Flow Formulation of Demand Propagation in Guaranteed Service Models ⋆

Dominik Kamp * **Jörg Rambau** **

* *Chair of Economathematics, University of Bayreuth, 95440 Bayreuth, Germany (e-mail: dominik.kamp@uni-bayreuth.de)*
** *Chair of Economathematics, University of Bayreuth, 95440 Bayreuth, Germany (e-mail: joerg.rambau@uni-bayreuth.de)*

Abstract: In this paper we present two alternative models for the Stochastic Guaranteed Service Model with Demand Propagation (SGSM-DP) in multi-echelon inventory theory. While the original SGSM-DP could so far not be solved for standard networks with more than 15 nodes in 1000 seconds, the new equivalent model F-SGSM-DP based on dynamic flows with arcs going backwards in time can be solved even for heavy instances with up to around 50 nodes.

Keywords: Operation Research, Logistics, and Planning; Comparison of Methods for Modeling; Optimization; Multi-Echelon Inventory Management; Stochastic Programming; Network Flow; Guaranteed Service Model; Mixed-Integer Linear Programming; Demand Propagation

1. INTRODUCTION

This paper is concerned with multi-echelon inventory management facing stochastic demands with deadlines where each stock point has an individual outsourcing option. The general task is to determine safety stocks and outsourcing quantities for all stock points in the network so that all demands can be satisfied in time as well as the expected total cost is minimal.

In the vast literature on multi-echelon inventory management (compare the survey by de Kok et al. (2018)) there are two principal modeling approaches: The *stochastic service models (SSM)* seek to exactly determine stock-out distributions and expected backorder costs. Thus, for large and complicated network structures it is very difficult to solve them exactly. In contrast, the *guaranteed service models (GSM)* use a bounded-demand assumption which allows for solving them even for large network structures (see the survey by Eruguz et al. (2016)). For general demands the GSM is often justified by assuming that unspecified *operational flexibility* (usually expediting and outsourcing) enables the network to act as if the demands were bounded, even if they are not. However, the means of operational flexibility are not accounted for inside the model. The recent line of research on the *Stochastic Guaranteed Service Model (SGSM)* tries to remedy this short-coming by considering expediting and outsourcing as penalized recourse actions in stock-out situations.

The first attempt in Rambau and Schade (2014) led to a basic two-stage stochastic integer linear program with recourse (SGSM), which can be solved by standard mixed-integer linear programming (MILP). The model-optimal solution of the SGSM improved the performance in simulations substantially compared to an ordinary GSM. However, the SGSM model-optimal objective was overestimating the expected cost because it did not account for the upstream demand reduction by outsourcing. Löhnert and Rambau (2018) showed that in order to keep track of the consequences of outsourcing correctly one must compute the demand propagation inside the model. The first logically sound proposition into this direction was the SGSM with Demand Propagation (SGSM-DP) by Löhnert and Rambau (2018). With the new model it could be shown that an endogenous modeling of demand propagation has a serious impact on optimality. We performed an additional test on random instances, where the SGSM optimum was converted to a feasible solution of the SGSM-DP. This experiment confirmed that ignoring this effect can lead to solutions that are more than 50% more expensive on average than the optimum. The problem with the SGSM-DP was its poor scalability. Instances with 20 nodes or more could not reliably be solved by standard MILP-solvers like `cplex`, `gurobi`, or `scip` in, say, 1000 seconds.

In this paper we present two alternative models: the T-SGSM-DP, which is an evolutionary tightening of the SGSM-DP, and the F-SGSM-DP, which uses an all-new modeling approach via variants of dynamic flows. Especially the F-SGSM-DP can solve hard instances with up to 50 nodes in 1000 seconds with the open-source standard MILP-solver `scip` (Gamrath et al., 2020).

2. PROBLEM STATEMENT

According to the short typology by de Kok et al. (2018) this paper studies an inventory problem of the type

$$\mathtt{n}^{\mathrm{ech}},\mathtt{D}^{\mathrm{net}}|\mathtt{I}^{\mathrm{cap}},\mathtt{C}^{\mathrm{del}}|\mathtt{U}^{\mathrm{dem}},\mathtt{G}^{\mathrm{cus}}|\mathtt{P}^{\mathrm{res}}|\mathtt{C}^{\mathrm{obj}}\ .$$

That is: Consider a single-rooted multi-echelon distribution network for a single product type with stochastic demand represented by a finite number of scenarios (which may be the outcome of sampling). The network is operated

⋆ Research supported by the German Research Foundation (DFG) and the Bayreuth research center for modeling and simulation (MODUS), project "The Stochastic Guaranteed Service Model with Demand Propagation" (RA 1033/3-1).

by periodic-review base-stock policies (i.e. an inventory is periodically replenished to a *base-stock level*) together with immediate, unbounded *outsourcing options* in each node. Informally, we search for scenario-independent base-stock levels and scenario-dependent outsourcing quantities such that the expected total cost consisting of holding costs and outsourcing costs is minimal.

Formally: Let $G := (N, A)$ be a *non-empty connected acyclic divergent graph* consisting of a node set N of *inventory locations* including the *unique root node* $1 \in N$ and an arc set A of *supply relations*. That is, for every node $i \in N \setminus \{1\}$ there is exactly one predecessor $j \in N$ supplying i along $(j, i) \in A$. The root node 1 has no predecessor and is supplied externally. For each node $i \in N$ the *lead time* denoted by $L_i \in \mathbb{Z}_{>0}$ is the time required to transport products to this node from its supplier. Let the *demand set* D consist of the nodes without successors. These are the nodes receiving external demand. Let Ω be the *finite scenario set* with *probabilities* $p^\omega \in \mathbb{Q}_{>0}$ for each scenario $\omega \in \Omega$. Any scenario is associated with an *external demand rate* $\alpha_i^\omega \in \mathbb{Z}_{\geq 0}$ for every demand node $i \in D$. Products demanded at $i \in D$ have to be delivered within the *guaranteed end-customer service time* $\overline{s}_i^{\mathrm{OUT}} \in \mathbb{Z}_{\geq 0}$ after the order was placed. At every node $i \in N$, products can be taken either from the internal inventory, which is kept at a *marginal holding cost* $h_i \in \mathbb{Q}_{\geq 0}$ per item or via outsourcing from an alternative supplier at a *marginal outsourcing cost* $c_i \in \mathbb{Q}_{\geq 0}$ per item.

The task is to find:

(1) *Guaranteed service times* s_i^{OUT} and s_i^{IN} for all $i \in N$. That is, node i delivers with a delay of s_i^{OUT} to its customers; its orders are delivered with a delay of at most s_i^{IN} from its regular supplier. This is the usual (S)GSM-setting.
(2) *Outsourcing quantities* q_i^ω for all $i \in N$ in all scenarios $\omega \in \Omega$ with the meaning that node i receives q_i^ω products from its alternative supplier during the time span it waits for replenishment. This is the setting of the (S)GSM with outsourcing options.

The idea of the GSM is that from these values one can derive base-stock levels necessary to guarantee the service times. The SGSM-DP by Löhnert and Rambau (2018) models this problem as two-stage stochastic mixed-integer linear program with recourse (2SMILP). The extensive form of the deterministic equivalent of this 2SMILP with finitely many scenarios is a MILP. However, the scalability of that model does not allow the computation of optimal solutions for networks with more than 15 nodes in reasonable time. This is due to the fact that the SGSM-DP is a "big-M" linearization of a non-linear intermediate model, which is usually cursed with a large integrality gap.

3. TIGHTENED FORMULATION OF THE SGSM-DP

In this section, we enhance the SGSM-DP while keeping its basic structure intact. It turns out that for this it is more instructive to use (a variant of) the non-linear model GSM-o-NL (Löhnert and Rambau, 2018) as our point of departure, since its logic is dominated by the original problem, whereas the SGSM-DP contains purely technical elements only for linearization.

Generalizing the structure of the GSM-o-NL by Löhnert and Rambau (2018) for stochastic demand rates we obtain the non-linear model SGSM-NL:

$$\min \sum_{i \in N} \Big(h_i \cdot y_i + c_i \cdot \sum_{\omega \in \Omega} p^\omega \cdot q_i^\omega \Big) \tag{1}$$

subject to

$$x_i - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} \geq L_i \qquad \forall i \in N \tag{2}$$

$$- s_j^{\mathrm{IN}} + s_i^{\mathrm{OUT}} \leq 0 \qquad \forall (i,j) \in A \tag{3}$$

$$s_i^{\mathrm{OUT}} \leq \overline{s}_i^{\mathrm{OUT}} \qquad \forall i \in D \tag{4}$$

$$y_i + q_i^\omega - x_i \cdot (n_i^\omega + m_i^\omega) \geq 0 \qquad \forall \omega \in \Omega, \forall i \in N \tag{5}$$

$$q_i^\omega - x_i \cdot m_i^\omega = 0 \qquad \forall \omega \in \Omega, \forall i \in N \tag{6}$$

$$q_i^\omega - m_i^\omega \geq 0 \qquad \forall \omega \in \Omega, \forall i \in N \tag{7}$$

$$- \sum_{j \in \delta_G^+(i)} n_j^\omega + n_i^\omega + m_i^\omega = 0 \qquad \forall \omega \in \Omega, \forall i \in N \setminus D \tag{8}$$

$$n_i^\omega + m_i^\omega = \alpha_i^\omega \qquad \forall \omega \in \Omega, \forall i \in D \tag{9}$$

$$s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i \in \mathbb{Z}_{\geq 0} \qquad \forall i \in N \tag{10}$$

$$q_i^\omega \in \mathbb{Z}_{\geq 0} \qquad \forall \omega \in \Omega, \forall i \in N \tag{11}$$

$$n_i^\omega, m_i^\omega \in \mathbb{Q}_{\geq 0} \qquad \forall \omega \in \Omega, \forall i \in N \tag{12}$$

We briefly recall the meanings of the various components: The independent decisions are the *inbound* and *outbound guaranteed service times* s_i^{IN} and s_i^{OUT} as well as the scenario-dependent *outsourcing quantities* q_i^ω for all nodes $i \in N$. The dependent decisions are the *replenishment delays* x_i during which all orders have to be satisfied by inventory and outsourcing, the *propagated demand rates* n_i^ω for the regular supplier, the *outsourced demand rates* m_i^ω for the alternative supplier, and the *base-stock levels* y_i necessary to bridge the time spans x_i. The outsourced demand rates have not appeared in the GSM-o-NL and are defined as

$$m_i^\omega := \begin{cases} \dfrac{q_i^\omega}{x_i} & \text{if } x_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

for all $\omega \in \Omega$ as well as $i \in N$, modeled by (6) and (7).

The objective function (1) consists of the strategic holding cost and the expected outsourcing cost. Restriction (2) bounds x_i to a sufficient replenishment delay, restriction (3) ensures that the transportation of products to the customer starts no earlier than the supplier sends them off, (4) guarantees the exogenous service times of the end customers. These are the restrictions of ordinary GSMs. Moreover, (5) enforces that the base-stock levels are sufficient for the propagated demands, (6) and (7) couple the outsourced demand rates to the outsourcing quantities, (8) and (9) is the demand propagation balance. Here, $\delta_G^+(i)$ is the set of arcs emanating from i. Finally, (10), (11), and (12) specify the domains of the variables.

The outsourcing variables can be eliminated from restriction (5) by subtracting equation (6) in order to handle inventory and outsourcing decisions separately. Experiments with the SGSM-DP on constructed and randomized instances with integral external demand rates led without exception to fully integral optimal solutions. Therefore, we assume that also the demand rates (12) are integral variables (though this is not guaranteed in general). Then, by the following procedure a feasible integral solution can be converted to a feasible integral solution satisfying inequations (2) and (3) as equations.

```
1:  procedure CANONICALIZE
2:      s_1^IN ← 0 ∈ ℤ_{≥0}
3:      for i ∈ N in topological order do
4:          x_i ← min(x_i, s_i^IN + L_i) ∈ ℤ_{≥0}
5:          s_i^OUT ← s_i^IN + L_i − x_i ∈ ℤ_{≥0}        ▷ (2) activated
6:          for j ∈ δ_G^+(i) do
7:              s_j^IN ← s_i^OUT ∈ ℤ_{≥0}                ▷ (3) activated
8:          end for
9:          for ω ∈ Ω do
10:             q_i^ω ← x_i · m_i^ω ∈ ℤ_{≥0}              ▷ (6) restored
11:         end for
12:     end for
13: end procedure
```

Since G is divergent, all time variables and outsourcing quantities are reassigned exactly once. At line 4 the replenishment delay is chosen in order to set at line 5 the corresponding outbound service time to a non-negative value. By induction can be observed that no variable is increased, and therefore this also holds for the objective. For the same reason, inequations (4) as well as (5) remain valid, and since the demand rates (12) are untouched, so are the conservation conditions (8) and (9). Moreover, this operation can not violate (7) because all lead times are assumed to be at least 1, and therefore in line 4 no replenishment delay is strictly decreased to 0, which finally implies (7) by (6). This shows that (2) and (3) can be restricted to equations as well as s_1^{IN} to 0.

This leads to a tightened version, called the T-SGSM-NL:

$$\min \sum_{i\in N}\Big(h_i\cdot y_i + c_i\cdot \sum_{\omega\in\Omega} p^\omega \cdot q_i^\omega\Big) \tag{14}$$

subject to

$$x_i - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} = L_i \qquad \forall i\in N \tag{15}$$

$$-s_j^{\mathrm{IN}} + s_i^{\mathrm{OUT}} = 0 \qquad \forall (i,j)\in A \tag{16}$$

$$s_i^{\mathrm{OUT}} \le \overline{s}_i^{\mathrm{OUT}} \qquad \forall i\in D \tag{17}$$

$$y_i - x_i\cdot n_i^\omega \ge 0 \qquad \forall \omega\in\Omega,\ \forall i\in N \tag{18}$$

$$q_i^\omega - x_i\cdot m_i^\omega = 0 \qquad \forall \omega\in\Omega,\ \forall i\in N \tag{19}$$

$$q_i^\omega - m_i^\omega \ge 0 \qquad \forall \omega\in\Omega,\ \forall i\in N \tag{20}$$

$$-\sum_{j\in\delta_G^+(i)} n_j^\omega + n_i^\omega + m_i^\omega = 0 \qquad \forall \omega\in\Omega,\ \forall i\in N\setminus D \tag{21}$$

$$n_i^\omega + m_i^\omega = \alpha_i^\omega \qquad \forall \omega\in\Omega,\ \forall i\in D \tag{22}$$

$$s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i \in \mathbb{Z}_{\ge 0} \qquad \forall i\in N \tag{23}$$

$$q_i^\omega, n_i^\omega, m_i^\omega \in \mathbb{Z}_{\ge 0} \qquad \forall \omega\in\Omega,\ \forall i\in N \tag{24}$$

With equations (15), (16), and $s_1^{\mathrm{IN}} = 0$, the time domains can be upper-bounded. To quantify this, let the *root lead time* $k_i \in \mathbb{Z}_{>0}$ of node $i\in N$ be defined as the sum of lead times on the unique path from 1 to i. Since all replenishment delays are non-negative, it is $x_i + s_i^{\mathrm{OUT}} \le k_i$. For this reason we call $K_i := \{0,\dots,k_i\}$ the *relevant time set* of i because considering times $x_i, s_i^{\mathrm{OUT}} \in K_i$ is enough to encounter an optimal solution.

Also the demand rates have natural upper-bounds. For $n_i^\omega + m_i^\omega$ in each scenario $\omega\in\Omega$ at every node $i\in N$ the *maximum demand rate* can be computed by

$$M_i^\omega := \begin{cases} \alpha_i^\omega & \text{if } i\in D \\ \sum_{j\in\delta_G^+(i)} M_j^\omega & \text{otherwise} \end{cases} \tag{25}$$

in reverse topological order for each scenario separately.

Using this within a strengthened linearization approach, a tightened version of the original SGSM-DP by Löhnert and Rambau (2018) can be formulated, called T-SGSM-DP:

$$\min \sum_{i\in N}\Big(h_i\cdot y_i + c_i\cdot \sum_{\omega\in\Omega} p^\omega \cdot q_i^\omega\Big) \tag{26}$$

subject to

$$\sum_{k\in K_i} z_{i,k}^{\mathrm{DEL}} = 1 \qquad \forall i\in N \tag{27}$$

$$\sum_{\substack{l\in K_i\\ l<k}} (l-k)\cdot z_{i,l}^{\mathrm{DEL}} + x_{i,k}^- = 0 \qquad \forall i\in N,\ \forall k\in K_i \tag{28}$$

$$\sum_{\substack{l\in K_i\\ l>k}} (k-l)\cdot z_{i,l}^{\mathrm{DEL}} + x_{i,k}^+ = 0 \qquad \forall i\in N,\ \forall k\in K_i \tag{29}$$

$$x_{i,0}^+ - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} = L_i \qquad \forall i\in N \tag{30}$$

$$-s_j^{\mathrm{IN}} + s_i^{\mathrm{OUT}} = 0 \qquad \forall (i,j)\in A \tag{31}$$

$$s_i^{\mathrm{OUT}} \le \overline{s}_i^{\mathrm{OUT}} \qquad \forall i\in D \tag{32}$$

$$y_i - k\cdot n_i^\omega + M_i^\omega\cdot x_{i,k}^- \ge 0 \qquad \forall \omega\in\Omega,\ \forall i\in N,\ \forall k\in K_i \tag{33}$$

$$q_i^\omega - k\cdot m_i^\omega + M_i^\omega\cdot x_{i,k}^- \ge 0 \qquad \forall \omega\in\Omega,\ \forall i\in N,\ \forall k\in K_i \tag{34}$$

$$q_i^\omega - k\cdot m_i^\omega - M_i^\omega\cdot x_{i,k}^+ \le 0 \qquad \forall \omega\in\Omega,\ \forall i\in N,\ \forall k\in K_i \tag{35}$$

$$q_i^\omega - m_i^\omega \ge 0 \qquad \forall \omega\in\Omega,\ \forall i\in N \tag{36}$$

$$-\sum_{j\in\delta_G^+(i)} n_j^\omega + n_i^\omega + m_i^\omega = 0 \qquad \forall \omega\in\Omega,\ \forall i\in N\setminus D \tag{37}$$

$$n_i^\omega + m_i^\omega = \alpha_i^\omega \qquad \forall \omega\in\Omega,\ \forall i\in D \tag{38}$$

$$z_{i,k}^{\mathrm{DEL}} \in \mathbb{B} \qquad \forall i\in N,\ \forall k\in K_i \tag{39}$$

$$x^-_{i,k}, x^+_{i,k} \in \mathbb{Z}_{\geq 0} \qquad \forall i \in N, \forall k \in K_i \tag{40}$$

$$s^{\mathrm{IN}}_i, s^{\mathrm{OUT}}_i, y_i \in \mathbb{Z}_{\geq 0} \qquad \forall i \in N \tag{41}$$

$$q^\omega_i, n^\omega_i, m^\omega_i \in \mathbb{Z}_{\geq 0} \qquad \forall \omega \in \Omega, \forall i \in N \tag{42}$$

Here, variables $z^{\mathrm{DEL}}_{i,k} \in \mathbb{B} := \{0,1\}$ indicate whether $x_i = k$. The new variables $x^-_{i,k} := \max(k - x_i, 0)$ and $x^+_{i,k} := \max(x_i - k, 0)$ measure, by (28) and (29), one-sided differences between prescribed values k and actual values x_i. In particular, $x^+_{i,0}$ is used as the original replenishment delay x_i in condition (30). This allows for a time- and demand-sensitive dimensioning of the "big-M"s as $M^\omega_i \cdot x^-_{i,k}$ and $M^\omega_i \cdot x^+_{i,k}$. And tighter "big-M"s can lead to a stronger formulation of the SGSM-DP.

Some of the constraints dedicated to demand propagation are reminiscent of *network flows* (Ahuja et al., 1993). Indeed, here we are confronted with conservation conditions (37) as well as demand requirements (38). This led to the idea to construct a flow network in which a feasible solution can be represented by constrained flows.

4. A FLOW-BASED REFORMULATION

Following the ideas in (Kamp, 2021), we can reformulate the SGSM-DP as a flow-based ILP using modified versions of commonly known graph transformations, namely *node splitting* and *time expansion* (Ahuja et al., 1993). In order to model the local outsourcing options by flow decisions, at first every node $i \in N$ is *split* into the *inventory node* i, the subsequent *dispatch node* i' and the *outsourcing node* i'' as additional source for the dispatch node. Regular supplies arrive at the inventory node i, whereas outsourced supplies arrive at the outsourcing node i''. Customers are supplied by the dispatch node i', that is, every original arc $(i,j) \in A$ is replaced by the split-arc (i', j). The latter arc is then associated with a duration which is the node's lead time L_i. The other arcs connecting the split-nodes internally obtain an arc duration of zero.

In order to represent the choice of service times and replenishment delays by an arc selection, we construct a *time-expanded network* from this split-node graph. For each split-node belonging to a node $i \in N$ we create for every relevant time slot in K_i an indexed copy. As usual, arcs in the time-expanded split-node graph connect all copies of split-nodes whose time-indices differ by the corresponding split-arc duration in the underlying graph. Additionally, we introduce a common flow source given by the global supply node $0'_0$ and define $K_0 := \{0\}$. This node is connected to the externally supplied root inventory node 1_{L_1} as well as to every outsourcing node for which there is an expanded split-arc terminating in an associated inventory node one time unit later.

The network so far is only able to model a physical product flow without any intermediate inventories. However, the whole point of an inventory network is to accelerate the supply process by the use of inventories. Our new idea is to represent these options by *acceleration arcs* going one unit *backwards* in time. We introduce these arcs between the inventory nodes and between the outsourcing nodes. This way, satisfying an order from inventory is represented by connecting the physical replenishment flow entering the inventory node after the replenishment delay logically to the physical supply flow leaving the inventory node at the outbound service time. The number of traversed acceleration arcs is equal to the replenishment delay. Moreover, we would like to have a canonical sink node for the flow satisfying a demand. This node is naturally given by the dispatch node copy at the guaranteed end-customer service time slot antedated into the relevant time set. Since products may arrive at an earlier copy, we introduce waiting arcs going forwards in time between consecutive demand dispatch nodes.

Figure 1 shows how a node is split into inventory, dispatch, and outsourcing node; moreover, it illustrates the time-expansion of a two-echelon serial inventory system including consistent flows for external demand rates $\alpha^1_2 := 1$ and $\alpha^2_2 := 2$. All drawn arcs carry unit flows. In the second scenario half of the external demand is immediately outsourced. This makes the depicted flows feasible for base levels at least $y_1 = 2$ and $y_2 = 1$ because the arcs going backwards in time represent inventory withdrawals, which are replenished two time units delayed at the supplier node and one time unit delayed at the customer node.

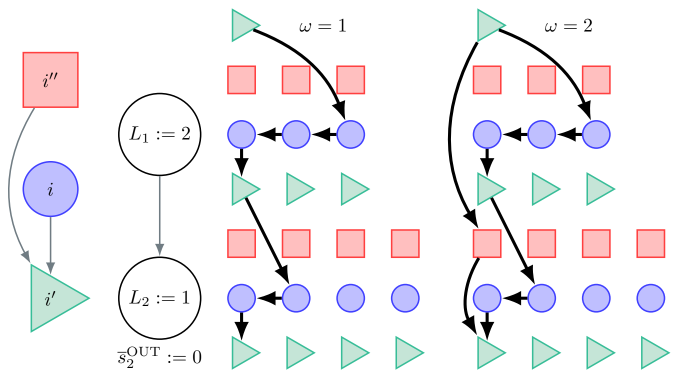

Fig. 1. Node split and two-echelon time-expanded flows

Formally, the resulting network G' is defined by:

$$G' := (N', A') \tag{43}$$

$$N' := \{i_k, i'_k, i''_k \mid i \in N, k \in K_i\} \dot\cup \{0'_0\} \tag{44}$$

$$A' := F \dot\cup E \tag{45}$$

$$F := \{(i_k, i'_k), (i''_k, i'_k) \mid i \in N, k \in K_i\} \tag{46}$$

$$\dot\cup \{(i'_k, j_{k+L_j}) \mid (i,j) \in A \dot\cup \{(0,1)\}, k \in K_i\} \tag{47}$$

$$\dot\cup \{(0'_0, j''_{k+L_j-1}) \mid (i,j) \in A \dot\cup \{(0,1)\}, k \in K_i\} \tag{48}$$

$$E := \{(i_k, i_{k-1}), (i''_k, i''_{k-1}) \mid i \in N, k \in K_i \setminus \{0\}\} \tag{49}$$

$$\dot\cup \{(i'_{k-1}, i'_k) \mid i \in D, k \in K_i \setminus \{0\}\} \tag{50}$$

$$D' := \{i'_{\min(\bar{s}^{\mathrm{OUT}}_i, k_i)} \mid i \in D\} \tag{51}$$

Additionally, for a compact demand notation we define

$$\alpha^\omega_i := \begin{cases} -M^\omega_1 & \text{if } i = 0'_0 \\ \alpha^\omega_j & \text{if } i = j'_k \in D' \\ 0 & \text{else} \end{cases} \tag{52}$$

in every scenario $\omega \in \Omega$ for all expanded nodes $i \in N'$.

Then, a flow-based formulation of the SGSM-DP with flow variables $x^{\omega}_{i_k,j_l}$, called F-SGSM-DP, is given by:

$$\min \sum_{i\in N}\Big(h_i \cdot y_i + c_i \cdot \sum_{\omega\in\Omega} p^{\omega} \cdot q^{\omega}_i\Big) \tag{53}$$

$$\sum_{k\in K_i} z^{\text{OUT}}_{i,k} = 1 \qquad \forall i \in N \tag{54}$$

$$x^{\omega}_{i_k,i'_k} + x^{\omega}_{i''_k,i'_k} - M^{\omega}_i \cdot z^{\text{OUT}}_{i,k} \le 0 \qquad \forall \omega \in \Omega,\ \forall i \in N,\ \forall k \in K_i \tag{55}$$

$$x^{\omega}_{i'_k,j_{k+L_j}} + x^{\omega}_{0'_0,j''_{k+L_j-1}} - M^{\omega}_j \cdot z^{\text{OUT}}_{i,k} \le 0 \qquad \forall \omega \in \Omega,\ \forall (i,j) \in A,\ \forall k \in K_i \tag{56}$$

$$\sum_{k\in K_i\setminus\{0\}} x^{\omega}_{i_k,i_{k-1}} \le y_i \qquad \forall \omega \in \Omega,\ \forall i \in N \tag{57}$$

$$\sum_{k\in K_i\setminus\{0\}} (x^{\omega}_{i''_k,i''_{k-1}} + x^{\omega}_{i''_{k-1},i'_{k-1}}) = q^{\omega}_i \qquad \forall \omega \in \Omega,\ \forall i \in N \tag{58}$$

$$\sum_{j\in\delta^-_{G'}(i)} x^{\omega}_{j,i} - \sum_{j\in\delta^+_{G'}(i)} x^{\omega}_{i,j} = \alpha^{\omega}_i \qquad \forall \omega \in \Omega,\ \forall i \in N' \tag{59}$$

$$z^{\text{OUT}}_{i,k} \in \mathbb{B} \qquad \forall i \in N,\ \forall k \in K_i \tag{60}$$

$$y_i \in \mathbb{Z}_{\ge 0} \qquad \forall i \in N \tag{61}$$

$$q^{\omega}_i \in \mathbb{Z}_{\ge 0} \qquad \forall \omega \in \Omega,\ \forall i \in N \tag{62}$$

$$x^{\omega}_{i,j} \in \mathbb{Z}_{\ge 0} \qquad \forall \omega \in \Omega,\ \forall (i,j) \in A' \tag{63}$$

Here, (53) is as before; (54), (55), and (56) allow physical flow only at the time slots corresponding to the outbound service times selected by $z^{\text{OUT}}_{i,k}$; (57) ensures sufficient inventory for withdrawals along the acceleration arcs; (58) sums up outsourcing rates to outsourcing quantities; (59) is the flow balance with in- and outgoing arc sets $\delta^-_{G'}(i)$ and $\delta^+_{G'}(i)$ respectively; finally, (60), (61), (62), and (63) specify the domains of the variables.

Theorem 1. T-SGSM-DP and F-SGSM-DP are equivalent.

Proof. We restrict ourselves to a rough sketch. Given an optimal solution of one of the models, it is possible to

(1) construct a (not necessarily feasible) solution of the other model with identical objective.
(2) apply a variant of CANONICALIZE to find a feasible solution without increasing the cost.

Technical details are straight-forward and are omitted. □

5. PERFORMANCE AND SCALABILITY

We generated two sets of benchmark instances for multi-echelon networks with an increasing number of nodes, random topology, and random data. The random topology given a number of nodes is generated by adding the nodes one-by-one, starting at the root node 1 and connecting each new node to one already present node chosen uniformly at random. The result is a divergent network, where the demand nodes are chosen to be the nodes without successors. The random data concerns lead times, end-customer service times, probabilities, demands, holding costs, and outsourcing costs. The latter are chosen to be strictly larger than the former to avoid trivial instances. We restrict ourselves to three scenarios because Rambau and Schade (2014) have shown that few scenarios (obtained by heavy sampling and subsequent heavy, asymmetric scenario reduction) can accurately represent the stochasticity of the system.

The solutions were computed on a standard MacBook Air (11 Inch, Mid 2012, macOS Catalina 10.15.7, 2.6 GHz Core i5-3317U, 4 GB DDR3-RAM) using the open-source MILP-solver SCIP 7.0.3 including the sub-LP-solver SoPlex 5.0.2 (Gamrath et al., 2020) with a time limit of 1000 seconds for each problem instance.

The first set of instances was generated for 2 through 30 nodes and three scenarios. The remaining data was drawn from the following sets:

$$L_i \in \{1, \ldots, 4\} \tag{64}$$

$$h_i \in \{1, 2\} \tag{65}$$

$$c_i \in \{h_i + 1, \ldots, h_i + 8\} \tag{66}$$

$$\bar{s}^{\text{OUT}}_i \in \{0, 1\} \tag{67}$$

$$p^{\omega} \in \{1, \ldots, 100\} \text{ normalized} \tag{68}$$

$$\alpha^{\omega}_i \in \{1, \ldots, \lceil 4 \cdot \tfrac{\omega}{|\Omega|} \rceil\} \tag{69}$$

Figures 2 and 3 show the total computation times and relative integrality gaps $(1 - \frac{\text{LP-Opt}}{\text{ILP-Opt}})$ of the models.

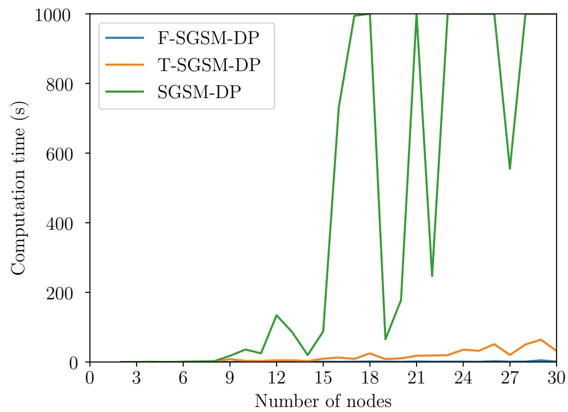

Fig. 2. Comparison of computation times I

The result is that, in contrast to the original SGSM-DP, the new models of this paper, the T-SGSM-DP and the F-SGSM-DP, can reliably solve the instances with more than 15 nodes far below the time limit. A plausible reason is the consistent difference among the integrality gaps.

The second set of instances was generated for 2 through 50 nodes and again three scenarios. For a "stress-test" the data was drawn from the substantially larger sets:

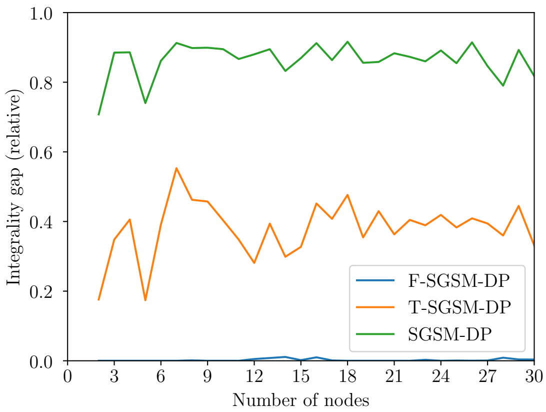

Fig. 3. Comparison of integrality gaps I

$$L_i \in \{1, \ldots, 31\} \tag{70}$$
$$h_i \in \{1, \ldots, 31\} \tag{71}$$
$$c_i \in \{h_i + 1, \ldots, h_i + 279\} \tag{72}$$
$$\overline{s}_i^{\mathrm{OUT}} \in \{0, \ldots, 30\} \tag{73}$$
$$p^\omega \in \{1, \ldots, 100\} \text{ normalized} \tag{74}$$
$$\alpha_i^\omega \in \{1, \ldots, \lceil 62 \cdot \tfrac{\omega}{|\Omega|} \rceil\} \tag{75}$$

Figures 4 and 5 show the resulting times and gaps.

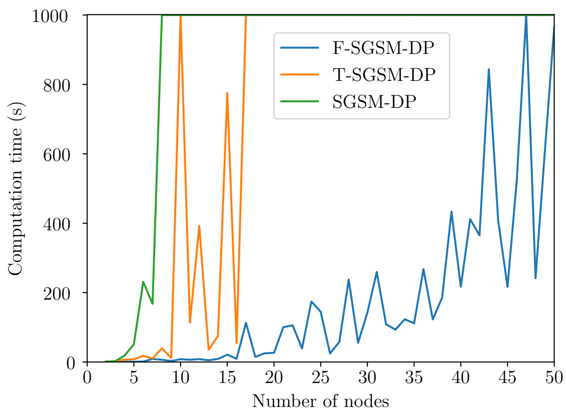

Fig. 4. Comparison of computation times II

Here, not only the SGSM-DP but also the T-SGSM-DP takes a serious dip in performance, while the F-SGSM-DP maintains a smaller integrality gap which leads to substantially lower computation times throughout reaching the time limit only for one of the largest instances. And this even though the number of variables in the F-SGSM-DP grows as large as 100,000 as opposed to at most 6,000 in the SGSM-DP. This shows again that the tightness of a model is sometimes more important than a compact size.

6. CONCLUSION

We presented two new mixed-integer linear models for the SGSM-DP with outsourcing option, one as a technically tightened version (T-SGSM-DP) and one based on time-expanded flows (F-SGSM-DP). The main idea is that inventory withdrawals can be modeled as logical flows backwards in time because they accelerate the net delivery from the unique supplier to the end customers. The flow-based model outperforms both the original and the tightened model by a large margin. Using the F-SGSM-DP, for the first time optimal solutions for multi-echelon inventory networks with 50 nodes and lead times up to 31 time units can be computed. Future research goes into two directions: First, investigate dynamic column generation in order to mitigate the influence of the number of variables in the F-SGSM-DP. Second, generalize the F-SGSM-DP for inventory networks containing convergent substructures.

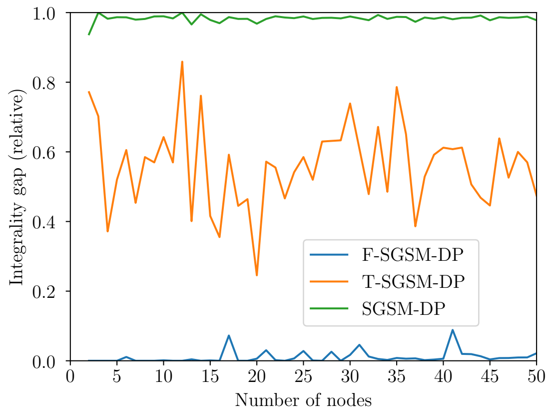

Fig. 5. Comparison of integrality gaps II

REFERENCES

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993). *Network Flows*. Prentice Hall, New Jersey.

de Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J., and Schade, K. (2018). A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3), 955–983. doi:10.1016/j.ejor.2018.02.047.

Eruguz, A.S., Sahin, E., Jemai, Z., and Dallery, Y. (2016). A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization. *International Journal of Production Economics*, 172, 110–125.

Gamrath, G., Anderson, D., Bestuzheva, K., Chen, W.K., Eifler, L., Gasse, M., Gemander, P., Gleixner, A., Gottwald, L., Halbig, K., Hendel, G., Hojny, C., Koch, T., Le Bodic, P., Maher, S.J., Matter, F., Miltenberger, M., Mühmer, E., Müller, B., Pfetsch, M.E., Schlösser, F., Serrano, F., Shinano, Y., Tawfik, C., Vigerske, S., Wegscheider, F., Weninger, D., and Witzig, J. (2020). The scip optimization suite 7.0. Technical report, Zuse Institut Berlin.

Kamp, D. (2021). *The Guaranteed Service Model with Demand Propagation – New Models based on Flow Networks*. Master's thesis, University of Bayreuth.

Löhnert, M. and Rambau, J. (2018). Modeling demand propagation in guaranteed service models. *International Federation of Automatic Control - PapersOnLine*, 51(2), 284–289. doi:10.1016/j.ifacol.2018.03.049.

Rambau, J. and Schade, K. (2014). The stochastic guaranteed service model with recourse for multi-echelon warehouse management. *Mathematical Methods of Operations Research*, 79(3), 293–326. doi:10.1007/s00186-014-0464-y.