP solver.

We now compare the models on the basis of the computational results obtained using the Cplex solver. The four instance sets have been solved using models SND and SND-TS with and without the inequalities previously defined. The analysis is performed here by adding subsets of valid inequalities to both formulations. This is done to avoid a lengthy presentation of the computational results, which would inevitably follow from measuring the impact of the inequalities by including them one at the time. In the following, SND refers to the original model (1)-(20) whereas $\mathrm{SND}^{\mathrm{Ineq1}}$, $\mathrm{SND}^{\mathrm{Ineq2}}$, and $\mathrm{SND}^{\mathrm{IneqAll}}$ refer to model SND when supplemented with valid inequalities

(21)-(26), (27)-(29), and (21)-(29), respectively. Accordingly, SND-TS refers to (30)-(38) whereas SND-TS^{Ineq1}, SND-TS^{Ineq2}, and SND-TS^{IneqAll} refer to SND-TS supplemented with (21)-(26), (39)-(42), and both sets of valid inequalities, respectively. Hence, the Ineq1-models include valid inequalities previously proposed in the context of network design problems. The Ineq2-models are obtained by adding new inequalities developed specifically for the considered problem. Finally, the IneqAll-models make use of all valid inequalities available.

Table 6 reports detailed results for the ten instances of set *A*, when solved under models SND^{IneqAll} and SND-TS^{IneqAll}. Included within the table are the following informations: both the total cost *Z* of the solutions obtained and the lower bound *LB* returned by Cplex, the relative $gap = (Z - LB)/Z$ in percent, and the required computation time *CPU* in seconds. Following the suggestion made in Cordeau et al (2006), when solving a problem instance, Cplex is terminated as soon as an integer solution is found within an optimality gap of 1%. The reasoning behind the use of this stopping criterion is that the additional computational time required to close the gap is unjustified in light of the fact that errors in the estimations of the input data for the problem are often greater than 1% in practice. Considering this threshold, whenever the solver terminates based on the defined criterion, we nevertheless refer to the solution obtained as an optimal solution. As previously mentioned, the maximum allotted time for solving an instance is fixed to 10 hours. If '-' appears in column *CPU*, the instance has not been solved to optimality within the runtime limit. Also, one should recall that model SND-TS is a generalization of model SND, given that facilities can produce a commodity under MTO and MTS fashions, whereas in the SND formulation one production strategy must be used exclusively for a commodity at a facility. For this reason, given a problem instance, the optimal solution for the SND-TS model may have a total cost *Z* which falls below that obtained for the optimal solution to model SND. In certain cases, this value may even be lower than *LB* obtained for model SND. Therefore, we additionally report the relative improvement *imp* in percent of the cost *Z* of a solution for SND-TS^{IneqAll} over a solution for SND^{IneqAll}. As such, values *imp* measure the added advantage of adopting a more flexible production scheme.

When examining the results reported in Table 6, one observes that even the small instances of set *A* are difficult to solve using model SND^{IneqAll}. When solving the ten instances, Cplex terminates ahead of the runtime limit only three times. The observed gaps range from 1.0% to 6.9%, the average being 2.6%. Compared to these results, solving model SND-TS^{IneqAll} improves the computational performance of the MIP solver tremendously. Each instance is solved using less than a second of computation time with the average gap obtained being 0.5%. Moreover, the cost of the solutions found using model SND-TS^{IneqAll} are, on average, 4.0% below the cost of the solutions produced by solving SND^{IneqAll}. These results tend to show that by formulating the supply network design problem as a time-space network flow problem, one obtains a model that can be solved much more efficiently. Also, solving model SND-TS^{IneqAll} improves the solution quality produced when compared to solving model SND^{IneqAll}. Thus, showing that there is an added advantage in using the more flexible production scheme defined in model SND-TS,

Table 6: Computational results for instances in set A under models SND$^{\text{IneqAll}}$ and SND-TS$^{\text{IneqAll}}$.

	SND$^{\text{IneqAll}}$				SND-TS$^{\text{IneqAll}}$				
#	Z	LB	gap [%]	CPU [sec.]	Z	LB	gap [%]	imp [%]	CPU [sec.]
1	714655	707508	1.0	3127	706382	704778	0.2	1.2	<1
2	596331	555722	6.8	36000	559403	556927	0.4	6.2	<1
3	733563	715843	2.4	36000	701342	694734	0.9	4.4	<1
4	863781	840358	2.7	36000	842803	840243	0.3	2.4	<1
5	1241766	1227140	1.2	36000	1183244	1179473	0.3	4.7	<1
6	525942	515946	1.9	36000	512286	510538	0.3	2.6	<1
7	319312	316119	1.0	219	306457	304738	0.6	4.0	<1
8	794525	739612	6.9	36000	727327	724441	0.4	8.5	<1
9	642854	636425	1.0	28594	626953	621563	0.9	2.5	<1
10	1283250	1265712	1.4	36000	1238781	1230344	0.7	3.5	<1
avg.	771598	752039	2.6	28394	740497	736778	0.5	4.0	<1

at least when addressing small sized instances.

In Table 7 are summarized the results obtained for all four instance sets, when solved using all considered configurations for models SND and SND-TS. The informations reported on each row are the following: the number of instances for which an integer feasible solution was found in the maximum allotted time (column $\#f$), the number of instances solved to optimality within the runtime limit ($\#o$), the average relative gap in percent for the solutions to the ten instances within the set (gap), and the average computation time in seconds (CPU). It should be noted that, for each problem instance, the gap is measured using the best lower bound obtained by the four variants of the corresponding model. Moreover, the average CPU times consider only those instances that were solved to optimality within the runtime limit. A '-' indicates that none of the instances of a corresponding instance set was solved to optimality.

We first discuss the results obtained when using model SND and its variants. Table 7 reveals that the average gap of the solutions obtained when solving the instance set A is hardly affected by the selection of valid inequalities. However, three instances are solved within the maximum allotted time when all inequalities are employed. For set B, no optimal solutions are found, but feasible solutions are found for all instances. The lowest average gap is observed for SND$^{\text{Ineq2}}$. In the case of set C, Cplex fails to provide feasible solutions for all instances. The best results, measured both in terms of highest number of feasible solutions and lowest average gap obtained, are again produced by SND$^{\text{Ineq2}}$. For model SND$^{\text{Ineq1}}$, the same number of feasible solutions is found but a very high gap is observed. This result is explained by the fact that the large number of added valid inequalities increases the complexity of the associated model. Although a high number of feasible solutions is still generated under this model, the expected speed up of the solution process does not occur and good feasible solutions are not found

Table 7: Summarized results for the four instance sets.

	SND				SND^{Ineq1}				SND^{Ineq2}				$SND^{IneqAll}$				
	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	
set	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	
A	10	0	2.6	36000	10	0	2.7	36000	10	0	2.5	36000	10	3	2.6	29091	
B	10	0	8.1	36000	10	0	8.3	36000	10	0	6.5	36000	10	0	6.7	36000	
C	8	0	8.3	36000	9	0	46.0	36000	9	0	7.1	36000	4	0	10.6	36000	
D	10	0	8.5	36000	6	0	84.1	36000	6	0	11.3	36000	3	0	5.8	36000	
	SND-TS				$SND\text{-}TS^{Ineq1}$				$SND\text{-}TS^{Ineq2}$				$SND\text{-}TS^{IneqAll}$				
	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	$\#f$	$\#o$	gap	CPU	imp
set	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	[-]	[-]	[%]	[sec.]	[%]
A	10	10	0.4	<1	10	10	0.5	<1	10	10	0.4	<1	10	10	0.5	<1	4.0
B	10	2	0.8	34208	10	10	0.9	142	10	5	0.8	22761	10	10	0.8	124	2.8
C	10	0	2.3	36000	10	0	2.2	36000	10	0	2.3	36000	10	0	2.4	36000	1.6
D	10	0	3.8	36000	10	0	4.6	36000	10	0	6.5	36000	10	0	45.6	36000	2.8

within the alloted time. This is confirmed, when analyzing the results obtained for the large instances of set D. In this case, feasible solutions are obtained on all instances, if model SND is used without any inequalities. However, the average *gap* obtained (i.e., 8.5%) is comparably high. It can be concluded that the valid inequalities are helpful in the case of small and medium sized instances. In contrast, when solving large instances, they may hinder Cplex in finding good solutions.

We now analyze the results obtained when using model SND-TS. We first find that, when solving the instances in set A, Cplex is able to solve model SND-TS to optimality in less than a second under every configuration. For instances in set B, inequalities clearly pay off as they considerably reduce the computation times and increase the number of instances solved to optimality, when compared to the original model SND-TS. The best configuration is obtained when all inequalities are added to the model. Using $SND\text{-}TS^{IneqAll}$, optimal designs are found for all instances within, on average, 124 seconds. In the case of set C, Cplex is unable to solve the instances in the maximum allotted time under any model configuration. However, the solver produces feasible solutions on all instances and the average gaps obtained are low (i.e., between 2.2% and 2.4%). Interestingly, when addressing the large instances in set D, we again observe that the best configuration, considering average solution quality, is the model that excludes all additional valid inequalities (the average gap obtained being 3.8%). Therefore, when tackling large instances, valid inequalities should be tentatively added given that they might render the model too hard to solve.

We now compare the cost of solutions obtained by both types of models to assess the added advantages of using the more flexible production policy defined in the SND-TS model (i.e., a facility can produce a commodity under both MTS and MTO strategies), compared to the more strict policy used in the SND model (i.e., one strategy

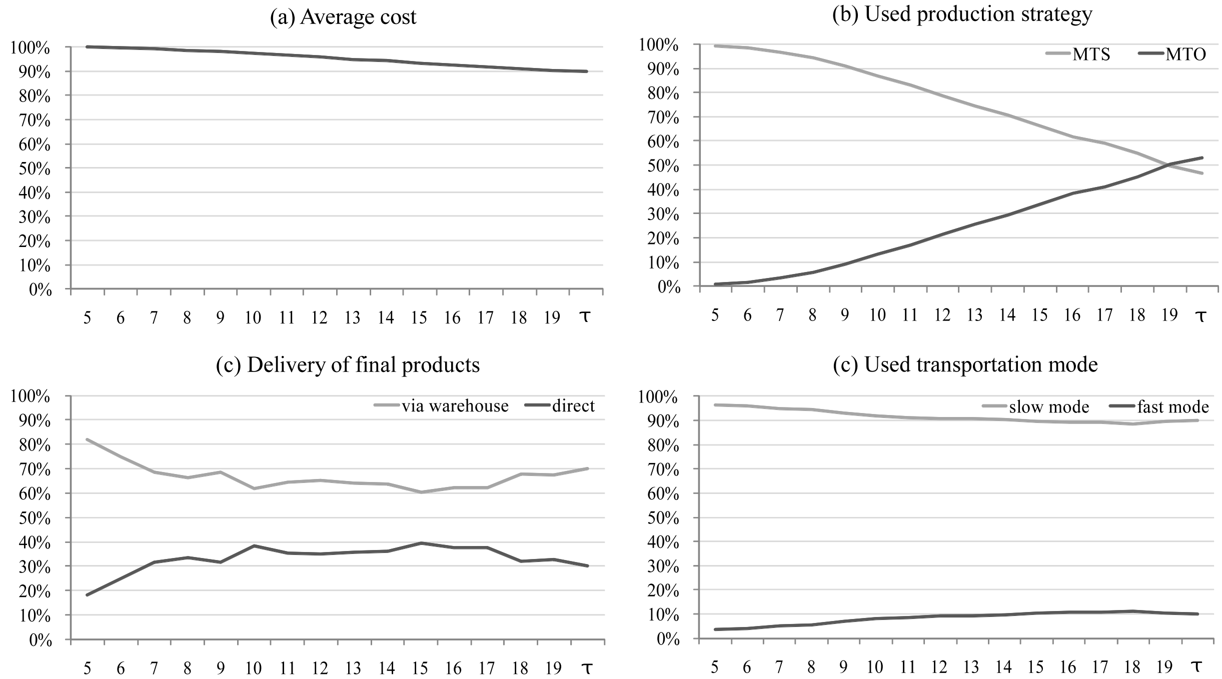

Figure 7: Change in performance measures under varied maximum order lead times.

has to be chosen for a commodity at a facility). In Table 7, values *imp* show that the best solutions produced by solving SND-TS have a total cost that is less than the one obtained for the best solutions produced by solving SND. The average relative improvements range from 1.6% to 4.0% for instances in sets *A* to *D*. These results prove that by using a more flexible production policy, one is able to obtain better quality solutions. Also, to summarize this test, one should notice the differences in terms of the efficiency of Cplex when solving both models. When solving the SND-TS model, Cplex produces feasible solutions on more problem instances when compared to the case where it solves the SND model. Moreover, on those instances solved to optimality, computation times are reduced drastically for problems of small and medium sizes when solving SND-TS compared to SND. Overall, these observations show not only the efficiency of the time-space formulation, but also the difficulty that is added to the network design problem, when the more restrictive production policy is used.

We next investigate the impact that the lead time restrictions have on the obtained networks. To do so, the instances of set *B* are solved under varied maximum lead times, i.e., $\tau_c^f = \tau = 5, 6, \ldots 20$. The choice of set *B* is motivated by the fact that it includes the largest instances that are consistently solved to optimality using model SND-TS$^{\text{IneqAll}}$. Figure 7 shows, for a series of selected measures, the performance of the networks obtained under each value of τ. Figure 7a illustrates the average costs observed for the instances in the set. The highest cost are observed under $\tau = 5$. They serve as a reference for assessing the relative cost decrease observed when customers accept looser lead times, i.e., when τ takes a higher value. As expected, the figure shows that, as the value of τ increases, the average cost decreases. One actually observes that for $\tau = 20$ the average cost represents 90% of the average cost obtained for $\tau = 5$. Hence, in the considered problem setting, costs can be reduced by about 10% if customers accept the looser maximum lead times of 20 time units. Figure 7b reports the total average commodity quantities that are produced under the MTS and MTO strategies. One clearly sees that when $\tau = 5$, all commodities are produced under MTS to meet the restrictive lead time requirements of customers. As lead time requirements are relaxed, there is a shift in production from MTS to MTO. When $\tau = 20$, the quantity of commodities produced under MTO even exceeds the MTS production. What these results show is the strong incentive that exists to propose higher lead times to customers in order to reduce production related costs. Figure 7c distinguishes between the average quantities of final products that are either delivered directly from plants to customers or through indirect deliveries via warehouses. One can observe that for low values of τ, the direct delivery option is surprisingly used less frequently. Although one would expect that direct deliveries are essential for serving customers within low lead times, warehouses actually allow to bring products closer to customers which, in combination with MTS driven production and distribution operations, is needed when accepted lead times are very low. When τ is increased, direct deliveries are used in a larger extend to compensate for the increase in the order lead times caused by switching from an MTS to an MTO driven production strategy. The small peaks observed for $\tau = 8$, 10, and 15 in Fig. 7c, show that direct deliveries are effectively used to fine-tune the realized order lead times. The use of fast modes of transportation also increases with maximum lead times, see Fig. 7d. Again, whenever production transfers from MTS to MTO, these modes are used to counterbalance the increasing lead times. However, when compared to the average quantity of commodities transported using slower modes, fast modes only account for approximately 10% of all transports. The reasons for this being that the fast mode is not available on every transport relation. Furthermore, using fast modes does not provide any advantages when the network operates under the MTS strategy. Overall, the summarized results reported in Fig. 7 show that the available options to control lead times can be effectively combined to meet the service requirements of customers while keeping costs low. The option which seems to be the most important is the selection of production strategies MTS and MTO. While the MTS strategy effectively shortens customer lead times, the MTO strategy can be used to decrease variable production costs whenever the lead time requirements are less restrictive. The options of direct deliveries and fast modes of transportation can be used to fine-tune the operations within the networks that are designed.

Finally, to provide further insight into the structure of optimal networks, Table 8 reports the opened facilities for solutions to set B for $\tau = 5$, 10, 15, and 20 time units. The table reveals that the set of opened warehouses changes, in nine out of the ten instances, when parameter τ varies. This illustrates that warehouses can be effectively used to align the distribution system with the service requirements defined. Changes to the supplier and plant layers are observed in only five and four instances, respectively. Opening decisions for these facilities appear comparably equivalent when considering different lead time requirements. At these layers, the option of applying MTS-driven

Table 8: Facilities opened within the optimal network designs for instances in set B under selected values of τ.

	Instance 1															Instance 2														
τ	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5
5	o		o		o		o	o	o	o		o				o		o	o	o	o		o		o			o		o
10	o		o	o	o		o	o	o	o		o				o		o	o	o	o	o		o	o			o		
15	o		o	o	o		o	o	o	o		o				o		o	o		o	o		o	o			o		
20	o		o	o	o		o	o	o	o			o			o		o	o		o	o		o	o			o		
	Instance 3															Instance 4														
τ	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5
5	o	o		o		o	o			o		o		o			o	o	o	o	o		o	o	o		o			
10	o	o		o		o	o			o		o				o	o		o	o	o	o	o	o				o		
15	o			o	o	o	o			o				o	o		o	o	o	o	o		o	o	o			o		
20	o	o		o		o	o			o				o		o	o	o	o	o	o		o	o	o			o		
	Instance 5															Instance 6														
τ	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5
5	o	o	o	o	o	o	o			o					o	o	o	o			o	o	o		o			o	o	
10		o	o	o	o	o	o		o						o	o	o	o			o	o	o		o	o				
15	o	o	o	o	o	o	o		o						o	o	o	o			o	o	o		o			o		
20	o	o	o	o	o	o	o		o						o	o	o	o			o	o	o		o				o	
	Instance 7															Instance 8														
τ	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5
5	o	o	o		o			o	o	o			o			o	o	o	o	o	o		o	o	o	o				
10	o	o	o		o			o	o	o			o			o	o	o	o	o	o		o	o	o		o			
15	o	o	o		o			o	o	o		o	o			o	o	o	o	o	o		o	o	o			o		
20	o	o	o		o			o	o	o		o	o			o	o	o	o	o	o		o	o	o		o			
	Instance 9															Instance 10														
τ	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5	S1	S2	S3	S4	S5	P1	P2	P3	P4	P5	W1	W2	W3	W4	W5
5	o	o		o	o		o	o	o	o	o					o	o		o	o	o	o	o	o	o	o			o	
10	o	o		o	o	o	o	o	o				o			o	o		o	o	o	o	o	o	o		o			
15	o	o		o	o		o	o	o	o			o			o	o		o	o	o	o	o	o	o					o
20	o	o		o	o		o	o	o	o			o			o	o		o	o	o	o	o	o	o		o			

production seems a more viable strategy for meeting maximum lead times.

Summarizing the computational tests, the obtained results reveal that restricted lead times clearly impact the decisions made within the supply network design. These result confirm the predominant importance of strategic network planning if customer order lead times are to be used as a competitive factor in the market. Towards this end, the proposed models represent useful decision support tools.

7 Conclusions

This paper investigated the design of supply networks under a restriction for customer order lead times. Two models were provided to address the general problem of designing multi-layer multi-product facility networks with complex product structures and multiple modes of transportation where customers receive products within guaranteed order lead times. Various strategies, aimed at controlling lead times, were considered within these models. These strategies included: the selection of MTS or MTO production, direct deliveries from plants to customers, and multiple modes of transportation that differ with respect to cost and transportation time.

Instances of various sizes were solved using ILOG Cplex in order to assess the efficiency of the two models proposed. It was shown that the new time-space network model proposed in this paper is much easier to solve for a standard MIP solver. The numerical analysis performed were also used to identify the interrelations that exist between the considered strategies and the order lead times, and to measure the impact that these strategies have on the optimal network designs obtained. It was generally found that the considered strategies for controlling lead times are used to different extents to achieve a given service level at lowest cost. The production strategy MTS is used to effectively reduce the customer order lead times, whereas MTO is more appropriate to efficiently lower production costs. Direct deliveries and fast transportation modes are used to adjust the lead times within a network. Furthermore, restrictions imposed on lead times at customers impact the set of facilities opened within the optimal network design. The configuration of the warehouse layer strongly depends on the quoted order lead times for customers. In addition, the selection of suppliers and the opening of production plants is also dependent on the lead time restrictions. Overall, these observations underline the need to explicitly incorporate lead time restrictions into strategic network design, when the customers of a company value this service criterion.

The presented models do not consider stochastic lead time components and their impact on network designs. Therefore, it is of interest to develop powerful solution methods for solving larger instances of the models presented here and for solving instances of stochastic network design problems that are to be developed.

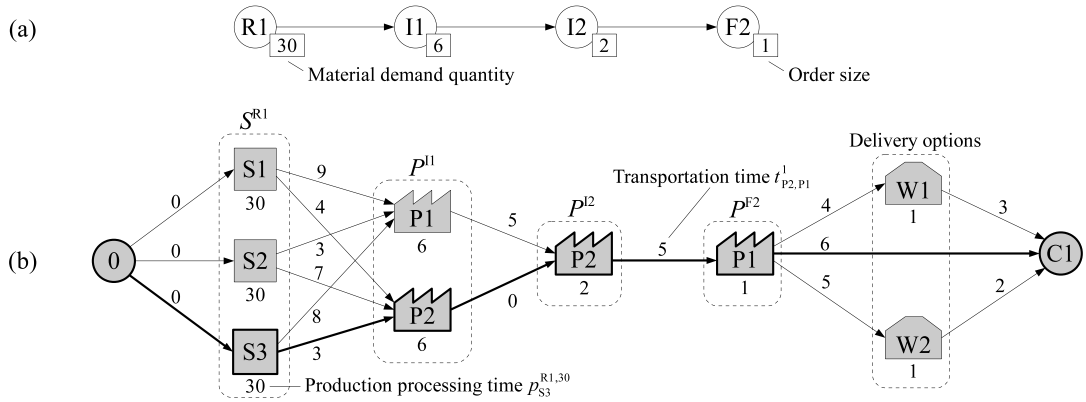

Figure 8: An example for the determination of a minimum order lead time.

Appendix A: An Example for the Determination of Shortest Possible Lead Times

Figure 8 is used for explaining the determination of $MTS_k^{\min}$ at the example of the R1-vertex found at the lowest stage of the MPBOM of Fig. 3a. We determine the shortest possible lead time for transforming this raw material into product F2 and for delivering F2 to customer C1 in the supply network of Fig. 2a. The sequence of production steps for transforming $k = \text{R1}$ into $f = \text{F2}$ corresponds to the highlighted path in Fig. 3a, which is reproduced in Fig. 8a. For each production stage k' involved in this sequence, we have a set of alternative facilities $S^{k'}$ or $P^{k'}$ that could potentially execute the production process. These sets are used to form the directed multi-stage graph of Fig. 8b. Arcs in this graph represent transport links among the facilities. The graph is supplemented by a dummy vertex '0' and by the available delivery options from the only available final-product plant P1 to the customer (either via warehouses W1 or W2 or via direct delivery from P1 to C1). Facility vertices in this graph are weighted with a corresponding production processing time $p_o^{k'q}$. In the example, we set the processing time for producing q units of k' at facility o to $p_o^{k'q} = q$ time units. Arcs are weighted with the transportation time t_{od}^m of the fastest transportation mode existing for a transport link. Given such a weighted directed graph, we determine the shortest path from the dummy source to the customer vertex. The length of this path corresponds to the shortest lead time for transforming k into f and for delivering f to the customer. In the example, the shortest lead time is then achieved if the required 30 units of R1 are produced at supplier S3, transformed into 6 units of I1 at plant P2, further processed into 2 units of I2 at the same plant, and finally transformed into one unit of F2 at plant P1 before being delivered directly from there to customer C1, see the bold path in Fig. 8b. If the length of this path exceeds the maximum lead time $\tau_{\text{C1}}^{\text{F2}}$, the 30 units of R1 needed per ordered unit of F2 must be produced under MTS. If, for example, the customer places orders for $d_{\text{C1}}^{\text{F2}} = 120$ units of F2, at least $MTS_{\text{R1}}^{\min} = 120 \cdot 30 = 3600$ units of R1 must be produced under MTS in any feasible solution to the network design problem.

Appendix B: An Example for the Generation of a Time-Space Network

To illustrate the procedures outlined in Section 5.1, we describe here the generation of the time-space network of Fig. 4c, which results from the problem instance represented by Figs. 4a/b. The sole customer C1 accepts a maximum lead time of $\tau_{\text{C1}}^{\text{F1}} = 8$ time units for receiving an order of $\overline{q} = 1$ unit of product F1. Hence, a vertex $(\text{C1}, \text{F1}, 1, 8)$ is initially created in the time-space network. In order to fulfill an order of customer C1 on time, plant P1 can provide the final product F1 from stock (vertex (P1,F1,1,0)). The replenishment of this production operation can only be made using raw material suppliers operating under MTS, represented by the arcs connecting vertex (P1,F1,1,0) with the supplier vertices at the time coordinate 0. Plant P1 cannot produce under MTO, because deliveries to C1 take $t_{\text{P1,C1}}^{1} = 5$ time units and the remaining 3 time units are insufficient for both replenishing P1 with material supplies from S1 ($t_{\text{S1,P1}}^{1} = 3$) and for processing the order at P1 ($p_{\text{P1}}^{\text{F1},1} = 1$). In the case of plant P2, it can deliver from stock (vertex (P2,F1,1,0)) or under MTO (vertex (P2,F1,1,6)). The MTO option is placed at time 6, given that the final delivery requires $t_{\text{P2,C1}}^{1} = 2$ time units and does not allow for a later finishing of the production operation. This option is feasible, considering that P2 can be replenished with material R1 within 1 time unit (using S2 with transportation mode 1) and with material R2 within 3 time units (using S1 with mode 1); and can also process the production order within $p_{\text{P2}}^{\text{F1},1} = 1$ time unit. For this MTO-operation, suppliers S1 and S2 can provide R1 under either MTS or MTO. Additionally, S2 can select between two available transportation modes. Raw material R2 can only be provided by S1 from stock, because the production of 3 units of R2 under MTO and the delivery to P2 does not allow to replenish the plant until time 6-$p_{\text{P2}}^{\text{F1},1}$=5.

The time-space network enables a timely routing of customer orders through the network facilities. The highlighted subgraph in Fig. 4d illustrates a feasible flow, which corresponds to the fulfillment of an order of product F1 on time. Here, S1 provides 3 units of R2 from stock. S2 produces 1 unit of R1 under MTO and delivers it to P2 within four time units using transportation mode 2. The service outbound time of P2 is 6, which follows from the maximum inbound time (5 time units) plus the production processing time (1 time unit). The final delivery takes two time units and arrives at the customer C1 within the accepted lead time.

Acknowledgments This work was supported by a fellowship within the Postdoc-Program D/09/49358 of the German Academic Exchange Service (DAAD). The authors would also like to gratefully acknowledge the ongoing support provided by the Natural Sciences and Engineering Research Council of Canada.

Bibliography

Arntzen BC, Brown GG, Harrison TP, Trafton LL (1995) Global Supply Chain Management at Digital Equipment Corporation. Interfaces 25(1):69–93

Balinski ML (1961) Fixed-cost transportation problems. Naval Research Logistics Quarterly 8(1):41–54

Cheong MLF, Bhatnagar R, Graves SC (2005) Logistics network design with differentiated delivery lead-time: Benefits and insights. In: Proceedings of 2005 SMA Conference, Singapore, pp 1–20

Cordeau JF, Pasin F, Solomon MM (2006) An integrated model for logistics network design. Annals of Operations Research 144(1):59–82

Daskin MS, Snyder LV, Berger RT (2005) Facility location in supply chain design. In: Langevin A, Riopel D (eds) Logistics Systems: Design and Optimization, Springer US, chap 2, pp 39–65

Erengüç SS, Simpson NC, Vakharia AJ (1999) Integrated production/distribution planning in supply chains: An invited review. European Journal of Operational Research 115(2):219 – 236

Geoffrion AM, Graves GW (1974) Multicommodity distribution system design by benders decomposition. Management Science 20(5):822–844

Goetschalckx M, Vidal CJ, Dogan K (2002) Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. European Journal of Operational Research 143(1):1 – 18

Graves SC, Willems SP (2005) Optimizing the supply chain configuration for new products. Management Science 51(8):1165–1180

Hall NG, Posner ME (2001) Generating experimental data for computational testing with machine scheduling applications. Operations Research 49(6):854–865

Holweg M (2005) The three dimensions of responsiveness. International Journal of Operations & Production Management 25(7):603–622

Klose A, Drexl A (2005) Facility location models for distribution system design. European Journal of Operational Research 162(1):4 – 29

Kohler K (2008) Global Supply Chain Design (in german). Center for Supply Management, Estenfeld (Germany)

Kreipl S, Pinedo M (2004) Planning and scheduling in supply chains: An overview of issues in practice. Production and Operations Management 13(1):77–92

Lin JR, Nozick LK, Turnquist MA (2006) Strategic design of distribution systems with economies of scale in transportation. Annals of Operations Research 144(1):161–180

Meixell MJ, Gargeya VB (2005) Global supply chain design: A literature review and critique. Transportation Research Part E: Logistics and Transportation Review 41(6):531 – 550

Melo MT, Nickel S, Saldanha-da-Gama F (2006) Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Computers & Operations Research 33(1):181 – 208

Melo MT, Nickel S, Saldanha-da-Gama F (2009) Facility location and supply chain management - a review. European Journal of Operational Research 196(2):401–412

Minner S (2003) Multiple-supplier inventory models in supply chain management: A review. International Journal of Production Economics 81-82(1):265–279

Paquet M, Martel A, Montreuil B (2006) Manufacturing network design with reliable promising capabilities. Tech. Rep. Working Paper DT-2006-AM-2, Network Organization Technology Research Center (CENTOR), Université Laval, Québec

ReVelle CS, Eiselt HA, Daskin MS (2008) A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research 184(3):817 – 848

Robinson AG, Bookbinder JH (2007) Nafta supply chains: facilities location and logistics. International Transactions in Operational Research 14(2):179–199

Sabri EH, Beamon BM (2000) A multi-objective approach to simultaneous strategic and operational planning in supply chain design. Omega 28(5):581 – 598

Sourirajan K, Ozsen L, Uzsoy R (2007) A single-product network design model with lead time and safety stock considerations. IIE Transactions 39(5):411–424

Verter V, Dasci A (2002) The plant location and flexible technology acquisition problem. European Journal of Operational Research 136(2):366 – 382

You F, Grossmann IE (2008) Design of responsive supply chains under demand uncertainty. Computers & Chemical Engineering 32(12):3090 – 3111