Cloud manufacturing service selection optimization and scheduling with transportation considerations

Int J Adv Manuf Technol (2018) 95:43–70
DOI 10.1007/s00170-017-1167-3

ORIGINAL ARTICLE

Cloud manufacturing service selection optimization and scheduling with transportation considerations: mixed-integer programming models

Hossein Akbaripour[1] · Mahmoud Houshmand[1] · Tom van Woensel[2] · Nevin Mutlu[2]

Received: 11 August 2017 / Accepted: 21 September 2017 / Published online: 11 October 2017

Abstract Cloud manufacturing is an emerging service-oriented manufacturing paradigm that integrates and manages distributed manufacturing resources through which complex manufacturing demands with a high degree of customization can be fulfilled. The process of service selection optimization and scheduling (SSOS) is an important issue for practical implementation of cloud manufacturing. In this paper, we propose new mixed-integer programming (MIP) models for solving the SSOS problem with basic composition structures (i.e., sequential, parallel, loop, and selective). Through incorporation of the proposed MIP models, the SSOS with a mixed composition structure can be tackled. As transportation is indispensable in cloud manufacturing environment, the models also optimize routing decisions within a given hybrid hub-and-spoke transportation network in which the central decision is to optimally determine whether a shipment between a pair of distributed manufacturing resources is routed directly or using hub facilities. Unlike the majority of previous research undertaken in cloud manufacturing, it is assumed that manufacturing resources are not continuously available for processing but the start time and end time of their occupancy interval are known in advance. The performance of the proposed models is evaluated through solving different scenarios in the SSOS. Moreover, in order to examine the robustness of the results, a series of sensitivity analysis are conducted on key parameters. The outcomes of this study demonstrate that the consideration of transportation and availability not only can change the results of the SSOS significantly, but also is necessary for obtaining more realistic solutions. The results also show that routing within a hybrid hub-and-spoke transportation network, compared with a pure hub-and-spoke network or a pure direct network, leads to more flexibility and has advantage of cost and time saving. The level of saving depends on the value of discount factor for decreasing transportation cost between hub facilities.

Keywords Cloud manufacturing · Service selection and scheduling · Service occupancy · Hub-and-spoke transportation network · Mixed-integer programming

✉ Mahmoud Houshmand
hoshmand@sharif.edu

[1] Industrial Engineering Department, Sharif University of Technology, Tehran, Iran

[2] School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

1 Introduction

The manufacturing industry is undergoing a major transformation due to evolving customer expectations. Traditional product-oriented manufacturing systems allow customers to purchase finished products for which customer involvement in design, engineering, or manufacturing process is at minimum. In today's customer-centric environment, though, customers expect to see customization options in order to have access to personalized items that fulfill their unique needs. For example, to meet customer expectations, the number of varieties in automobile vehicle models in the USA was increased from 44 in 1969 to 165 in 2006 [23]. According to a study by [7], almost 81% of motorcyclists prefer to have the motorcycle seats made-to-order. Magnusson and Pasche [40] investigated a forklift truck manufacturing company which has increased its customized products up to 25%. Small and medium-sized enterprises (SMEs) have proven to exhibit a great potential to fulfill customers' evolving expectations with their diverse functionalities, flexibility due to their smaller size compared to large manufacturing enterprises, and potential to develop service-oriented businesses [24, 25, 52, 64]. However, the lack of powerful platforms that allow (1) interactions between customers and SMEs [1], and (2) collaboration among geographically dispersed SMEs in a real-time, on-demand, dynamic setting [15, 21], poses challenges in the transformation process of shifting toward a service-oriented manufacturing paradigm. Cloud manufacturing provides an attractive solution for these problems by enabling integration of manufacturing resources of partner SMEs such that they can perform complex manufacturing tasks cooperatively with a high degree of customization [30, 42, 67], a process that is impossible to perform in the standard large-scale manufacturing systems. This platform is an on-demand search and recommendation tool that identifies all kinds of manufacturing services (such as design, engineering, machining, testing, and packing) in the product lifecycle to satisfy customer requirements for customized manufacturing tasks. However, optimal selection and scheduling of these services are still quite challenging, especially considering the transportation aspect across geographically dispersed SME locations. In this study, we attempt to answer the following research questions:

(1) How to model service selection and scheduling for cloud manufacturing tasks with different subtask composition structures.
(2) How to optimally route a shipment between geographically dispersed SMEs using hybrid hub-and-spoke transportation network.
(3) How to deal with service occupancy and how to schedule subtasks onto available services.

Cloud manufacturing, which is also called "cloud-based design and manufacturing" [63] or "cloud-based global supply chain" [1], is a new service-oriented business model, which is enabled by the recent advances in the field of information technology. Cloud manufacturing utilizes the Internet, big data, the Internet of Things (IoT), cloud computing, etc., to virtualize and integrate distributed manufacturing resources into cloud services and manage them under intelligent control of a centralized manufacturing platform [21, 39, 47, 48, 54, 62, 67, 71]. By interconnecting geographically distributed manufacturing resources of SMEs with various functionalities, cloud manufacturing facilitates cooperation, coordination, and information sharing among partner SMEs. This, in turn, promotes interoperability among SMEs through innovation and enables them to maintain a competitive advantage in the global marketplace as an integrated global virtual enterprise [21, 60].

In a cloud manufacturing setting, a manufacturing cloud service (MCS) is an encapsulation of one or more physical manufacturing resources [36]. A resource can perform a certain manufacturing function involved in the product lifecycle [47] and is in the form of hardware (e.g., welding machine) or software (e.g., data analysis tools) [24, 25, 67]. Cloud manufacturing systems can perform two types of manufacturing tasks: a single-functionality manufacturing task and a multi-functionality manufacturing task. While the former can be performed using a single MCS, the latter needs an iterative task decomposition process for the purpose of dividing a task into a series of subtasks, such that at least one candidate MCS for each subtask can be offered. The critical issue with cloud manufacturing is to complete a multi-functionality task optimally through selection and composition of MCSs among functionally equivalent MCSs within their available time frames by also satisfying constraints including transportation and quality of service (QoS). In a service-oriented environment, the aforementioned issues are tackled by solving QoS-aware service selection optimization and scheduling (SSOS) problem [13, 38]. Also, it is studied as service composition and optimal selection (SCOS) when the objective is service composition rather than service scheduling [65, 74].

As opposed to services in some service-oriented environments such as cloud computing, which can be delivered over the Internet or virtual network, cloud manufacturing necessitates physical transportation between MCSs, hence, additional constraints to existent service composition and scheduling problems, in various geographical locations [1, 30]. Although SSOS and SCOS problems in other service-oriented systems including cloud computing [28, 49, 75] and manufacturing grid [55, 57] have been well studied, majority of the proposed approaches cannot be applied to cloud manufacturing systems due to different QoS metrics and lack of transportation consideration between manufacturing services, which can change the results dramatically. Recently, transportation has been considered within the context of cloud manufacturing service composition [13, 38, 72]. However, these works address only the sequential subtask (or service) composition structure.

In this paper, the SSOS problem is studied with the consideration of transportation and different composition structures with service occupancy constraints. The main contributions of this work are as follows. We develop new mixed-integer programming (MIP) models for the SSOS problem in cloud manufacturing with not only a sequential composition structure but also with other three basic structures including parallel, loop, and selective, which are commonly observed in many manufacturing processes. Second, to the best of our knowledge, this paper is the first to incorporate routing optimization within a given hybrid hub-and-spoke transportation network in the context of cloud manufacturing. This idea allows the proposed models to trade-off cost, time, etc., and route a shipment between pairs of distributed manufacturing services using a direct or the hub-and-spoke transportation network. Third, we also incorporate service occupancy constraints in the proposed MIP models which enables us to obtain more realistic solutions.

The remainder of this paper is organized as follows. In Section 2, we provide a review of the related literature. A simple example is presented in Section 3 to motivate the structure of MIP models in the subsequent section. The SSOS problem formulation and mathematical models of the service selection and scheduling for the different subtask composition structures are established in Section 4. Computational results are given in Section 5 and Section 6 concludes this study and suggests directions for future research.

2 Literature review

In order to ensure that a cloud manufacturing system can perform a complicated and customized manufacturing task, solving the SSOS problem efficiently is becoming increasingly important. It is more complex compared to several kinds of traditional shop scheduling or task scheduling problems [11, 69], especially when transportation between different geographically dispersed locations needs to be taken into account in the problem solving. The related literature spans three main streams of research: (1) scheduling, (2) SCOS and SSOS, and (3) transportation and routing.

In general, scheduling deals with the assignment of limited resources to activities over time such that one or more predefined objectives are optimized. The research on scheduling originated from static scheduling approaches which usually were used to develop shop floor scheduling systems for mass production. Manne [43] proposed a discrete linear programming model for solving the typical job-shop scheduling problem. Branch-and-bound and branch-and-cut algorithms were developed by Laporte et al. [29] to solve job sequencing problem. Banaszak and Zaremba [9] proposed a heuristic method for integration of process planning and scheduling in virtual manufacturing systems. Kesen et al. [27] addressed job scheduling in virtual manufacturing cells as a multi-objective MIP model in which summation of weighted makespan and weighted total traveling distance were minimized. A bi-objective genetic algorithm was developed by Arkat and Ghahve [6] in order to minimize the makespan and the total cost of inter- and intra-plant transportation in virtual manufacturing cells. As traditional scheduling algorithms rarely consider the virtualization characteristics of MCSs such as quality and reliability, their practicality cannot be assumed to be effective in cloud manufacturing environment [73].

SCOS is a process of service composition and match between MCSs and (sub)tasks. In SSOS, in addition to this process, it is determined when a (sub)task starts or ends depending on some parameters such as its duration, predecessor (sub)tasks, composition structure, and MCSs' availability. Both of SCOS and SSOS are NP-complete [49, 70, 74]. Although many studies have been conducted on SCOS, research about SSOS in cloud manufacturing is just beginning. Table 1 summarizes the related literature of SCOS and SSOS. For solving the SCOS problem, Tao et al. [56] proposed full connection-based parallel adaptive chaos optimization with reflex migration as a novel parallel intelligent algorithm. Liu et al. [37] tackled the problem of multi-task oriented SCOS with a new composition pattern for improving the success rate of QoS requirement satisfaction. Tian et al. [58] solved the SCOS problem for motorcycle design and assembly application using a discrete hybrid bee colony algorithm. The SCOS problem based on QoS and energy consumption was modeled as a multi-objective problem by Xiang et al. [65]. Wang et al. [61] addressed the selection strategy of machining equipment in cloud manufacturing system and used an improved particle swarm optimization algorithm to help clients effectively select the machining equipment. In Jin et al. [26], a genetic algorithm-based approach was applied to investigate the SCOS when there are potential quality correlations among MCSs. Lartigau et al. [30] extended the scope of the SCOS problem to the geographical locations of manufacturing services taking geo-perspective transportation into consideration and optimized the problem using an adapted artificial bee colony algorithm. The issue of measuring fuzzy QoS and selecting the best MCSs considering design preference was studied by Zheng et al. [72]. Zhou and Yao [74] discussed reputation evaluation during QoS-aware service composition process by introduction of a time decay function. Liu and Zhang [33] provided an approach which allows functionally equivalent MCSs to establish a group of synergistic services for performing each manufacturing subtask cooperatively. For the purpose of considering both horizontal and sequential collaboration between MCSs, networked collaboration-based QoS evaluation model was proposed [68]. Xiang et al. [66] suggested two phases based on case library for the SCOS with large-scale composed solutions. Many of the above works have focused on the service composition (instead of service scheduling) perspective of cloud manufacturing only by supposing a common assumption that MCSs are available all the time and can be used upon request. However, this assumption does not hold true in many real-world scenarios because an MCS' availability is constantly changing over time based on previously assigned tasks, unpredictable failures, and so on.

Cheng et al. [17] investigated multi-task-oriented SSOS in cloud manufacturing system taking correlation among virtual resources into account to provide more profit for an enterprise with deadline-constrained delivery. In this work, the necessity of transportation between manufacturing services was neglected. Li et al. [31] addressed task scheduling for cloud manufacturing resources with cost minimization, load balancing, and processing time minimization as objective functions. Different from most of the literature on cloud manufacturing service composition, Cao et al. [13] explicitly considered service occupancy during the process of SSOS. They ingeniously converted the process into a multi-stage

Table 1 A chronological literature review on SCOS and SSOS at a glance

Author(s)	Consideration of		Composition structure				Objectives functions				Algorithm
	Scheduling	Transportation	Sequential	Parallel	Loop	Selective	Cost	Time	Quality	etc.	
Tao et al. [56]	–	–	✓	✓	–	–	✓	✓	–	✓	Chaos optimization
Liu et al. [37]	–	–	✓	✓	✓	✓	✓	✓	–	✓	Genetic algorithm
Tian et al. [58]	–	–	✓	✓	✓	✓	✓	✓	–	✓	Artificial bee colony
Xiang et al. [65]	–	–	✓	✓	✓	✓	✓	✓	–	✓	Group leader algorithm
Wang et al. [61]	–	✓	✓	–	–	–	✓	✓	✓	–	Particle swarm optimization
Cheng et al. [17]	✓	–	✓	–	–	–	✓	✓	–	–	Genetic algorithm
Jin et al. [26]	–	–	✓	✓	✓	✓	✓	✓	–	✓	Genetic algorithm
Lartigau et al. [30]	–	✓	✓	✓	✓	✓	✓	✓	–	✓	Artificial bee colony
Li et al. [31]	✓	✓	✓	✓	–	–	✓	✓	–	✓	Subtask-scheduling heuristic
Zheng et al. [72]	–	✓	✓	–	–	–	✓	✓	–	✓	Particle swarm optimization
Zhou and Yao [74]	–	–	✓	–	–	–	✓	✓	–	✓	Artificial bee colony
Liu and Zhang [33]	–	–	✓	✓	–	–	✓	✓	–	✓	Genetic algorithm
Xue et al. [68]	–	–	✓	–	–	–	✓	✓	✓	✓	Genetic-artificial bee colony
Xiang et al. [66]	–	–	✓	✓	–	–	–	–	–	✓	Case-library-based heuristic
Cao et al. [13]	✓	✓	✓	–	–	–	✓	✓	✓	✓	Ant colony optimization
Liu et al. [38]	✓	✓	✓	–	–	–	✓	✓	✓	✓	Workload-based heuristic
The current study	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	Branch-and-Cut

graph and solved it using a modified ant colony optimization algorithm. Very recently, based on service composition idea, a workload-based multi-task scheduling model for cloud manufacturing was designed by Liu et al. [38] that incorporates workload modeling, service efficiency, and quantity as well as transportation issues. The developed models in the last three studies dealt with only particular service composition structures and their credibility cannot be maintained for all basic structures especially in the present of transportation.

Although the notion of transportation is a fundamental requirement in the SCOS and SSOS, according to Table 1, there are few works that have considered its effect on QoS metric and the process of service selection and scheduling. In addition, these limited works usually address only a “direct transportation” network, in which a point-to-point transportation between different locations is realized [41, 45]. Another alternative strategy that often emerges in practical situations is a hub-and-spoke network, where hubs as consolidation facilities provide indirect transportation between spokes (i.e., non-hubs) [12, 46]. A direct and hub-and-spoke network have the advantage of fast speed [32] and economies of scale [20], respectively. To gain benefit from both types of transportation network, a hybrid hub-and-spoke network can be used wherein the most important decision is whether to route a shipment via a hub or directly to its destination [22]. Many authors have investigated this routing problem which bears some relevance to this study. Aykin [8] developed a mathematical formulation for determining the hub locations and routing decisions together. Liu et al. [34] studied a mixed truck delivery system that allows both direct and hub-and-spoke transportation and delineated vehicle routings using a heuristic algorithm. Hsu and Hsieh [22] formulated direct versus hub-and-spoke routing problem as a two-objective model and determined Pareto optimal solutions based on a trade-off between transportation and inventory costs. Çetiner et al. [14] established an iterative two-stage algorithm so that in the first stage, hub locations are specified and the second stage solves routing problems. Then, the algorithm iterates between two stages with a heuristic updating mechanism to attain a route-compatible hub network. Mahmutoğulları and Kara [41] presented mathematical models for the different versions of hub location problem with allowed direct transportation between spokes.

In conclusion, there is still a considerable gap between the SSOS requirements (regarding transportation involvement, basic composition structures, and service scheduling especially when services are not available all the time) and the efficiency of solutions to fulfill them. Thus, the SSOS needs further research and investigation and cloud manufacturing communities need to consider realistic points in their proposed models to make cloud manufacturing commercially feasible in the near future.

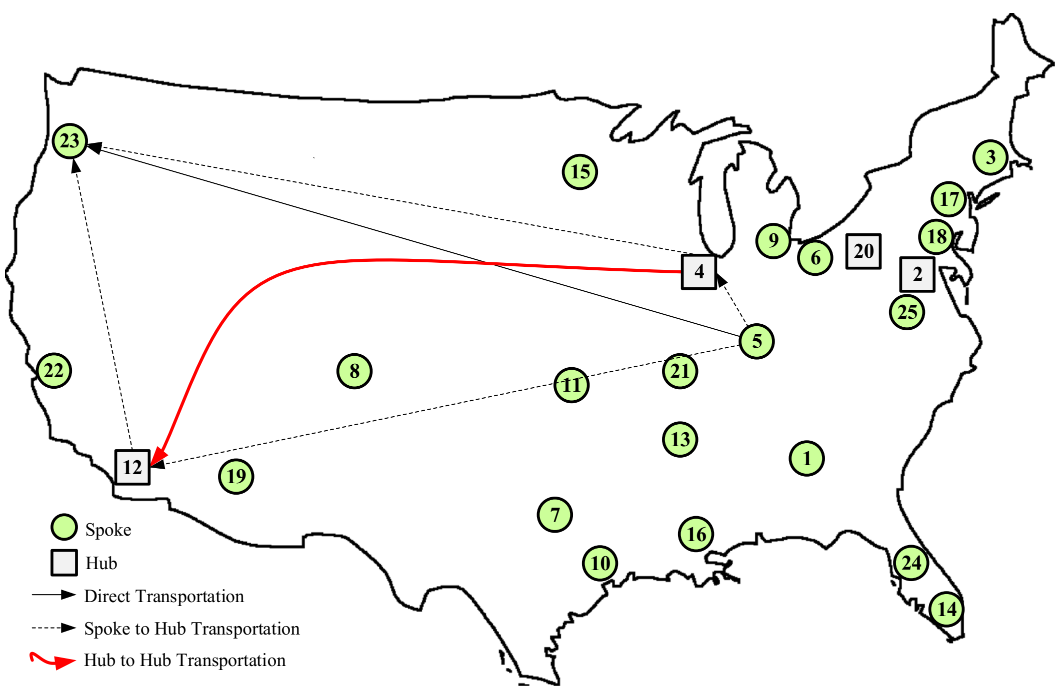
3 An example

Assume that a cloud manufacturing platform provides a service-oriented collaborative manufacturing system to 25

SMEs working in the motorcycle production industry in the USA. Each SME is located in a different city and is equipped with one or two manufacturing resources. We use the well-known Civil Aeronautics Board (CAB) dataset [46] for city locations and inter-city distances; see the Appendix. As shown in Fig. 1, Baltimore, Chicago, Los Angeles, and Pittsburg are hub cities and a shipment between two different cities can be routed through direct or hub-and-spoke network. An example of online motorcycle production (OMP) is considered to illustrate the problem setting. A customer submits an OMP task, denoted by T_{OMP}, and requests customized motorcycle for less than \$$C_{\max}$, within $T_{\max}$ days, and average quality level of at least $Q_{\min}$% which for instance can be obtained using quality grading system shown in Table 2.

For satisfying the T_{OPM}, according to Fig. 2, the platform executes the below three main processes:

- *Task decomposition*: the task is firstly decomposed into several subtasks such that at least one candidate MCS for each subtask can be provided. For the T_{OMP} example, 11 subtasks have been extracted (Fig. 3) such that some subtasks have parallel structures, some are in loop and so on.
- *Service discovery and matching*: the platform performs this process to find out all candidate MCSs for each subtask. Next, they are pooled into Cloud Manufacturing Qualified Service Set (CMQSS). In Table 3, we have designed the output of service discovery and matching process for the T_{OMP}. Note that the CMQSS for each subtask includes those MCSs with given manufacturing cost and/or manufacturing time.
- *Service selection optimization and scheduling* (*SSOS*): the platform simultaneously selects an MCS from corresponding CMQSS of each subtask, schedules subtasks onto the available time frames of selected MCSs, and routes required transportations in a given hybrid hub-and-spoke network such that the overall QoS of the manufacturing composite cloud service (MCCS) is optimized and the associated constraints are fulfilled. Theoretically, by multiplying the number of candidate MCSs for all the subtasks, the total number of possible MCCSs can be calculated which is equal to $\prod_{s=1}^{11}|R_s| = 4 \times 5 \times \cdots \times 2 = 172,800$ for the T_{OMP} example. Note that R_s denotes the CMQSS for the sth subtask. In order to find the optimal MCCS among all possible MCCSs, the platform needs a suitable model that not only is valid for a task with a mixed subtasks' composition structure, but also can optimize transportation routes and is able to consider service occupancy.



01 Atlanta	06 Cleveland	11 Kansas City	16 New Orleans	21 ST. Louis
02 Baltimore	07 Dallas-FW	12 Los Angeles	17 New York	22 San Francisco
03 Boston	08 Denver	13 Memphis	18 Philadelphia	23 Seattle
04 Chicago	09 Detroit	14 Miami	19 Phoenix	24 Tampa
05 Cincinnati	10 Houston	15 Minneapolis	20 Pittsburg	25 Washington D.C.

Fig. 1 Location of the SMEs on the US map: direct vs. hub-and-spoke transportation between Cincinnati and Seattle

Table 2 An example of quality grading system for the T_{OPM}

Letter grade	Quality level range (%)	Description
A+	90–100	Excellent: an excellent product, only marginal errors
A	75–89	Very good: some errors, but overall still outstanding product
A−	60–74	Good: good and sound understanding but some basic errors
B	45–60	Fair: product fulfill minimum requirements
C	Below 45	Fail: product does not fulfill even minimum requirements

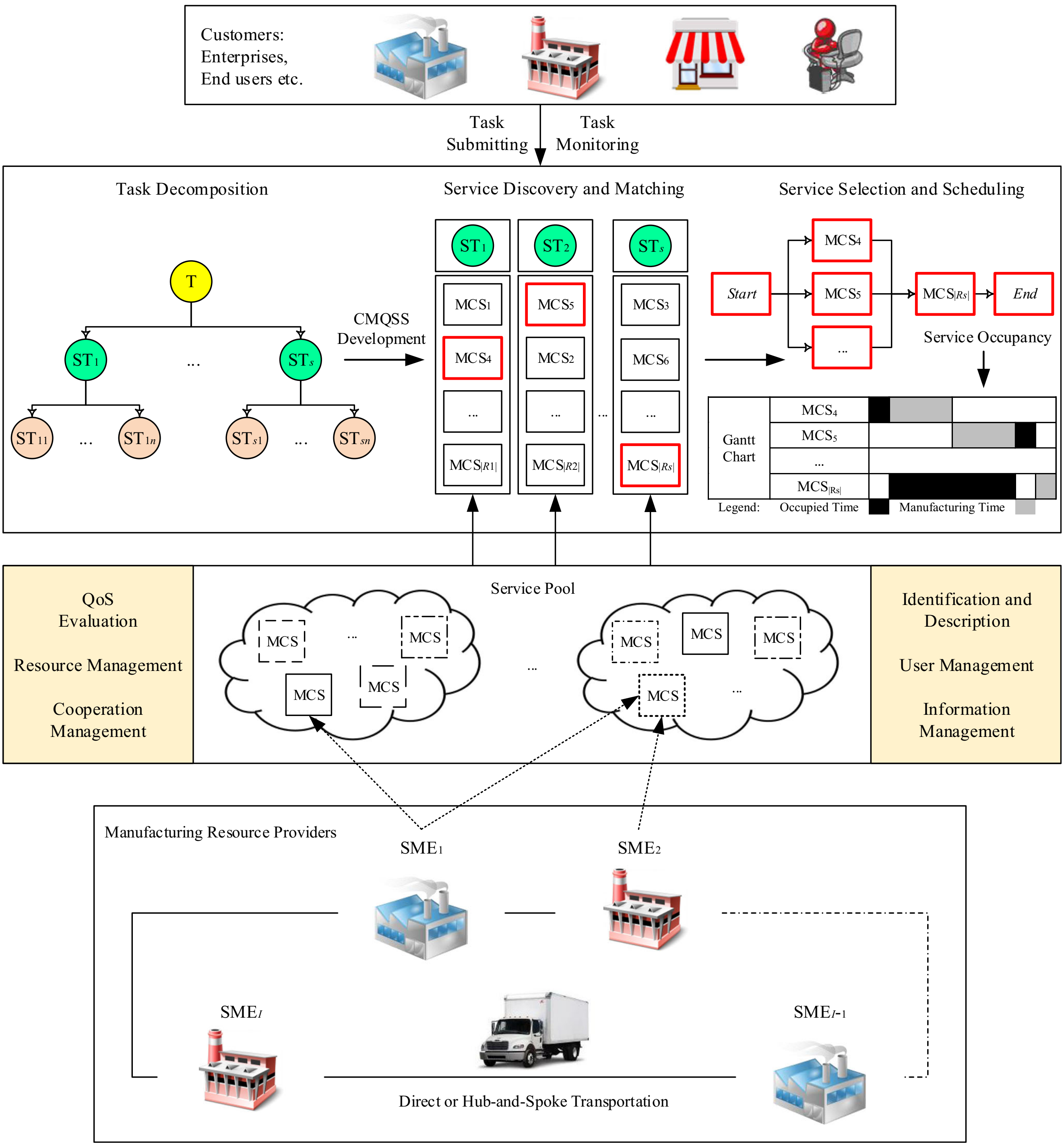


Fig. 2 The framework of the SSOS process in cloud manufacturing

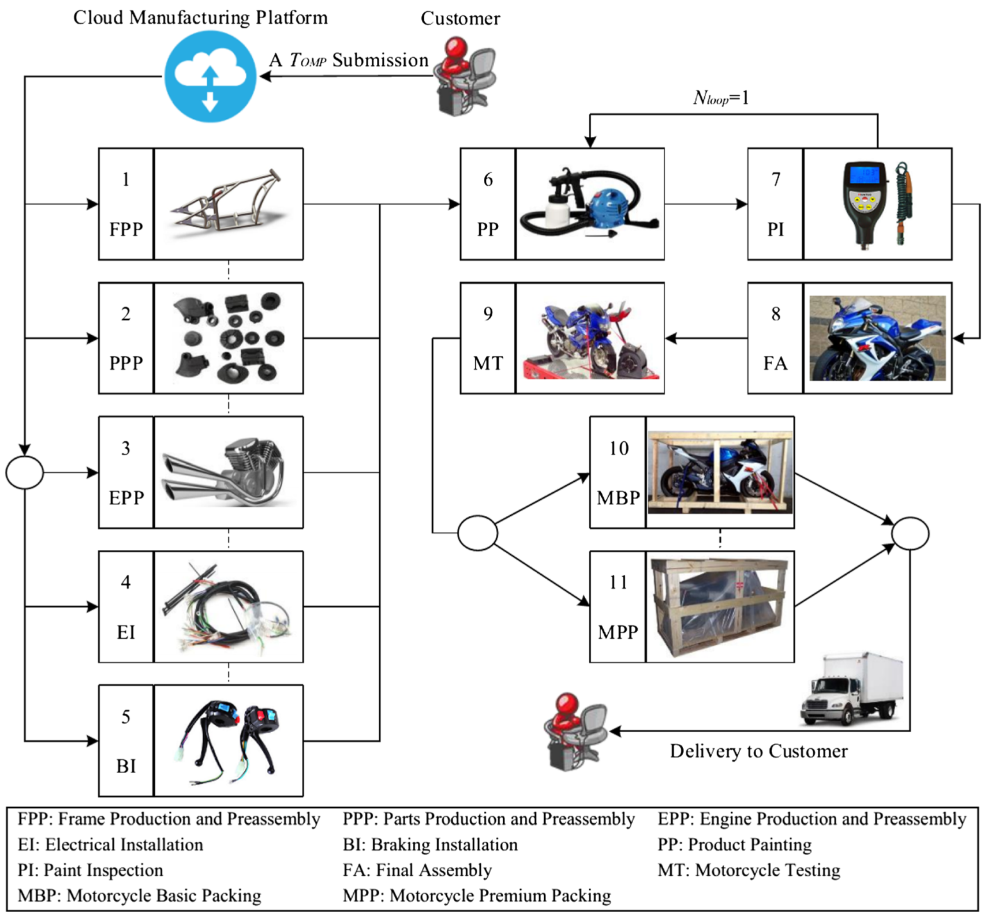


Fig. 3 Manufacturing process in the customized motorcycle production

4 The SSOS problem formulation

In formulating the SSOS, we assume that the task decomposition and service discovery processes have already been completed since they are broad topics and largely beyond the scope of the current study. The indices, parameters, and variables used to formulate the SSOS process are described in Table 4.

4.1 Manufacturing tasks, SMEs, and services

A manufacturing task is decomposed into a number of subtasks such that each one can be performed using an MCS. It has a certain subtask composition structure which is usually a combination of sequential, parallel, loop, and selective structures. In this paper, a task is denoted as *Task* (L_{Start}, L_{End}, $S_{\text{Comp\_Str}}$) in which L_{Start} and L_{End} denote the location of the workblank supplier and customer, respectively, and $S_{\text{Comp\_Str}}$ expresses subtasks' composition structure. Also, it is assumed that there are I registered SMEs, i.e., SME$_i$ ($1 \leq i \leq I$), where SME$_i$ provides NS_i different MCSs in the product lifecycle. Hence, the total number of services is equal to $TNS = \sum_{i \in I} NS_i$. MCS_r ($1 \leq r \leq TNS$) is formulated as MCS_r (L_r, PR_r, [SOT_r EOT_r]) where L_r denotes its location, PR_r is its pass rate and [SOT_r EOT_r] indicates its occupied time interval. Note that for the sake of simplicity but without loss of generality, we assume that at most there is only one interval of service occupancy. The CMQSS for sth subtask is defined as R_s (MCS_r (MC_{sr}, MT_{sr}), MCS_k (MC_{sk}, MT_{sk}), …) in which MC_{sr} and MT_{sr} are manufacturing cost and time for performing sth subtask using MCS_r, respectively.

4.2 Service selection optimization

For each subtask, several functionally equivalent MCSs but with different QoS metrics can exist. It is a realistic point of view and accords with the fact that different SMEs may have different performance in fulfilling a given subtask in terms of various QoS metrics [19, 38] even if they use the same resources. As a result, many possible solutions may be discovered for constructing an MCCS which makes service selection optimization difficult. To differentiate MCS candidates in a CMQSS, we consider cost, time, and quality metrics which are among the most important optimization criteria of the SSOS process in service-oriented manufacturing systems [13, 35]. For an MCS, cost and time metrics are defined as the cost associated with utilizing of the MCS and the time

Table 3 The generated data for the case SMEs and their MCSs in the customized motorcycle manufacturing example

SME_i	Location ($city_i$)	MCS index	Pass rate (%)	Occupied time	Manufacturing cost ($)/manufacturing time (day) for the T_{OMP}'s subtasks											NS_i
					FPP	PPP	EPP	EI	BI	PP	PI	FA	MT	MBP	MPP	
SME_{01}	Atlanta	01	0.93	[0 2]	–	–	880/0.9	–	–	–	–	–	–	–	–	1
SME_{02}	Baltimore	02	0.96	[3 5]	–	500/1.2	–	–	–	–	–	–	–	–	–	1
SME_{03}	Boston	03	0.65	[10 15]	–	612/1.3	–	–	–	–	–	–	–	–	–	1
SME_{04}	Chicago	04	0.97	[6 12]	–	–	–	–	250/0.4	–	–	–	–	–	–	1
		05	0.85	[1 4]	–	–	–	–	–	150/1.0	–	–	–	–	–	1
SME_{05}	Cincinnati	06	0.64	[10 12]	1200/1.2	–	–	–	–	–	–	1950/1.8	–	–	–	2
SME_{06}	Cleveland	07	0.71	[1 6]	–	–	–	–	–	–	–	1800/1.6	–	–	–	1
SME_{07}	Dallas FW	08	0.82	[6 8]	–	–	–	–	375/0.7	–	–	–	–	–	–	1
SME_{08}	Denver	09	0.98	[0 9]	–	–	–	–	–	–	50/0.6	–	–	–	–	1
SME_{09}	Detroit	10	0.99	[1 4]	–	535/.6	–	–	–	–	–	–	–	–	–	1
SME_{10}	Houston	11	0.66	[8 9]	–	–	–	–	–	–	–	1650/1.9	–	–	–	1
SME_{11}	Kansas City	12	0.99	[10 11]	–	–	–	–	315/0.9	–	–	–	–	–	–	1
SME_{12}	Los Angeles	13	0.98	–	980/2.5	–	–	–	–	–	–	–	–	–	–	1
		14	0.79	[0 7]	–	–	–	–	–	–	–	–	–	–	270/0.6	1
SME_{13}	Memphis	15	0.92	[1 7]	–	–	–	300/0.5	–	–	–	–	–	–	–	1
SME_{14}	Miami	16	0.66	[3 8]	–	–	–	–	–	–	–	–	–	–	330/0.5	1
SME_{15}	Minneapolis	17	0.77	[1 2]	–	455/2.1	–	–	–	–	–	–	–	–	–	1
SME_{16}	New Orleans	18	0.97	[13 15]	–	–	–	465/0.7	–	–	–	–	–	–	–	1
SME_{17}	New York	19	0.92	[11 14]	–	–	–	–	–	–	–	2000/2.0	–	–	–	1
		20	0.98	[2 5]	1005/1.8	–	–	–	–	–	–	–	–	–	–	1
SME_{18}	Philadelphia	21	0.86	[6 7]	–	–	–	–	–	–	65/0.4	–	–	–	–	1
SME_{19}	Phoenix	22	0.61	[6 10]	–	–	–	–	–	–	–	–	900/0.7	100/0.2	–	2
SME_{20}	Pittsburgh	23	0.94	[9 12]	–	–	1100/1.1	–	–	–	–	–	–	–	–	1
SME_{21}	St. Louis	24	0.97	[7 8]	–	–	–	–	–	–	–	–	–	95/0.3	–	1
SME_{22}	San Francisco	25	0.87	[9 14]	–	–	–	–	–	225/0.8	130/0.5	–	–	–	–	2
		26	0.90	[3 8]	–	680/1.7	–	–	–	–	–	–	–	–	–	1
SME_{23}	Seattle	27	0.90	[8 10]	–	–	–	–	–	–	–	–	840/0.9	–	–	1
SME_{24}	Tampa	28	0.76	[4 10]	1310/1.9	–	–	–	–	–	–	1600/2.2	–	–	–	2
SME_{25}	Washington D.C.	29	0.86	[5 11]	–	–	–	–	–	–	80/0.3	–	720/0.6	–	–	2
		30	0.67	[7 14]	–	–	–	385/0.6	–	–	–	–	–	–	–	1
$\|R_s\|$					4	5	2	3	3	2	4	5	3	2	2	TNS = 35

Table 4 The notations used in formulating the SSOS problem

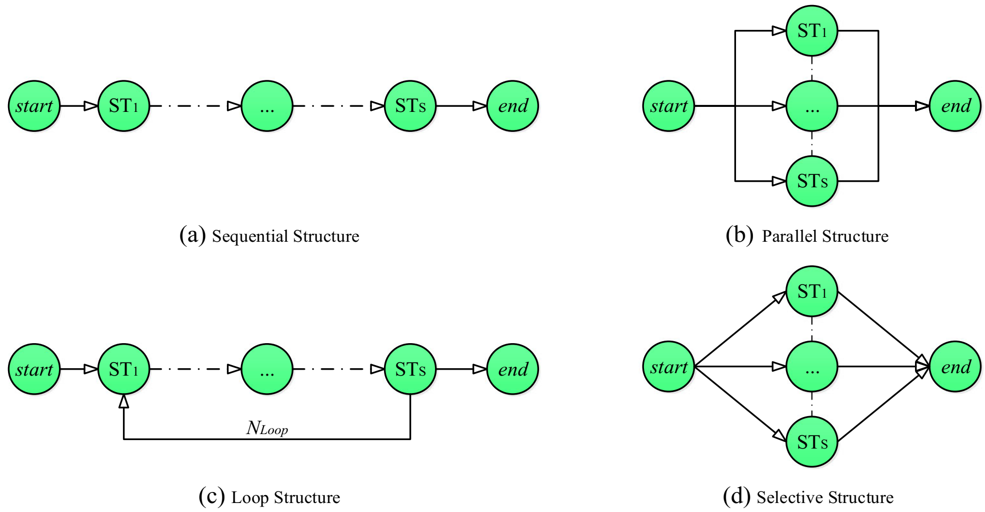
Indices	
a, b	Index of cost, time, and quality objective functions ($a, b = C, T, Q$)
i, j	Index of geographical location of manufacturing SMEs ($i, j = 1, \ldots, I$)
r, k	Index of MCSs ($r, k = 1, \ldots, K$)
s, u	Index of manufacturing subtasks ($s, u = 1, \ldots, S$)
o, p	Index of supplying and delivering subtasks in addition to manufacturing subtasks ($o, p = start, 1, \ldots, S, end$)
h, g	Index of hub facility locations in the hybrid hub-and-spoke transportation network ($h, g = 1, \ldots, H$)
n, v	Index of number of cycles in the loop composition structure ($n, v = 1, \ldots, N_{\mathrm{Loop}} + 1$)
Parameters	
MC_{sr}	Manufacturing cost for performing sth subtask using rth MCS (\$)
MT_{sr}	Manufacturing time for performing sth subtask using rth MCS (day)
Dist_{ij}	Geographical distance between city i and j (km)
TC_{ij}	Transportation cost between city i and j for unit weight (\$/kg)
TT_{ij}	Transportation time between city i and j (day)
W_{op}	Transportation weight between oth and pth subtasks (kg)
PR_r	Pass rate for rth MCS (%)
R_s	Cloud manufacturing qualified service set for sth subtask
L_r	Location of rth MCS
L_{Hub}	Location of hub facilities in the hybrid hub-and-spoke transportation network
L_{Start}	Location of the workblank supplier (where transportation starts)
L_{End}	Location of the customer (where transportation ends)
VC_h	Variable cost of using hub facility h (\$)
SOT_r	Start of the occupied time for rth MCS
EOT_r	End of the occupied time for rth MCS
NS_i	Number of provided MCSs by the SME located in city i
TNS	Total number of provided MCSs by all SMEs
N_{Loop}	Number of cycles in the loop composition structure
Pro_s	Corresponding probability for selecting sth subtask in the selective composition structure
w_C	Cost preference weight
w_T	Time preference weight
w_Q	Quality preference weight
$C_{\max}$	Maximum cost that the customer is willing to pay
$T_{\max}$	Product delivery deadline specified by the customer
$Q_{\min}$	Minimum acceptable quality level specified by the customer
α	Discount factor to decrease transportation cost between hub facilities
U_{TC}	Transportation cost for unit distance (\$/km)
U_{TT}	Transportation time for unit distance (day/km)
$S_{\mathrm{Comp\_Str}}$	Subtasks' composition structure
M	A large positive number
Binary decision variables	
x_{sr}	1 if sth subtask is performed using rth MCS; 0 otherwise
y_{ih}	1 if city is by hub h; 0 otherwise
λ_{sr}	1 if sth subtask is performed using rth MCS before starting its occupied time; 0 otherwise
λ_{srn}	1 if sth subtask is performed using rth MCS before starting its occupied time and in the nth cycle of the loop composition structure; 0 otherwise
z_{op}^{ij}	1 if oth and pth subtasks are performed in city i and j respectively and hence there is a shipment between these cities; 0 otherwise
d_{op}^{ij}	1 if oth and pth subtasks are performed in city i and j respectively and there is a direct transportation between these cities; 0 otherwise
f_{op}^{ijhg}	1 if oth and pth subtasks are performed in city i and j respectively and there is a hub-and-spoke transportation between these cities with the routing $i \rightarrow$hub $h \rightarrow$hub $g \rightarrow j$; 0 otherwise
Continuous decision variables	
TPS	Task completion time in the parallel composition structure (day)
STS_s	Start time of manufacturing process of sth subtask
ETS_s	End time of manufacturing process of sth subtask
wt_s	Waiting time before starting manufacturing process of sth subtask (day)
CSS_s	Corresponding cost in the selective composition structure if sth subtask is selected among alternative subtasks (\$)
TSS_s	Task completion time in the selective composition structure if sth subtask is selected among alternative subtasks (day)
wt_{sn}	Waiting time before starting manufacturing process of sth subtask in the nth cycle of the loop structure (day)
STS_{sn}	Start time of manufacturing process of sth subtask in the nth cycle of the loop composition structure
ETS_{sn}	End time of manufacturing process of sth subtask in the nth cycle of the loop composition structure

interval between manufacturing start time and manufacturing end time, respectively. Also, quality metric can be measured based on pass rate [13]. In this paper, we use average pass rate for evaluating the product quality level. In order to find the optimal MCSS, a QoS metric for an MCSS should be derived from aggregating the corresponding metrics of all selected MCSs. Aggregation of QoSs (without involvement of transportation and service occupancy) for basic subtask composition structures (Fig. 4) is summarized in Table 5. While the sequential structure means subtasks perform in turns, the parallel structure implies that subtasks can be performed at the same time. In the loop structure, subtasks should be performed repetitively $N_{\mathrm{Loop}} + 1$ times and the selective structure represents one of subtasks is selected according to the specific evaluation. A manufacturing task which combines more than one basic structure has a mixed composition structure.

4.3 Routing in a hybrid hub-and-spoke transportation network

The most elementary method in developing transportation networks is establishing a direct connection between each origin–destination (O–D). The dramatic growth of network development costs is a major weakness of this approach. A

Fig. 4 Four basic subtask composition structures (adapted from [55])



direct transportation should be utilized when there is a tight delivery lead-time or when the commodities need to be isolated. If these criteria are not satisfied, then transportation costs can be reduced by utilizing a hub-and-spoke network which consolidates commodities from different origin cities at hubs and transports them in bulk between hubs. This consolidation and increased volume of hub to hub commodities bring the advantage of economies of scale for transportation and discounts its cost between two hub cities by a factor α ($0 \leq \alpha \leq 1$). However, due to commodities consolidation at hub cities and routing each O–D transportation via hub(s), some commodities travel longer paths when compared to the distance between their O–D [41]. Obviously, this also results in longer transportation time.

Since a cloud manufacturing platform deals with several manufacturing tasks over time, the amount of transportation between different cities can be very significant and hard to manage. Thus, unlike the literature of the SSOS that neglects transportation or just considers a direct transportation between each O–D, we have supposed that the platform routes all required transportation in a given hybrid hub-and-spoke network. In this network, for transporting commodities between city i and j (where oth and pth subtasks of a manufacturing task are performed respectively), there are three routing alternative as follows (see Fig. 1):

- Routing directly from city i to j. In order to select this alternative, d_{op}^{ij} as a binary decision variable will be equal to 1. Note that we include subscript of subtasks to differentiate probable multiple transportations between two cities.
- Routing through one hub, which can be used when both city i and j are served by the same hub facility, e.g., hub h. This means the routing is $i \rightarrow h \rightarrow j$ and f_{op}^{ijhh} will be equal to 1.
- Routing through two hubs, which can be used when city i and j are served by different hubs, e.g., hub h and g, respectively. The routing consists of three parts: collecting commodities from city i to hub h, transporting commodities between hub h and hub g, and distributing commodities from hub g to the destination city j. This means the routing is $i \rightarrow h \rightarrow g \rightarrow j$ and f_{op}^{ijhg} will be equal to 1.

Table 5 The aggregation of QoS metrics for the basic subtask composition structures: transportation and service occupancy have not been considered (adapted from [55])

QoS metric	Sequential structure	Parallel structure	Loop structure	Selective structure
Cost	$\sum_{s=1}^{S} \sum_{r \in R_s} MC_{sr} x_{sr}$	$\sum_{s=1}^{S} \sum_{r \in R_s} MC_{sr} x_{sr}$	$N_{Loop} \sum_{s=1}^{S} \sum_{r \in R_s} MC_{sr} x_{sr}$	$\sum_{s=1}^{S} Pro_s \sum_{r \in R_s} MC_{sr} x_{sr}$
Time	$\sum_{s=1}^{S} \sum_{r \in R_s} MT_{sr} x_{sr}$	$\text{Max}\left\{ \sum_{s=1}^{S} \sum_{r \in R_s} MT_{sr} x_{sr} \right\}$	$N_{Loop} \sum_{s=1}^{S} \sum_{r \in R_s} MT_{sr} x_{sr}$	$\sum_{s=1}^{S} Pro_s \sum_{r \in R_s} MT_{sr} x_{sr}$
Quality	$\frac{1}{S} \sum_{s=1}^{S} \sum_{r \in R_s} PR_{sr} x_{sr}$	$\frac{1}{S} \sum_{s=1}^{S} \sum_{r \in R_s} PR_{sr} x_{sr}$	$\frac{1}{S} \sum_{s=1}^{S} \sum_{r \in R_s} PR_{sr} x_{sr}$	$\frac{1}{S} \sum_{s=1}^{S} Pro_s \sum_{r \in R_s} PR_{sr} x_{sr}$

The platform needs an efficient way to explore route decision-making on whether a shipment should be routed directly or through hubs facilities. A suitable approach for this decision-making can be realized based on a trade-off between cost and time. Transportation cost (for unit weight) and transportation time between city i and j are calculated through the following formulas where Dist_{ij} is geographical distance between the cities, and U_{TC} and U_{TT} are transportation cost and transportation time for unit distance, respectively.

$$TC_{ij} = Dist_{ij} \times U_{TC} \quad (1)$$

$$TT_{ij} = Dist_{ij} \times U_{TT} \quad (2)$$

Taking for example the routing through two hubs, transportation cost is $TC_{ih} + \alpha TC_{hg} + TC_{gj}$ and transportation time equals to $TT_{ih} + TT_{hg} + TT_{gj}$. It is worth to mention that transportation cost and time will be zero if two successive manufacturing subtasks are performed within the same city.

4.4 Service occupancy consideration

Most past literature on the SSOS or SCOS problems assumes that MCSs are continuously available all the time. This availability assumption cannot be justified if maintenance requirements, breakdowns, or other constraints (that cause services be partially occupied) have to be addressed. With consideration of service occupancy in the process of service selection and scheduling, not only the value of some QoS metrics can be changed but also the difficulty in finding a feasible solution may be increased. However, the obtained solution would be more realistic and practical due to taking the availability of services into account.

For the explanation of service occupancy involvement, suppose that MCS_r has been selected to perform sth subtask of a manufacturing task. As depicted in Fig. 5, six different cases can be imagined when the subtask manufacturing interval and the service occupancy interval are compared. Although there is no overlap between mentioned intervals in the first and the last cases, the other cases at least have a partial overlap area. The existence of overlap means that the current service scheduling is invalid. In the case 1, manufacturing process is finished before the service occupancy starts. Also, in the case 6, manufacturing process is started after the service occupancy ends. These two cases provide valid service scheduling and can be formulated as a simple if-then rule: IF $ETS_s \leq \mathrm{SOT}_r$ OR $STS_s \geq \mathrm{EOT}_r$ THEN service scheduling is valid.

4.5 The mathematical models for the SSOS problem

In this subsection, the developed MIP models for the SSOS problem with basic subtask composition structures (shown in Fig. 4) and consideration of transportation and service occupancy are presented. The assumptions in the models are listed below:

- For all manufacturing services, there is need for transportation.
- Transportation cost and transportation time within the same city are zero.
- Hub facilities have already been established and there is a variable cost for using a hub facility.

The main decisions which are addressed by the models include the following: selecting the optimal MCS for performing each subtask, the optimal routing alternative for transporting commodities between different cities and scheduling subtasks over available time of the selected MCSs. Also, the objective functions (corresponding to QoS metrics) for solving the SSOS problem are to fulfill a manufacturing task with the lowest possible cost, in the shortest completion time and with the highest quality product.

4.5.1 The MIP model for the sequential subtask composition structure

The objective function (3) minimizes the total cost which includes both manufacturing cost and transportation cost. While the former usually is calculated based on the quoted price by an SME to perform a given subtask, the latter depends on different factors such as transportation distance and transportation cost for unit distance (as formulated in (1)), weight of commodities needed to be transported, and more importantly, the way commodities are routed to their destination. Also, we have defined a

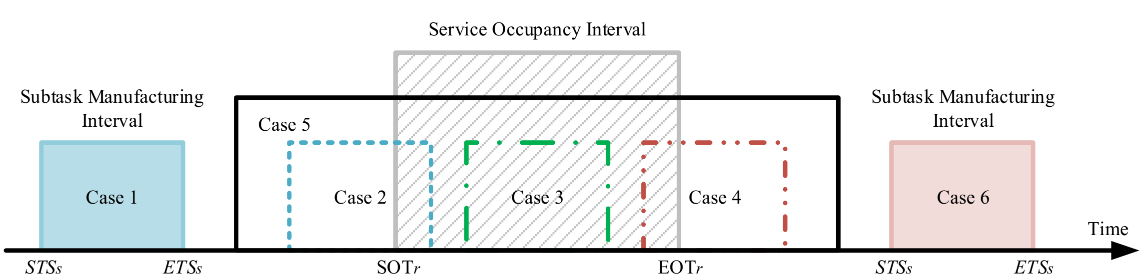


Fig. 5 Different cases in the comparison of manufacturing and service occupancy intervals: obviously, just one of these cases occurs in a time

variable cost for using a hub facility in routing a shipment through a hub-and-spoke network which is calculated through the last term in (3).

$$\begin{aligned}\text{Min } C_{\text{Seq}} = &\sum_{s=1}^{S}\sum_{r\in R_s} MC_{sr}x_{sr} + \sum_{o=start}^{S}\sum_{i=1}^{I}\sum_{j=1}^{I} TC_{ij}W_{o,o+1}d_{o,o+1}^{ij} \\ &+ \sum_{o=start}^{S}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TC_{ih}+\alpha TC_{hg}+TC_{gj}\right)W_{o,o+1}f_{o,o+1}^{ijhg} \\ &+ \sum_{h=1}^{H} VC_h y_{hh}\end{aligned} \tag{3}$$

The objective function (4) minimizes the task completion time which encompasses three parts: manufacturing time, transportation time, and waiting time. Manufacturing time, which is promised by SMEs after sending manufacturing subtask invitations, includes setup time, processing time, and maintenance time. According to formula (2), transportation time itself depends on the geographical distance that should be traveled, transportation time for unit distance and how to route a shipment in a hybrid hub-and-spoke transportation network. Also, service occupancy can result in waiting time before the start of a manufacturing process. In some cases, it would be a time-saving method to wait for utilizing a currently occupied MCS instead of having transportation to use another MCSs located in distant places.

$$\begin{aligned}\text{Min } T_{\text{Seq}} = &\sum_{s=1}^{S}\sum_{r\in R_s} MT_{sr}x_{sr} + \sum_{o=start}^{S}\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij}d_{o,o+1}^{ij} \\ &+ \sum_{o=start}^{S}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{ih}+TT_{hg}+TT_{gj}\right)f_{o,o+1}^{ijhg} + \sum_{s=1}^{S} wt_s\end{aligned} \tag{4}$$

The quality of the product is maximized through calculating average pass rate of the selected MCSs as presented in the objective function (5).

$$\text{Max } Q_{\text{Seq}} = \frac{1}{S}\sum_{s=1}^{S}\sum_{r\in R_s} PR_r x_{sr} \tag{5}$$

The aim of the model is to simultaneously optimize the mentioned objective functions subject to the following constraints. Constraint (6) imposes that there is a shipment from the location of the workblank supplier to the location where the first subtask is performed. Constraint (7) states that transportation should be considered between the locations of selected MCSs for performing two successive subtasks. The necessity of transportation from the location of the last MCS to the location of the customer is expressed using constraint (8). Constraint (9) ensures that a subtask should be performed by exactly one MCS.

$$x_{sr} \le z_{start,s}^{L_{start},L_r} \qquad s=1, \forall r\in R_s \tag{6}$$

$$x_{sr} + x_{s+1,k} \le 1 + z_{s,s+1}^{L_r,L_k} \qquad \forall s < S, r\in R_s, k\in R_{s+1} \tag{7}$$

$$x_{sr} \le z_{s,\text{end}}^{L_r,L_{\text{end}}} \qquad s=S, \forall r\in R_s \tag{8}$$

$$\sum_{r\in R_s} x_{sr} = 1 \qquad \forall s \tag{9}$$

Constraint (10) is equality for decision-making on routing a shipment directly or through hub facilities. For the routing via hubs, constraint (11) assures that the origin and destination cities are served by hub facilities. The inequality constraint (12) indicates that city i is served by hub h only if hub h selected to be used in routing shipments between different cities.

$$\sum_{h=1}^{H}\sum_{g=1}^{H} f_{o,o+1}^{ijhg} + d_{o,o+1}^{ij} = z_{o,o+1}^{ij} \qquad \forall o,i,j \tag{10}$$

$$y_{ih} + y_{jg} \ge 2f_{o,o+1}^{ijhg} \qquad \forall o,i,j,h,g \tag{11}$$

$$y_{ih} \le y_{hh} \qquad \forall i,h \tag{12}$$

In the sequential subtask composition structure (Fig. 4a), sth subtask (i.e., ST_s) cannot be performed until completion of the preceding subtask (i.e., ST_{s-1}). As Fig. 6 shows, transportation time and waiting time may be required between two consecutive subtasks. Constraints (13) and (14) calculate the starting time and ending time of subtasks, respectively. In order to determine the starting time of a given subtask, manufacturing time, transportation time, and waiting time for all previous subtasks should be summed with its waiting time. Also, ending time of a subtask equals its starting time plus the corresponding manufacturing time.

$$\begin{aligned}&\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij}d_{start,1}^{ij} + \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{ih}+TT_{hg}+TT_{gj}\right)f_{start,1}^{ijhg} + \sum_{s=1}^{u-1}\sum_{r\in R_s} MT_{sr}x_{sr} \\ &\quad + \sum_{o=1}^{u-1}\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij}d_{o,o+1}^{ij} + \sum_{o=1}^{u-1}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{ih}+TT_{hg}+TT_{gj}\right)f_{o,o+1}^{ijhg} + \sum_{s=1}^{u} wt_s = STS_u \qquad \forall u\end{aligned} \tag{13}$$

$$STS_u + \sum_{k\in R_u} MT_{uk}x_{uk} = ETS_u \qquad \forall u \tag{14}$$

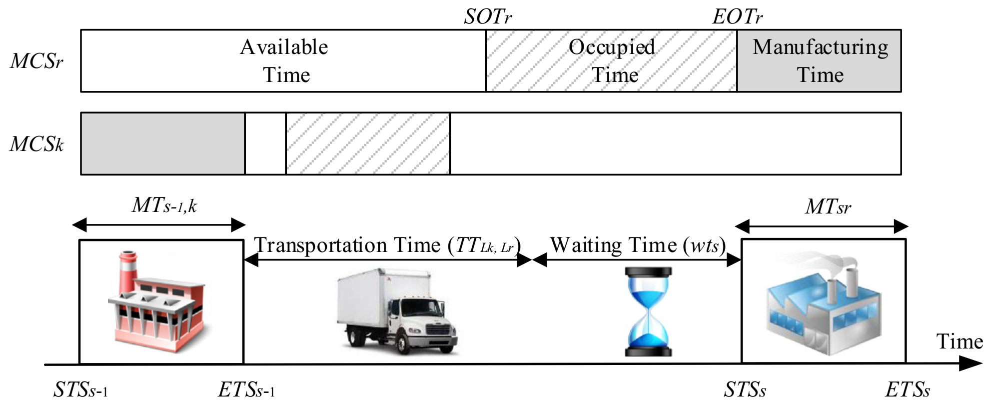


Fig. 6 Illustration of transportation and waiting time between two consecutive subtasks in the sequential structure

Constraints (15) and (16) schedule a subtask to start after or to finish before occupancy of the selected MCS, respectively. Using variable λ, the model guarantees that only one of these constraints can be active for each subtask. Constraints (17–20) express that the QoS requirements (specified by the customer) in terms of cost, time, and quality should be satisfied. Constraints (20) enforces the binary and non-negativity restrictions on the related decision variables.

$$STS_u + M\lambda_{uk} \geq EOT_k x_{uk} \qquad \forall u, k \in R_u \qquad (15)$$

$$ETS_u \leq SOT_k x_{uk} + M(1-\lambda_{uk}) \qquad \forall u, k \in R_u \qquad (16)$$

$$C_{\mathrm{Seq}} \leq C_{\max} \qquad (17)$$

$$T_{\mathrm{Seq}} \leq T_{\max} \qquad (18)$$

$$Q_{\mathrm{Seq}} \geq Q_{\min} \qquad (19)$$

$$x_{sr}, \lambda_{uk}, y_{ih}, z_{op}^{ij}, d_{op}^{ij}, f_{op}^{ijhg} \in \{0,1\} \& wt_s, STS_u, ETS_u \geq 0 \qquad \forall i,j,r,k,s,u,o,p,h,g \qquad (20)$$

4.5.2 The MIP model for the parallel subtask composition structure

In the parallel subtask composition structure (Fig. 4b), all subtasks (i.e., ST_1, …, ST_s) can be performed concurrently. The objective function (21) minimizes the maximum time required for performing parallel subtasks. The completion time of each subtask is calculated trough summation of transportation time between the supplier and the subtask, waiting time before the related manufacturing process starts, manufacturing time and the required time for transporting the subtask's output to the delivery location. Note that the objective function (21) is of the mini-max type and imposes non-linearity to the model.

Constraint (22) ensures that there are shipments from the location of the workblank supplier to the location where subtasks are performed and also from these locations to the delivery location. In the parallel structure, an MCS can perform at most one subtask which is guaranteed through constraint (23). This constraint is considered due to the possibility of concurrent subtasks and differs from the exclusivity of services in cloud computing environment which means a service is utilized at most once, regardless of subtask composition structure and number of subtasks in the SSOS process. The exclusivity of cloud computing services can be justified by the fact that the unit of time in the process is usually seconds or even milliseconds [50]. However, in a cloud manufacturing system, the unit of time is usually days and reutilizing of an MCS for performing different subtasks not only is possible when the MCS is available but also can improve some QoS metrics. For example, the completion time and cost objective functions decrease significantly if two consecutive subtasks can be performed by MCSs located within the same city. Therefore, there is no constraint related to the exclusivity of MCSs in the proposed MIP models. Constraint (24) calculates the starting time of parallel manufacturing subtasks. Moreover, constraints (9–12) as well as constraints (14–20) are used in a similar manner as the MIP model for the sequential structure.

$$\text{Min } C_{Par} = (3)$$

$$\text{Min } T_{Par} = \left(\underset{\forall s}{\text{Max}} \left(\begin{array}{l} \sum_{j=1}^{I} TT_{L_{start},j} d_{start,s}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{L_{start},h} + TT_{hg} + TT_{g,j}\right) f_{start,s}^{L_{start},jhg} \\ + \ wt_s + \sum_{r \in R_s} MT_{sr} x_{sr} \\ + \sum_{i=1}^{I} TT_{i,L_{end}} d_{s,end}^{i,L_{end}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{i,h} + TT_{hg} + TT_{g,L_{end}}\right) f_{s,end}^{i,L_{end},hg} \end{array} \right) \right) \qquad (21)$$

Max Q_{Par} = (5)

Subject to:

$$z_{start,s}^{L_{start},L_r} + z_{s,\mathrm{end}}^{L_r,L_{\mathrm{end}}} \geq 2x_{sr} \qquad \forall s, r \in R_s \tag{22}$$

$$\sum_{s=1}^{S} x_{sr} \leq 1 \qquad \forall r \tag{23}$$

$$\sum_{j=1}^{I} TT_{L_{start},j} d_{start,s}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{L_{start},h} + TT_{hg} + TT_{g,j}\right) f_{start,s}^{L_{start},jhg} + wt_s = STS_s \qquad \forall s \tag{24}$$

(9–12) and (14–20).

To linearize the objective function (21), its mini-max structure should be removed. For this purpose, we use a common trick in linear programming. First, a new continuous and non-negative variable *TPS* that denotes task completion time in the parallel structure is added to the model. Then, in addition to optimizing (3) and (5), the model minimizes *TPS* subject to the conditions that *TPS* must be at least as large as the completion time for the first subtask, *TPS* must be at least as large as the completion time for the second subtask, and so on, for each subtask. Therefore, the objective function (21) is replaced by (25) and a set of new constraints, as formulated in (26), are added to the original constraints of the model.

$$\mathrm{Min} T_{\mathrm{Par}} = TPS \tag{25}$$

$$\left(\begin{array}{l} \sum_{j=1}^{I} TT_{L_{start},j} d_{start,s}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{L_{start}},h + TT_{hg} + TT_{g,j}\right) f_{start,s}^{L_{start},jhg} \\ + \ wt_s + \sum_{r\in R_s} MT_{sr} x_{sr} \\ + \sum_{i=1}^{I} TT_{i,L_{end}} d_{s,end}^{i,L_{end}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H}\left(TT_{i,h} + TT_{hg} + TT_{g,L_{end}}\right) f_{s,end}^{i,L_{end},hg} \end{array}\right) \leq TPS \qquad \forall s \tag{26}$$

4.5.3 The MIP model for the loop subtask composition structure

Figure 4c illustrates that manufacturing subtasks in the loop subtask composition structure are performed N_{Loop} times and there are more than one entry or exit point. From transportation perspective in the loop structure, two different situations can be investigated: (1) forward transportation toward the locations where the next subtask is performed (like as the sequential structure), and (2) reverse transportation from the location where the last subtask (i.e., ST_S) is performed to the location where the first subtask (i.e., ST_1) is accomplished.

The objective function (27) minimizes the total cost. The first two terms calculate transportation cost between the supplier and the first subtask. Since all subtasks as well as the forward transportation between consecutive subtasks in the loop structure are performed N_{Loop} + 1 times, the related manufacturing and transportation cost are multiplied by N_{Loop} + 1. Also, the reverse transportation cost between the last and the first subtask is multiplied by N_{Loop}. The last two terms in (27) determine transportation cost between the last subtask and the delivery location, and variable cost of using hub facilities, respectively. The task completion time is minimized using the objective function (28) which can be described by replacing the cost in explanation of (27) by time. Note that waiting time of subtasks in all cycles are considered in calculating the completion time. The product quality in the loop structure is evaluated similar to the sequential or parallel structure.

Constraint (29) ensures that there is a shipment between the locations where the last and the first subtasks are performed. Constraint (30) determines the starting time of manufacturing subtasks in all cycles of the loop structure. To calculate the starting time for a given subtask in the *n*th cycle, following terms must be summed: transportation time between the supplier and the first subtask; manufacturing time, forward transportation time, waiting time, and reverse transportation time in the all *n*-1 previous cycles; manufacturing time, transportation time, and waiting time of the all previous subtasks in the current cycle and finally waiting time for that task in the current cycle. Constraints (31–33) are similar to constraints (14–16) and constraint 34 enforces the binary and non-negativity restrictions on the related decision variables. Constraints (6–12) remain valid in the loop structure as well.

$$
\begin{aligned}
C_{Loop} = & \left(\left(\sum_{j=1}^{I} TC_{L_{start},j} W_{start,1} d_{start,1}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{L_{start},h} + \alpha TC_{hg} + TC_{gj} \right) W_{start,1} f_{start,1}^{L_{start},jhg} \right) \right. \\
& + \left(N_{Loop}+1 \right) \left(\sum_{s=1}^{S}\sum_{r\in R_s} MC_{sr}x_{sr} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I} TC_{ij} W_{o,o+1} d_{o,o+1}^{ij} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{ih} + \alpha TC_{hg} + TC_{gj} \right) W_{o,o+1} f_{o,o+1}^{ijhg} \right) \\
& + N_{Loop} \left(\sum_{i=1}^{I}\sum_{j=1}^{I} TC_{ij} W_{S1} d_{S1}^{ij} + \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{ih} + \alpha TC_{hg} + TC_{gj} \right) W_{S1} f_{S1}^{ijhg} \right) \\
& \left. + \left(\sum_{i=1}^{I} TC_{i,L_{\text{end}}} W_{S,\text{end}} d_{S,\text{end}}^{i,L_{\text{end}}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{i,h} + \alpha TC_{hg} + TC_{g,L_{\text{end}}} \right) W_{S,\text{end}} f_{S,\text{end}}^{i,L_{\text{end}}hg} \right) + \sum_{h=1}^{H} VC_h y_{hh} \right)
\end{aligned}
\tag{27}
$$

$$
\begin{aligned}
T_{Loop} = & \left(\left(\sum_{j=1}^{I} TT_{L_{start},j} d_{start,1}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{L_{start},h} + TT_{hg} + TT_{gj} \right) f_{start,1}^{L_{start},jhg} \right) \right. \\
& + \left(N_{Loop}+1 \right) \left(\sum_{s=1}^{S}\sum_{r\in R_s} MT_{sr}x_{sr} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij} d_{o,o+1}^{ij} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{ih} + TT_{hg} + TT_{gj} \right) f_{o,o+1}^{ijhg} \right) \\
& + N_{Loop} \left(\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij} d_{S1}^{ij} + \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{ih} + TT_{hg} + TT_{gj} \right) f_{S1}^{ijhg} \right) \\
& \left. + \left(\sum_{i=1}^{I} TT_{i,L_{\text{end}}} d_{S,\text{end}}^{i,L_{\text{end}}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{i,h} + TT_{hg} + TT_{g,L_{\text{end}}} \right) f_{S,\text{end}}^{i,L_{\text{end}},hg} \right) + \sum_{s=1}^{S}\sum_{n=1}^{N_{Loop}+1} wt_{sn} \right)
\end{aligned}
\tag{28}
$$

Max Q_{Loop} = (5).
Subject to:

$$
x_{sr} + x_{uk} \le 1 + z_{su}^{L_r,L_k} \qquad s = S, u = 1, \forall r\in R_s, k\in R_u \tag{29}
$$

$$
\begin{aligned}
& \left(\left(\sum_{j=1}^{I} TT_{L_{start},j} d_{start,1}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{L_{start},h} + TT_{hg} + TT_{gj} \right) f_{start,1}^{L_{start},jhg} \right) \right. \\
& + (n-1) \left(\sum_{s=1}^{S}\sum_{r\in R_s} MT_{sr}x_{sr} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij} d_{o,o+1}^{ij} + \sum_{o=1}^{S-1}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{ih} + TT_{hg} + TT_{gj} \right) f_{o,o+1}^{ijhg} \right) + \sum_{s=1}^{S}\sum_{v=1}^{n-1} wt_{sv} \\
& + (n-1) \left(\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij} d_{S1}^{ij} + \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{ih} + TT_{hg} + TT_{gj} \right) f_{S1}^{ijhg} \right) \\
& \left. + \left(\sum_{s=1}^{u-1}\sum_{r\in R_s} MT_{sr}x_{sr} + \sum_{o=1}^{u-1}\sum_{i=1}^{I}\sum_{j=1}^{I} TT_{ij} d_{o,o+1}^{ij} + \sum_{o=1}^{u-1}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{ih} + TT_{hg} + TT_{gj} \right) f_{o,o+1}^{ijhg} + \sum_{s=1}^{u} wt_{sn} \right) \right) = STS_{u,n} \qquad \forall u, n
\end{aligned}
\tag{30}
$$

$$
STS_{un} + \sum_{k\in R_u} MT_{uk}x_{uk} = ETS_{un} \qquad \forall u, n \tag{31}
$$

$$
STS_{un} + M\lambda_{ukn} \ge EOT_k x_{uk} \qquad \forall u, n, k\in R_u \tag{32}
$$

$$
ETS_{un} \le SOT_k x_{uk} + M(1-\lambda_{ukn}) \qquad \forall u, n, k\in R_u \tag{33}
$$

$$
x_{sr}, y_{ih}, \lambda_{ukn}, z_{op}^{ij}, d_{op}^{ij}, f_{op}^{ijhg} \in \{0,1\} \,\&\, wt_{sn}, STS_{un}, ETS_{un} \ge 0 \qquad \forall i, j, r, k, s, u, o, p, n, h, g \tag{34}
$$

(6–12) and (17–19).

4.5.4 The MIP model for the selective subtask composition structure

Figure 4d depicts the selective subtask composition structure in which one subtask is selected among number of alternative subtasks. Indeed, manufacturing subtasks are performed exclusively and it is unknown which subtask will be chosen. However, we know that sth subtask is selected with the probability of Pro_s such that $\sum_{s=1}^{S} Pro_s = 1$. Therefore, the expected cost, time, and quality level can be calculated through the objective functions (35–37), respectively. Two new continuous and non-negative variables, i.e., CSS_s and TSS_s, have been defined in the selective structure which are determined based on constraints (38–39). CSS_s stands for imposed cost when sth subtask is selected (to accomplish the requested manufacturing task) and includes transportation cost between the supplier and the location where the subtask is performed, subtask manufacturing cost, and transportation cost between the subtask and the delivery location. Also, TSS_s expresses the task completion time if sth subtask is selected. It comprises waiting time and manufacturing time as well as the corresponding time for transportation between the supplier, the subtask and the delivery location. As formulated in the model, some constraints of the sequential and parallel structures are repeated in the selective structure, such as routing in the hybrid hub-and-spoke transportation network and service scheduling.

$$\text{Min } C_{\text{Sel}} = \sum_{s=1}^{S} Pro_s CSS_s + \sum_{h=1}^{H} VC_h y_{hh} \tag{35}$$

$$\text{Min } T_{\text{Sel}} = \sum_{s=1}^{S} Pro_s TSS_s \tag{36}$$

$$\text{Max } Q_{\text{Sel}} = \sum_{s=1}^{S} Pro_s \sum_{r\in R_s} PR_r x_{sr} \tag{37}$$

$$\left(\begin{array}{l} \sum_{j=1}^{I} TC_{L_{start},j} W_{start,s} d_{start,s}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{L_{start},h} + \alpha TC_{hg} + TC_{g,j}\right) W_{start,s} f_{start,s}^{L_{start},jhg} \\ + \sum_{r\in R_s} MC_{sr} x_{sr} \\ + \sum_{i=1}^{I} TC_{i,L_{\text{end}}} W_{s,\text{end}} d_{s,\text{end}}^{i,L_{\text{end}}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TC_{i,h} + \alpha TC_{hg} + TC_{g,L_{\text{end}}}\right) W_{s,\text{end}} d_{s,\text{end}}^{i,L_{\text{end}},hg} \end{array}\right) = CSS_s \quad \forall s \tag{38}$$

$$\left(\begin{array}{l} \sum_{j=1}^{I} TT_{L_{start},j} d_{start,s}^{L_{start},j} + \sum_{j=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{L_{start},h} + TT_{hg} + TT_{g,j}\right) f_{start,s}^{L_{start},jhg} \\ + wt_s + \sum_{r\in R_s} MT_{sr} x_{sr} \\ + \sum_{i=1}^{I} TT_{i,L_{\text{end}}} d_{s,\text{end}}^{i,L_{\text{end}}} + \sum_{i=1}^{I}\sum_{h=1}^{H}\sum_{g=1}^{H} \left(TT_{i,h} + TT_{hg} + TT_{g,L_{\text{end}}}\right) d_{s,\text{end}}^{i,L_{\text{end}},hg} \end{array}\right) = TSS_s \quad \forall s \tag{39}$$

(9–12), (14–20), (22), (24), $CSS_s \geq 0$, and $TSS_s \geq 0$.

4.5.5 Dealing with the mixed subtask composition structure

Usually, computing and manufacturing tasks—especially those that are more complex—have the mixed subtask composition structure which means a combination of four basic structures, namely sequential, parallel, loop, and selective. In cloud computing environment, a task with the mixed structure can be simplified or transformed to the sequential structure using the methods proposed in Ardagna and Pernici [5]. However, their proposed methods cannot be applied directly to cloud manufacturing environment due to necessity of transportation between MCSs.

In this paper, the MIP models for the basic subtask composition structures have been proposed which consider transportation and service occupancy. The vast majority of manufacturing tasks with the mixed structure can be modeled through incorporation of these models. Figure 7 illustrates how the T_{OMP} task with the mixed structure can be transformed to the sequential using the proposed MIP models. To this end, although constraints (9–12) are valid for all basic structures, a number of dedicated constraints (as discussed in the previous subsections) must be considered for the subset of subtasks with a given basic composition structure. Also, for the mixed structure, some obvious modifications have to be made in scheduling subtasks over (available) MCSs. For example, as shown in Fig. 7, the loop structure comes after the parallel structure.

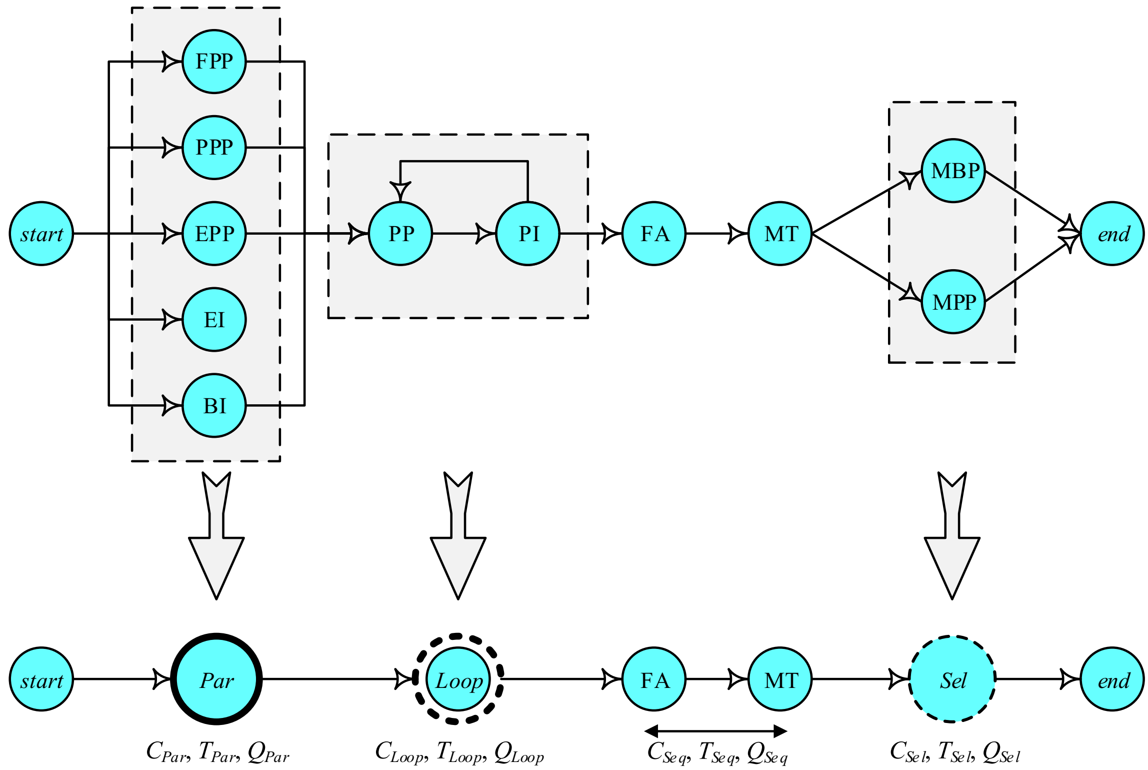


Fig. 7 Transforming the T_{OMP} with the mixed subtask composition structure to the sequential structure

Thus, T_{Par} (which denotes the completion time of subtasks in the parallel structure) is added to left-hand side of constraint (30) when this constraint is used for calculating start (and also end) time of manufacturing process for subtasks in the loop structure. Such modifications are necessary to apply for subtasks in the sequential and selective structures as well. Note that the cost objective function for the mixed structure (C_{Mix}) can easily be obtained by summation of cost objective functions of all basic structures, i.e., $C_{\mathrm{Mix}} = C_{\mathrm{Par}} + C_{\mathrm{Loop}} + C_{\mathrm{Seq}} + C_{\mathrm{Sel}}$. The same manner is applied for calculating the time and quality objective functions.

4.5.6 Formulating the overall objective function

The established MIP models for solving the SSOS are multi-objective optimization problems in which they simultaneously minimize cost and time, and maximize the product quality. Generally, in this kind of problem, it is not possible to find a solution where all the three objective functions are optimized. Instead, the overall objective function (OOF) can be optimized. For this purpose, simple additive weighting (SAW) technique as one of the simplest, natural and most widely used techniques is applied in this study. SAW includes two steps: normalization and aggregation.

Normalization: since there may be a significant difference in the objective functions' range and their unit of measurement, various objective functions (OFs) cannot be summed directly when calculating the OOF. Also, the objective functions are categorized in two, namely positive OF (e.g., quality) and negative OF (e.g., cost and time). A positive OF means that the bigger the value of the OF, the higher the interest of customers, and it is normalized using (40). A negative OF means that the lower the value of the OF, the higher the interest of customers, and it is normalized using (41). Note that through (41) and (42), the value of OFs are scaled into real value between 0 and 1 where $of_{a,\max}$ as the maximum and $of_{a,\min}$ as the minimum value of of_a in all possible MCCSs need to be determined [26, 56]. To this end, the payoff table method [10] is used in which each OF is first optimized individually. Then, the table is established such that ath row of the table represents value of all OFs calculated at the point where the of_a obtained its optimal value [18]. Therefore, as indicated in Table 6, the diagonal of the payoff table includes components of the ideal OFs. On the other hand, the worst value of the ath column can be considered as an estimate of the lower bound (resp., upper bound) of the of_a if it is a positive (resp., negative) objective function.

$$OF_a^+ = \begin{cases} \dfrac{of_a - of_{a,\min}}{of_{a,\max} - of_{a,\min}} & of_{a,\max} \neq of_{a,\min} \\ 1 & of_{a,\max} = of_{a,\min} \end{cases} \quad (40)$$

Table 6 The payoff table in the SSOS: $of_{a,b}$ is the value of of_a at the point that of_b obtained its optimal value

–	Cost ($)	Time (day)	Quality (%)
Min (cost)	$of_{C,C} = of_{C,\min}$	$of_{T,C}$	$of_{Q,C}$
Min (Time)	$of_{C,T}$	$of_{T,T} = of_{T,\min}$	$of_{Q,T}$
Max (quality)	$of_{C,Q}$	$of_{T,Q}$	$of_{Q,Q} = of_{Q,\max}$
Upper/lower bound	$of_{C,\max}$ = Max $\{of_{C,T}, of_{C,Q}\}$	$of_{T,\max}$ = Max $\{of_{T,C}, of_{T,Q}\}$	$of_{Q,\min}$ = Min $\{of_{Q,C}, of_{Q,T}\}$
Range value of the ideal solution	$[of_{C,\min}\ of_{C,\max}]$	$[of_{T,\min}\ of_{T,\max}]$	$[of_{Q,\min}\ of_{Q,\max}]$

$$OF_a^- = \begin{cases} \frac{of_{a,\max} - of_a}{of_{a,\max} - of_{a,\min}} & of_{a,\max} \neq of_{a,\min} \\ 1 & of_{a,\max} = of_{a,\min} \end{cases} \quad (41)$$

Aggregation: in this step, a weighted summation is used in order to optimize all OFs simultaneously subject to constraints of the SSOS problem. For this purpose, the scaled values of the OFs are multiplied with a weight preference and then are summed. Therefore, the OOF of an MCCS solution is calculated based on (42) with the new constraint $\sum_{a=C}^{Q} w_a = 1$, in which the weights are adopted to control the relative importance of the OFs.

$$\text{Max } OFF = w_C OF_C + w_T OF_T + w_Q OF_Q \quad (42)$$

5 Computational results

In order to evaluate the performance and the efficiency of the proposed MIP models, the T_{OMP} manufacturing task (presented in Section 3) as an example of customized motorcycle production is considered. The models for tackling the corresponding SSOS problems are coded in GAMS 23.5.2 optimization software and solved to global optimality using CPLEX solver on a 2.27 GHz CPU personal computer with 4 GB of RAM, and in Windows 7 Operating System. The aim of the study is to determine the tactical level decisions related to service selection and scheduling which include: selecting the optimal MCS for performing each subtask, routing optimization in the given hub-and-spoke transportation network, and delineating when each subtask starts and ends with the consideration of service occupancy.

Table 7 shows the default parameters in solving the SSOS for the T_{OMP}. Although the value of Wop, C_{mac}, $T_{\max}$, $Q_{\min}$, U_{TC}, and U_{TT} parameters have been selected based on the real-world setting, we have determined the value of remaining parameters randomly. According to Table 7, the task is defined as T_{OMP} (23, 10, [5 Par, 2 Loop, 2 Seq, 2 Sel]) that means transportation starts from Seattle (the supplier location) and 11 different manufacturing subtasks need to be performed before transportation ends in Houston (the customer location). In the T_{OMP}, FPP, PPP, EPP, EI, and BI subtasks have the parallel structure, PP and PI subtasks are in the loop structure, FA and MT are performed in the sequential structure, and MBP and MPP have the selective structure. In addition, data of the case SMEs, and their MCSs as well as inter-city geographical distances between the USA cities have been presented in Table 3 and Appendix, respectively. These tables include only atomic parameters which can be used to calculate other composite parameters such as $TC_{i,j}$ and $TT_{i,j}$ based on their definition. For instance, $TC_{ij} = \text{Dist}_{ij} \times U_{TC}$, then $TC_{25,\,14} = 923.2 \times 0.003 = 2.77$ which is equal to transportation cost for unit weight between Washington D.C. and Miami.

To examine the robustness of results to some key parameters or constraints in the SSOS, the following six scenarios are taken into account.

- Scenario 1 is a general case which is used as a benchmark for comparing the results of all scenarios.
- Transportation is excluded in scenario 2 to investigate whether it has any effect on the results of SSOS.
- In scenario 3, it is assumed that MCSs are available all the time and the cloud platform have access to them when required. In other words, there is no occupancy interval for MCSs.
- In scenario 4, the deadline for delivering the product is $T_{\max} = 12$ which means the optimal SSOS solution must be selected among the ones that satisfy the tight time constraint.
- Routing through the hub-and-spoke (H&S) transportation network is not allowed in scenario 5 and all shipments need to be routed directly.
- Routing through the direct transportation network is not allowed in scenario 6 and all shipments need to be routed in the H&S network.

Table 8 demonstrates range value of the ideal solution, obtained from the corresponding payoff table, as well as the value of OOF for all scenarios. As an example, in scenario 1, the cost OF lower bound is $9822.53 that can be calculated when $w_C = 1$, and $w_T = w_Q = 0$ and its upper bound equals to $C_{\max}$ = $12,000. Moreover, the time OF is between 11.66 and 14 days, and the quality OF falls in the range of 81 to 87%. Using SAW technique and according to formula (42), the value of OOF is equal to 0.73 for scenario 1. Among all

Table 7 Default atomic parameters in solving the SSOS for the T_{OMP} (except for special statements)

Parameters	Value	Unit	Type	Parameters	Value	Unit	Type
I	25		Integer	Pro_{sel}	[0.6, 0.4]		Decimal
K	30		Integer	w_C	0.4		Decimal
S	11		Integer	w_T	0.4		Decimal
H	4		Integer	w_Q	0.2		Decimal
W_{op}	40 if $o \leq 6$ and 200 if $o \geq 7$	kg	Decimal	C_{max}	12,000	$	Decimal
L_{Hub}	2, 4, 12, 20		Integer	T_{max}	14	Day	Decimal
L_{Start}	23		Integer	Q_{min}	75	%	Decimal
L_{End}	10		Integer	α	0.6		Decimal
VC_h	100 for all h	$	Decimal	U_{TC}	0.003	$/km	Decimal
N_{Loop}	1		Integer	U_{TT}	463×10^{-6}	Day/km	Decimal
TNS	35		Integer	S_{Comp_Str}	[5 Par, 2 Loop, 2 Seq, 2 Sel]		Integer

scenarios, scenario 2 leads to the higher value of OOF, which is followed by scenario 3 and scenario 1.

Table 9 provides the CPU time and details of cost, time and quality OFs in solving the SSOS problem for different scenarios. The following noteworthy points can be deduced from this table.

- The global optimal solutions are achieved under 200 s for all scenarios. This amount of time is appropriate as only tactical level of decisions are made by the proposed MIP models. The CPU time in scenarios 2 and 5 where transportation has been excluded or is possible only through direct network is significantly less than those scenarios benefiting from H&S network. Also, it is inferred that the models need the most CPU time to find the optimal solution if both direct and H&S transportation networks can be used (as in scenario 1).
- The time complexity of the proposed MIP models in the worst case is the same as that for exhaustive search [51] which means it is exponential on the number of subtasks, the size of qualified service sets, etc. However, when using CPLEX, the running time is usually better than exhaustive search as CPLEX uses cutting planes (in branch-and-cut algorithms) to tighten the linear programming relaxations. Also, it may apply certain heuristics for finding an integer-feasible solution. That is why CPLEX can solve small and medium-sized SSOS problems optimally in reasonable computational time.

Table 8 Range value of the ideal solution and the value of OOF for all scenarios

Scenario number	Range value of the ideal solution			OOF
	Cost ($)	Time (day)	Quality (%)	
Scenario 1	[9822.53 12,000]	[11.66 14]	[0.81 0.87]	0.73
Scenario 2	[5780 12,000]	[6.76 14]	[0.77 0.92]	0.83
Scenario 3	[9616.58 12,000]	[9.58 14]	[0.81 0.91]	0.81
Scenario 4	[10,968 12,000]	[11.66 12]	[0.79 83]	0.71
Scenario 5	[10,064.10 12,000]	[11.66 14]	[0.82 0.87]	0.72
Scenario 6	[10,392.80 12,000]	[11.96 14]	[0.80 0.87]	0.64

- When no transportation is involved (scenario 2), cost and time OFs have the lowest values (i.e., $6900 and 7.24 days) and the highest quality OF (i.e., 87%) can be obtained. Two main reasons can be put forward to explain this outcome. The first one is that transportation cost and transportation time are omitted that results in a considerable decrease in the value of corresponding OFs. Secondly, the model is able to select MCSs completely based on QoS of MCSs themselves without the consideration of transportation and its effect on the value of OFs. For example, MCSs with the highest pass rate can be selected for consecutive subtasks even if they are geographically very distant from each other. However, physical transportation between MCSs is an important characteristic of cloud manufacturing and the role of transportation in the SSOS process is not deniable. Therefore, the results of scenario 2 lack practicality for cloud manufacturing environment.
- A comparison between scenarios 1 and 3 can shed light on the influence of incorporating service occupancy on the outcome results of the SSOS process. As expected, the CPU time has been decreased and the value of all OFs have been improved when service occupancy is not considered due to the fact that more flexibility leads to better overall performance of cloud manufacturing system. It is also worth to notice that since all MCSs are available upon request, scenario 3 has zero waiting time which is the most important reason for the decrease in the time OF. Nonetheless, as MCSs may not be available during certain time interval(s) due to stochastic or deterministic reasons, scenario 3 dose not accord with most real industrial

Table 9 Computational results for solving the SSOS problem on all scenarios

Scenario number	CPU (s)	Cost ($)					Time (day)					Quality (%)
		Manufacturing	Transportation		Variable	Total	Manufacturing	Transportation		Waiting	Total	
			Direct	H&S				Direct	H&S			
Scenario 1	198	6369.00	2560.26	835.38	200.00	9964.64	7.88	1.68	0.98	2.17	12.71	85
Scenario 2	38	6900.00	0.00	0.00	0.00	6900.00	7.02	0.00	0.00	0.22	7.24	87
Scenario 3	137	6139.00	2685.04	793.42	200.00	9817.46	6.98	2.13	1.27	0.00	10.38	87
Scenario 4	166	6833.00	3281.48	1199.58	200.00	11,514.06	7.76	2.03	1.12	0.77	11.68	82
Scenario 5	80	6369.00	3775.88	0.00	0.00	10,144.88	7.88	2.62	0.00	2.21	12.71	85
Scenario 6	169	6345.00	0.00	3969.69	300.00	10,614.69	7.92	0.00	3.61	1.76	13.29	86

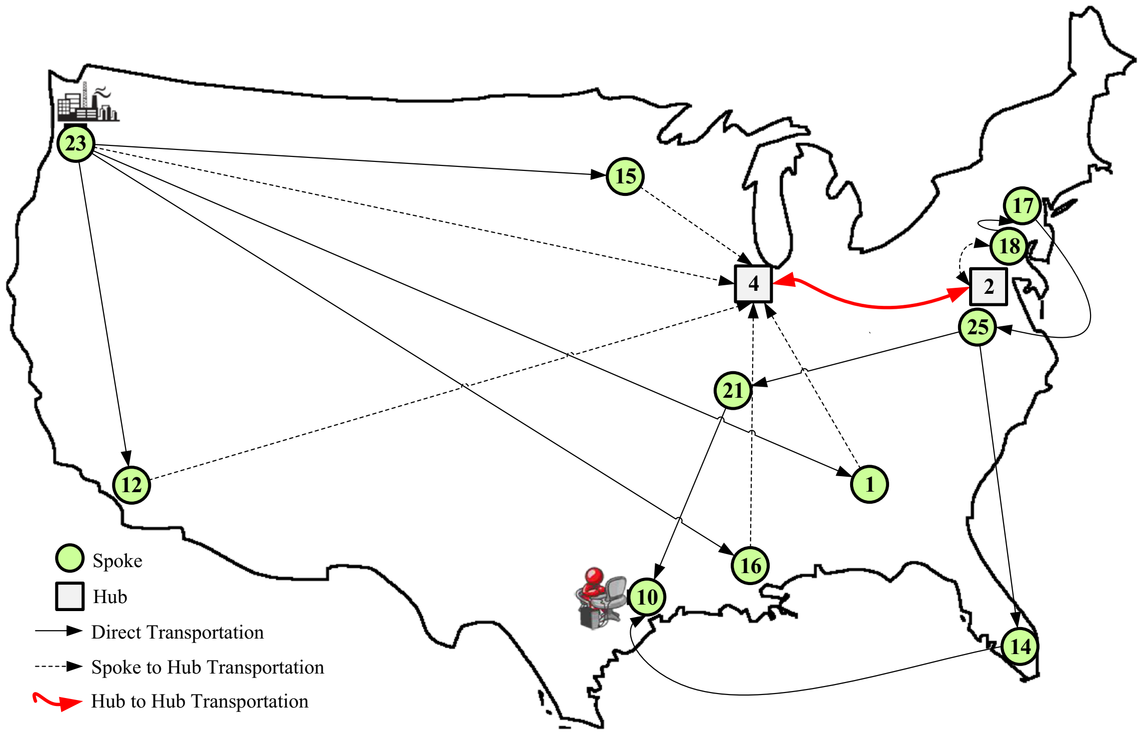
settings and its results lack practicality for cloud manufacturing environment, as like as scenario 2.

- In order to investigate the effect of the tight time constraint on the results of SSOS process, scenarios 1 and 4 can be compared. Although the time OF in scenario 4 is almost equal to its ideal value (i.e., 11.66 as shown in Table 8) and satisfies the time constraint, the cost OF has been increased by almost 16%. This means that minimization of cost OF and time OF are in conflict with each other. In addition, the quality OF has been decreased from 85% in scenario 1 to 82% in scenario 4. The results indicate that for shortening the product delivery time, the model has to worsen the value of other OFs.
- As is clear from comparing scenarios 1, 5, and 6, the hybrid H&S transportation network is more efficient than pure direct or pure H&S networks. It can be considered as an H&S network, which gains from economies of scale, allowing some transportations to be routed directly whenever beneficial, e.g., transporting commodities in the shorter time for completing a subtask before the related service occupancy gets started. Thus, the hybrid H&S potentially has advantages in terms of cost and time savings. The results of Table 9 reveal that it reduces almost 2 and 6% of the cost OF compared with pure direct and pure H&S networks, respectively. It has also more than 4% improvement in the time OF in comparison to pure H&S network.

In the rest of the paper, we only report results for scenario 1 which not only considers service occupancy and transportation, but also routes shipments in the hybrid H&S network. Figure 8 shows how the model routes shipments between different cities. In Section 4.3, we have mentioned three different alternatives for transporting commodities between a pair of cities, namely, routing directly, through one hub, and through two hubs. Since distances in the CAB dataset (which has been used in this paper for city locations and inter-city distances) satisfy triangular inequality (i.e, $\mathrm{Dist}_{ij} \le \mathrm{Dist}_{ih} + \mathrm{Dist}_{hj}$, for all i, j, h), the model will never route shipments through one hub. This is because according to (1), transportation cost is a function of distance and routing through one hub cannot decrease cost in comparison with direct routing. However, in routing through two hubs, α as a discount factor decreases transportation cost between hubs. In these cases, the model trade-off cost and time to select the best alternative. For instance, in order to transport commodities between city 4 and 18 (where PP and PI subtasks are performed, respectively), the optimal routing is $4 \rightarrow 4 \rightarrow 2 \rightarrow 18$, as indicated in Fig. 8. On the other hand, for transporting commodities between city 12 and 4 (where FPP and PP subtasks are performed, respectively), the best alternative is direct routing $12 \rightarrow 4$ regardless both of the cities can be hub facilities. In this case, $\alpha \times TC_{12,4} + VC_{12} \ge TC_{12,4}$ and that is why the model has selected direct routing alternative.

According to the results of Table 9 for scenario 1, more than 36% of the cost OF and almost 21% of the time OF are related to transportation. This reveals the great impact of transportation in the SSOS process. For the purpose of mitigating the impact of transportation, there is an overall trend that the subtasks are performed by geographically close MCSs. Figure 8 illustrates, although few subtasks are performed by MCSs located in the western part of the USA, the majority of them are fulfilled by MCSs located in the eastern half of the country. Also, it is preferred that adjacent subtasks are performed within the same city as this approach does not impose transportation cost and transportation time. Indeed, only when an SME cannot provide all required MCSs with acceptable QoS, some of subtasks will be performed by other SMEs. Accordingly, as shown in Fig. 8, both BI and PP subtasks are performed by MCSs located in city 4 (Chicago).

Figure 9 demonstrates the Gantt chart for the T_{OMP} after completion of the SSOS process. It shows what state (occupied, busy with a subtask, idle) each MCS is in at any given time, and selected MCSs for performing all the subtasks. The Gantt chart also indicates schedule of the subtasks onto MCSs,



x_{sr}= $[x_{\text{FPP},13}\ \ x_{\text{PPP},17}\ \ x_{\text{EPP},1}\ \ x_{\text{EI},18}\ \ x_{\text{BI},4}\ \ x_{\text{PP},5}\ \ x_{\text{PI},21}\ \ x_{\text{FA},19}\ \ x_{\text{MT},29}\ \ x_{\text{MBP},24}\ \ x_{\text{MPP},16}]$

L_r= $[L_{13}=12, L_{17}=15, L_1=1, L_{18}=16, L_4=4, L_5=4, L_{21}=18, L_{19}=17, L_{29}=25, L_{24}=21, L_{16}=14]$

Fig. 8 Routing shipments in the hybrid H&S transportation network for scenario 1: transportation starts from city 23 (Seattle) and ends in city 10 (Houston)

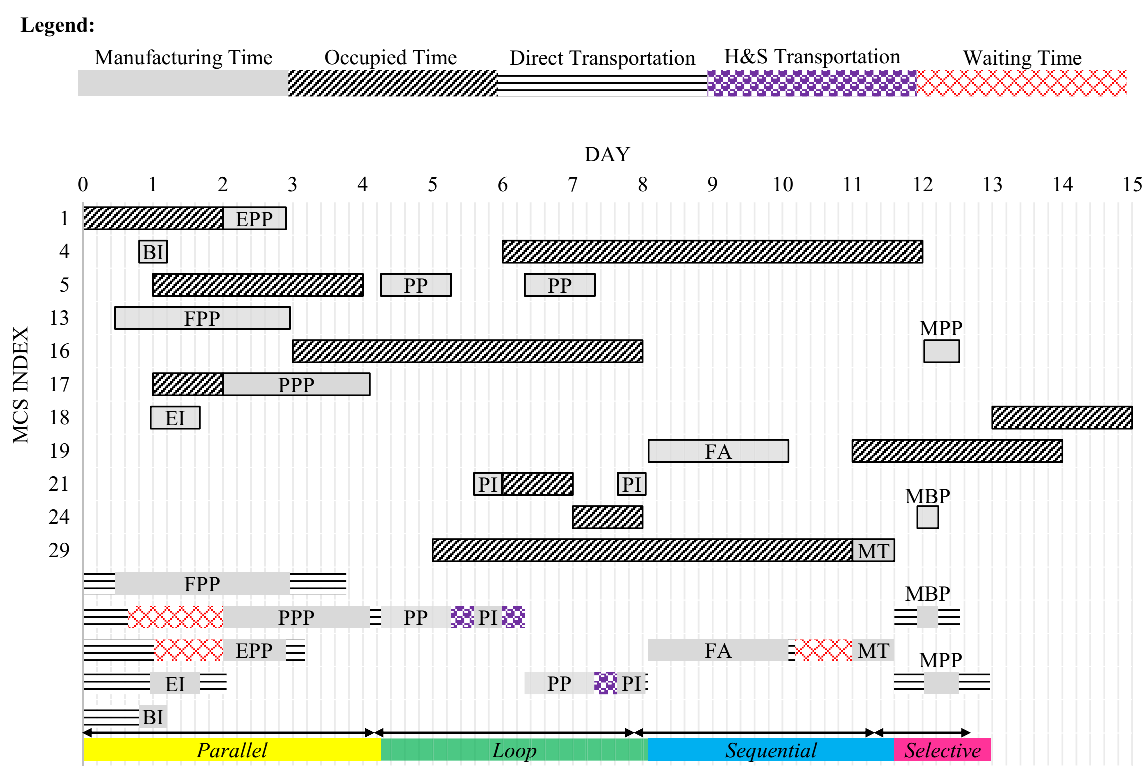


Fig. 9 The Gantt chart for the SSOS of the T_{OMP}: only involved MCSs have been shown

start time and finish time of all the subtasks, and the completion time of the T_{OMP}. It can be seen that there is no time conflict between manufacturing times and MCS's occupied time. Moreover, in the bottom of the figure, the details of transportation time and waiting time have been shown for all the subtasks with different composition structures.

5.1 Sensitivity analysis on key parameters

Input data and parameters can have a great influence on the results of heuristics and exact algorithms [3, 44]. Sensitivity analysis is conducted in this section to analyze the robustness of the proposed MIP models and gain greater insight regarding the influence of some key parameters on the results of the SSOS process. Among the numerous input parameters of the proposed models, MC_{sr}, MT_{sr}, $C_{\max}$, $T_{\max}$, $Q_{\min}$, α, L_{Hub}, and w_c are considered here for the purpose of sensitivity analysis.

In order to explore whether variations in MC_{sr} and MT_{sr} parameters for the selected MCSs (in scenario 1) can change the optimal solution, we have increased them by $\beta, \gamma \in \{5, 10, 20, 50\%\}$, respectively. When MC_{sr} parameters increase by β, the increase in the cost OF is between 3 and 14% and the quality OF decreases between 4 and 7% (see Fig. 10a). The interesting point here is that the time OF remains constant irrespective of the value of β. The phenomenon can be explained as follows. For $\beta \in \{5, 10\%\}$, the optimal MCSs are same with scenario 1 and, therefore, variation in both the time and quality OFs must be zero. Also, the increase in the cost OF is equal to β times of manufacturing cost (i.e., \$6369.00) divided by the cost OF (i.e., \$9964.64). However, for $\beta \in \{20, 50\%\}$, decrease in the quality OF means different MCSs have been selected and, hence, the increase in the cost OF would not necessarily be equal to β times of manufacturing cost divided by the cost OF. Why variation in time OF is still zero? According to Table 3, the MT subtask can be performed using three different MCSs, two of them located in the western and one of them in the eastern part of the USA. As indicated in Fig. 8, the predecessor subtask of MT are performed in the east and it would not be cost-effective (regarding transportation cost) if the MT is done in the west of the country. Therefore, the model selects MCS_{29} located in Washington D.C. (on the east) for performing the MT. Since this MCS is not available during [5 11] interval and the predecessor subtask is performed after the 5th day, the MT cannot be started sooner than the 11th day. On the other hand, with the consideration of transportation cost, the model also selects those MCSs located in the east for performing the successor subtasks of MT. Note that for these subtasks, only one MCS located in the east exists in the relevant CMQSS. Consequently, the time OF dose not undergo any change even when MC_{sr} parameters increase up to 50%. The abovementioned phenomenon reveal the importance and non-negligible influence of transportation and service occupancy considerations in the SSOS. Figure 10b demonstrates that increasing MT_{sr} parameters by γ increases the time OF by about 0.4–4% and the cost OF by about 1.3–6.8%. The effect of variation in MT_{sr} parameters on the quality OF is less than 1.2% increase or decrease which can be caused by selecting different MCSs compared to scenario 1.

In the real-world situations, customers' requirements usually include some constraints regarding cost, delivery time, and quality. There is an implicit assumption that a faster product delivery imposes more cost or decreases quality, creating a trade-off between them affecting the SSOS process. The results of sensitivity analysis on $C_{\max}$, $T_{\max}$, and $Q_{\min}$ have been tabulated in Table 10 which are consistent with the results of range value of the ideal solution shown in Table 8. Indeed, there is no feasible solution for scenario 1 if $C_{\max} < \$9822.53$ or $T_{\max} < 11.66$ days, or $Q_{\min} > 87\%$ and the T_{OMP} is regarded as being unsuccessfully performed. Table 10 also indicates less CPU time is needed to detect the infeasibility and the required CPU time will be increased when $C_{\max}$, $T_{\max}$, and $Q_{\min}$ approach the upper bound of their range value. In addition, although increase in $C_{\max}$ (resp. $T_{\max}$) causes the higher cost (resp. time) OF, it improves the value of other two OFs. When the value of $Q_{\min}$ is increased from 75 to 85%, not only

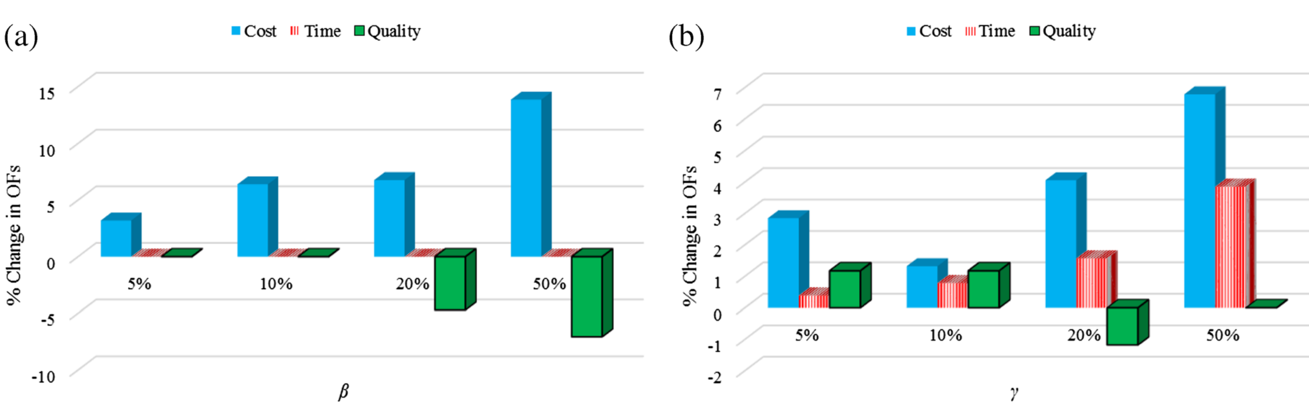


Fig. 10 The effect of increase in **a** MC_{sr} parameters by β% and **b** MT_{sr} parameters by γ% on the OFs

Table 10 The results of sensitivity analysis on $C_{\max}$, $T_{\max}$, and $Q_{\min}$

Parameter	Value	CPU (s)	Cost ($)	Time (day)	Quality (%)
$C_{\max}$	9750	108	Infeasible	Infeasible	Infeasible
	10,000	196	9822.53	12.77	83
	10,250	188	9839.20	12.71	83
	10,500	204	9964.64	12.71	85
$T_{\max}$	11.5	125	Infeasible	Infeasible	Infeasible
	12.0	187	11,514.06	11.68	82
	12.5	245	11,514.06	11.68	82
	13.0	247	9964.64	12.71	85
$Q_{\min}$	75	198	9964.64	12.71	85
	80	230	10,248.56	12.71	86
	85	245	10,248.56	12.71	86
	90	101	Infeasible	Infeasible	Infeasible

the time OF remains constant, but also only about 1% improvement in the quality OF can be obtained that costs almost 3% increase in the cost OF.

One of the parameters in the proposed MIP models that could affect the outcomes of routing optimization in the hybrid H&S transportation network is α which was set at 0.6 throughout the computational results. The impacts of changing α on the cost and time OFs as well as the percentage of using the direct or H&S transportation networks have been depicted in Fig. 11. The smaller α is the less the cost OF and the higher the time OF as there are more shipments that are routed through the H&S network. Conversely, by $\alpha \geq 0.8$, cost saving is zero and the model routes all shipments using the direct network. This results in higher cost OF and lower time OF. Therefore, the value of α has a significant impact on the OFs and routing decisions.

We have also checked the impact of changing L_{Hub} on the results of the SSOS. To this end, for different value of α, the optimal location of hub facilities (in the CAB dataset) have been extracted from the literature of uncapacitated single allocation hub location problem. In order to investigate further details of this problem, one can refer to [2, 4, 59]. Table 11 shows by increasing the number of hub facilities, the proposed model needs more CPU time. Moreover, the results demonstrate that the cost OF is more sensitive to L_{Hub} compared to the time and quality OFs. A comparison between the cost OF in Fig. 11a and Table 11 reveals that the lower cost can be obtained when city 2 (i.e., Baltimore) is a hub.

Finally, we have investigated the impact of increasing w_C on the value of all OFs for which the constraint $\sum_{a=C}^{Q} w_a = 1$ should be satisfied. According to the results of scenario 1 in Table 8, the percentage variation in the range value of the ideal solution for quality OF is significantly lower than other two OFs. We use this fact to fix w_Q at its default value, i.e., $w_Q = 0.2$, and therefore $w_T = 0.8 - w_C$. Consequently, increasing w_C means decreasing w_T. Figure 12 illustrates that increasing w_C improves that value of cost and quality OFs and worsens the value of time OF. Also, for $w_C \geq 0.2$, the OFF has an ascending trend. The figure, which actually depicts a subset of Pareto optimal solutions, can be beneficial for the customer when trading-off between different OFs. For instance, in the case of changing preference weights from $w_C = 0.2$, $w_T = 0.6$, $w_Q = 0.2$ to $w_C = w_T = 0.4$ and $w_Q = 0.2$, the product delivery time is increased by almost 1 day but the customer can save more than $1500 and will receive the product with almost 3.5% higher quality.

6 Conclusion

Cloud manufacturing paradigm has become a hot topic in the last 5 years and has attracted the attention of both academia and industry. It represents the future directions of advance manufacturing technologies in which, on the one hand, customers can request customized and complex manufacturing tasks on demand, and on the other hand, integration of geographically dispersed manufacturing resources is enabled for the purpose of collaboration of partner SMEs. In this way, they can fulfill the requested complex tasks cooperatively with a high degree of customization, a process that is impossible

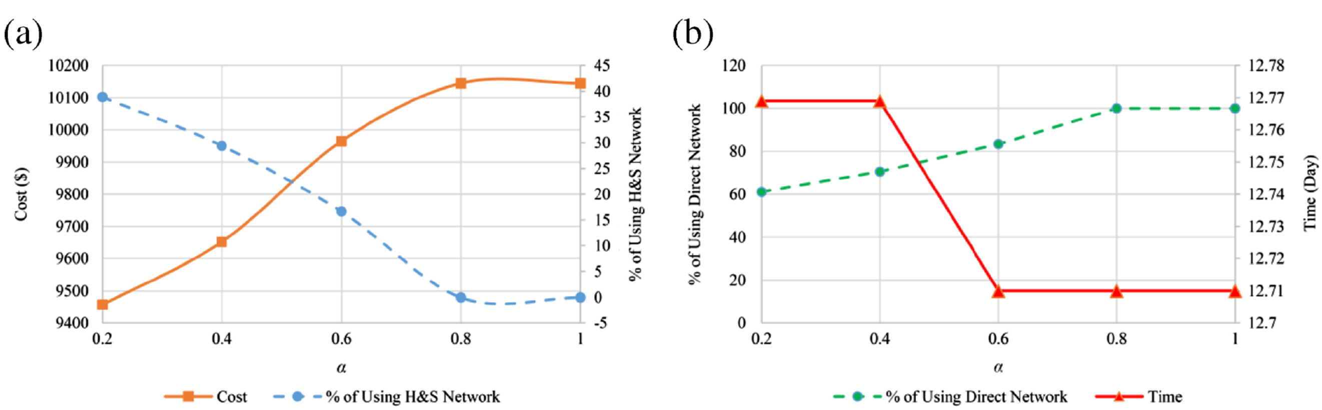


Fig. 11 The impact of changing α on **a** the cost OF and the percentage of using H&S network **b** the time OF and the percentage of using direct network

Table 11 The results of sensitivity analysis on L_{Hub}

α	L_{Hub}	CPU (s)	Cost ($)	Time (day)	Quality (%)
0.2	4, 12, 17, 20, 24	343	9579.19	12.71	86
0.4	1, 4, 12, 17, 18, 20	564	9796.27	12.71	86
0.6	2, 4, 12, 20	198	9964.64	12.71	85
0.8	2, 4, 5, 12, 20	293	10,144.88	12.71	85
1.0	4, 5, 8, 20	204	10,144.88	12.71	85

nowadays due to product-oriented business models of the standard large-scale manufacturing systems. Meanwhile, the process of SSOS is considered as an important step toward practical implementation of cloud manufacturing so that its ultimate goals are performing a task optimally through selection of MCSs, routing shipments between them and scheduling subtasks into the available time frames of the selected MCSs.

Although complex manufacturing tasks usually have a mixed structure, it is not possible to provide a general mathematical model for such tasks as changing the order of composition structures will results in significantly different outcomes. In this paper, motivated from shortcomings in the literature, we have proposed four multi-objective MIP models for solving the SSOS with a sequential, parallel, loop, and selective subtasks' composition structures, respectively. These MIP models act as elementary models such that through their incorporation, the SSOS of any task with a mixed structures can be solved. As mentioned previously, the models also consider service occupancy and transportation between MCSs. For service occupancy, we have assumed that an MCS is subject to at most one occupied interval which its start time and end time are known in advance. From transportation point of view, for the first time in the context of cloud manufacturing, the proposed models optimize routing decisions within a given hybrid hub-and-spoke network. In order to convert the multi-objective models into single objective problems, SAW technique is applied. To this end, we use the payoff table method for normalization of the OFs. Afterwards, their scaled values are multiplied with a weight preference and then are summed into the OOF.

For evaluating the performance of the models, an example of customized motorcycle production was considered. Also, six scenarios as well as a series of sensitivity analysis were designed to examine the robustness of the results. The outcomes demonstrated the usefulness and applicability of the proposed models for solving the SSOS problem. Furthermore, the results showed the great influence of transportation and service occupancy considerations in the SSOS such that without them the obtained solutions lack practicality for cloud manufacturing environment. It was also revealed that the hybrid H&S, in comparison with pure transportation networks, potentially has advantages in terms of cost and time savings. The level of savings mainly depends on the value of α.

The main limitation of this paper is that the proposed MIP models do not consider dynamic nature of the cloud-based environment. In a cloud manufacturing system, the services are dynamic which means the total number of services as well as parameters (such as cost and quality) for a given service are changed over time. On the other hand, customer requirements are subject to continuous changes. Therefore, for solving the SSOS problem, the dynamics from both parameters of services and customer requirements must be taken into account [16]. As a consequence, cloud manufacturing service selection becomes a question not only of "who" but also of "when". To answer this question, supply–demand matching simulator (SDMSim) developed by Tao et al. [53] and the proposed MIP models can be combined where, for each time period, the SDMSim reveals correlations between services and subtasks, and the MIP models find optimal MCCS, routing and subtasks scheduling. Another limitation of the current study relates to deterministic MIP models and a

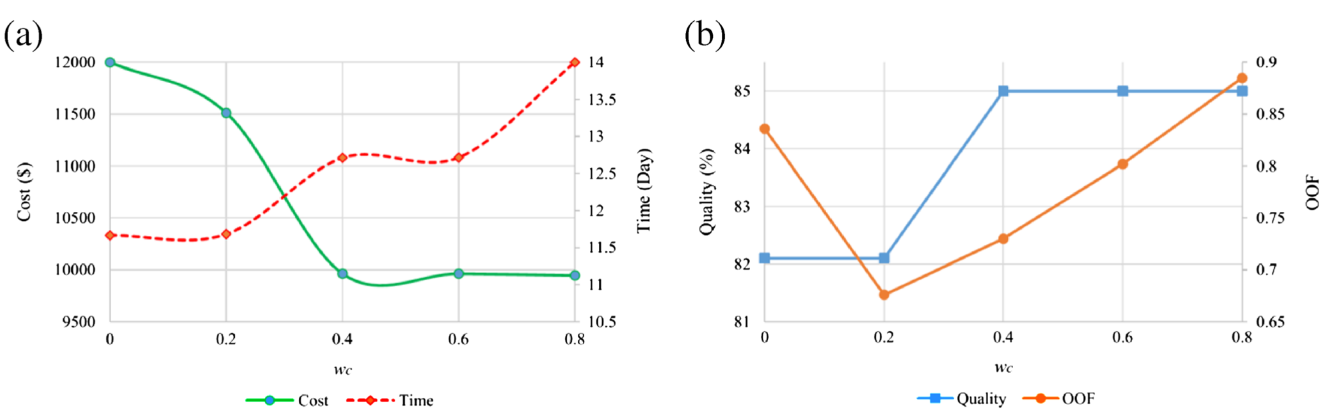


Fig. 12 The impact of increasing w_c on **a** the value of cost and time OFs **b** the value of quality OF and OOF

Appendix

Table 12 Inter-city geographical distances (km) between the US cities in the CAB dataset: the distances satisfy triangular inequality

$City_i$	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.0	577.0	946.5	597.6	373.8	559.8	709.0	1208.3	603.6	695.2	680.7	1936.6	332.5
2	577.0	0.0	369.5	613.0	429.1	312.9	1196.5	1502.1	405.9	1242.0	960.3	2318.1	786.6
3	946.5	369.5	0.0	858.3	749.6	556.1	1541.3	1764.8	621.3	1603.2	1251.0	2600.1	1137.3
4	597.6	613.0	858.3	0.0	255.0	311.3	790.1	907.4	237.1	932.2	406.3	1741.9	485.6
5	373.8	429.1	749.6	255.0	0.0	225.9	794.2	1080.4	238.9	879.6	533.2	1889.5	402.3
6	559.8	312.9	556.1	311.3	225.9	0.0	1009.7	1216.9	94.3	1104.6	694.9	2047.1	627.1
7	709.0	1196.5	1541.3	790.1	794.2	1009.7	0.0	663.9	982.7	221.4	447.8	1249.8	411.1
8	1208.3	1502.1	1764.8	907.4	1080.4	1216.9	663.9	0.0	1143.8	874.5	551.6	841.6	880.1
9	603.6	405.9	621.3	237.1	238.9	94.3	982.7	1143.8	0.0	1094.9	636.9	1978.9	620.5
10	695.2	1242.0	1603.2	932.2	879.6	1104.6	221.4	874.5	1094.9	0.0	642.2	1375.6	477.5
11	680.7	960.3	1251.0	406.3	533.2	694.9	447.8	551.6	636.9	642.2	0.0	1358.2	378.6
12	1936.6	2318.1	2600.1	1741.9	1889.5	2047.1	1249.8	841.6	1978.9	1375.6	1358.2	0.0	1608.1
13	332.5	786.6	1137.3	485.6	402.3	627.1	411.1	880.1	620.5	477.5	378.6	1608.1	0.0
14	592.6	949.6	1266.9	1186.9	947.3	1084.5	1097.6	1714.7	1151.9	963.7	1236.2	2335.8	858.3
15	908.8	938.7	1124.8	345.9	598.5	626.2	851.8	694.0	535.0	1046.1	405.1	1530.6	700.8
16	426.2	999.5	1368.3	830.4	700.4	922.3	423.7	1066.6	936.3	305.3	674.5	1661.8	348.3
17	756.2	179.2	190.3	720.5	578.3	409.4	1362.9	1625.9	489.6	1417.1	1096.7	2453.4	955.6
18	672.6	96.3	274.3	675.3	512.4	365.7	1289.0	1574.8	453.3	1337.6	1038.6	2396.8	880.0
19	1590.2	1999.6	2299.4	1447.1	1570.7	1743.4	895.1	593.4	1682.5	1017.3	1048.5	358.4	1265.6
20	527.3	210.8	494.2	403.9	255.7	104.6	1049.3	1301.5	198.9	1125.0	768.2	2125.5	651.1
21	483.5	736.4	1043.5	255.9	307.3	491.1	537.6	781.0	450.3	677.1	229.5	1582.4	255.0
22	2141.0	2456.3	2703.4	1853.6	2036.1	2164.9	1493.8	955.8	2086.8	1649.6	1506.5	361.5	1808.5
23	2184.4	2339.5	2503.8	1733.1	1967.3	2027.3	1686.7	1024.6	1936.3	1891.2	1503.8	986.8	1872.7
24	408.2	844.2	1188.5	1005.8	775.2	933.2	912.2	1519.2	992.3	795.2	1038.6	2157.5	660.5
25	540.7	36.5	405.8	592.0	399.2	298.8	1161.7	1475.5	392.9	1205.7	931.7	2288.7	751.5

$City_i$	14	15	16	17	18	19	20	21	22	23	24	25
1	592.6	908.8	426.2	756.2	672.6	1590.2	527.3	483.5	2141.0	2184.4	408.2	540.7
2	949.6	938.7	999.5	179.2	96.3	1999.6	210.8	736.4	2456.3	2339.5	844.2	36.5
3	1266.9	1124.8	1368.3	190.3	274.3	2299.4	494.2	1043.5	2703.4	2503.8	1188.5	405.8
4	1186.9	345.9	830.4	720.5	675.3	1447.1	403.9	255.9	1853.6	1733.1	1005.8	592.0
5	947.3	598.5	700.4	578.3	512.4	1570.7	255.7	307.3	2036.1	1967.3	775.2	399.2
6	1084.5	626.2	922.3	409.4	365.7	1743.4	104.6	491.1	2164.9	2027.3	933.2	298.8
7	1097.6	851.8	423.7	1362.9	1289.0	895.1	1049.3	537.6	1493.8	1686.7	912.2	1161.7
8	1714.7	694.0	1066.6	1625.9	1574.8	593.4	1301.5	781.0	955.8	1024.6	1519.2	1475.5
9	1151.9	535.0	936.3	489.6	453.3	1682.5	198.9	450.3	2086.8	1936.3	992.3	392.9
10	963.7	1046.1	305.3	1417.1	1337.6	1017.3	1125.0	677.1	1649.6	1891.2	795.2	1205.7
11	1236.2	405.1	674.5	1096.7	1038.6	1048.5	768.2	229.5	1506.5	1503.8	1038.6	931.7

Table 12 (continued)

$City_i$	14	15	16	17	18	19	20	21	22	23	24	25
12	2335.8	1530.6	1661.8	2453.4	2396.8	358.4	2125.5	1582.4	361.5	986.8	2157.5	2288.7
13	858.3	700.8	348.3	955.6	880.0	1265.6	651.1	255.0	1808.5	1872.7	660.5	751.5
14	0.0	1500.8	675.8	1098.3	1021.6	1977.6	1015.2	1065.6	2591.4	2725.8	197.8	923.2
15	1500.8	0.0	1039.8	1018.4	987.9	1280.7	728.4	450.4	1589.8	1401.3	1311.2	922.3
16	675.8	1039.8	0.0	1178.4	1095.7	1304.0	918.6	602.0	1916.6	2090.1	496.4	963.0
17	1098.3	1018.4	1178.4	0.0	84.3	2143.6	328.8	880.5	2574.1	2415.5	1008.2	215.6
18	1021.6	987.9	1095.7	84.3	0.0	2082.3	273.4	818.1	2526.6	2388.7	926.6	132.8
19	1977.6	1280.7	1304.0	2143.6	2082.3	0.0	1814.8	1264.2	661.7	1129.3	1800.1	1968.7
20	1015.2	728.4	918.6	328.8	273.4	1814.8	0.0	552.4	2253.2	2128.8	875.3	194.6
21	1065.6	450.4	602.0	880.5	818.1	1264.2	552.4	0.0	1735.9	1712.1	871.6	706.5
22	2591.4	1589.8	1916.6	2574.1	2526.6	661.7	2253.2	1735.9	0.0	694.9	2404.8	2430.3
23	2725.8	1401.3	2090.1	2415.5	2388.7	1129.3	2128.8	1712.1	694.9	0.0	2528.5	2321.9
24	197.8	1311.2	496.4	1008.2	926.6	1800.1	875.3	871.6	2404.8	2528.5	0.0	813.6
25	923.2	922.3	963.0	215.6	132.8	1968.7	194.6	706.5	2430.3	2321.9	813.6	0.0

future research direction is developing stochastic models to evaluate how sensitive the results are to variations or uncertainty in parameters. Moreover, as CPLEX solver cannot solve the large-sized SSOS problems in reasonable computational time, heuristic, and metaheuristic algorithms can be applied to search for optimal or near-optimal solution(s).

References

1. Akbaripour H, Houshmand M, Valilai OF (2015) Cloud-based global supply chain: a conceptual model and multilayer architecture. J Manuf Sci Eng 137(4):31–36
2. Akbaripour H, Masehian E (2013) Efficient and robust parameter tuning for heuristic algorithms. Int J Ind Eng Prod Res 24(2):143–150
3. Akbaripour H, Masehian E (2017) Semi-lazy probabilistic roadmap: a parameter-tuned, resilient and robust path planning method for manipulator robots. Int J Adv Manuf Technol 89(5–8):1401–1430
4. Akbaripour H, Masehian E, Roostaei A (2017) Landscape analysis and scatter search metaheuristic for solving the uncapacitated single allocation hub location problem. Int J Ind Syst Eng 26(4):425–459
5. Ardagna D, Pernici B (2007) Adaptive service composition in flexible processes. IEEE Trans Softw Eng 33(6):369–384
6. Arkat J, Ghahve H (2014) Scheduling of virtual manufacturing cells with outsourcing allowed. Int J Comput Integr Manuf 27(12):1079–1089
7. Arora N et al (2008) Putting one-to-one marketing to work: personalization, customization, and choice. Mark Lett 19(3–4):305
8. Aykin T (1995) The hub location and routing problem. Eur J Oper Res 83(1):200–219
9. Banaszak ZA, Zaremba MB (2006) Project-driven planning and scheduling support for virtual manufacturing. J Intell Manuf 17(6):641–651
10. Benayoun R et al (1971) Linear programming with multiple objective functions: Step method (STEM). Math Program 1(1):366–375
11. Bennett DP, Yano CA (2004) A decomposition approach for an equipment selection and multiple product routing problem incorporating environmental factors. Eur J Oper Res 156(3):643–664
12. Campbell JF (1994) Integer programming formulations of discrete hub location problems. Eur J Oper Res 72(2):387–405
13. Cao Y et al (2016) A TQCS-based service selection and scheduling strategy in cloud manufacturing. Int J Adv Manuf Technol 82(1–4):235–251
14. Çetiner S, Sepil C, Süral H (2010) Hubbing and routing in postal delivery systems. Ann Oper Res 181(1):109–124
15. Chen X-J et al (2012) Resource reconstruction algorithms for on-demand allocation in virtual computing resource pool. Int J Autom Comput 9(2):142–154
16. Cheng Y et al (2017) Modeling of manufacturing service supply–demand matching hypernetwork in service-oriented manufacturing systems. Robot Comput Integr Manuf 45:59–72
17. Cheng Z et al (2014) Multitask oriented virtual resource integration and optimal scheduling in cloud manufacturing. J Appl Math 2014(1):1–9
18. Deb K, Miettinen K & Sharma D (2009) A hybrid integrated multi-objective optimization procedure for estimating nadir point. In International Conference on Evolutionary Multi-Criterion Optimization. pp. 569–583

19. Duflou JR et al (2012) Towards energy and resource efficient manufacturing: a processes and systems approach. CIRP Ann-Manuf Technol 61(2):587–609
20. Ebery J et al (2000) The capacitated multiple allocation hub location problem: formulations and algorithms. Eur J Oper Res 120(3): 614–631
21. He W, Xu L (2014) A state-of-the-art survey of cloud manufacturing. Int J Comput Integr Manuf 28(3):239–250
22. Hsu C, Hsieh Y (2004) Direct versus hub-and-spoke routing on a maritime container network. J Mar Sci Technol 13(3):209–217
23. Hu SJ et al (2011) Assembly system design and operations for product variety. CIRP Ann-Manuf Technol 60(2):715–733
24. Huang B, Li C, Yin C et al (2013a) Cloud manufacturing service platform for small- and medium-sized enterprises. Int J Adv Manuf Technol 65(9–12):1261–1272
25. Huang B, Li C, Tao F (2013b) A chaos control optimal algorithm for QoS-based service composition selection in cloud manufacturing system. Enterp Inf Syst 8(4):445–463
26. Jin H, Yao X & Chen Y, (2015) Correlation-aware QoS modeling and manufacturing cloud service composition. J Intell Manuf, Available at: https://doi.org/10.1007/s10845-015-1080-2
27. Kesen SE, Das SK, Güngör Z (2010) A genetic algorithm based heuristic for scheduling of virtual manufacturing cells (VMCs). Comput Oper Res 37(6):1148–1156
28. Kurdi H et al (2015) A combinatorial optimization algorithm for multiple cloud service composition. Comput Electr Eng 42(0): 107–113
29. Laporte G, Salazar-Gonzalez JJ, Semet F (2004) Exact algorithms for the job sequencing and tool switching problem. IIE Trans 36(1): 37–45
30. Lartigau J et al (2015) Cloud manufacturing service composition based on QoS with geo-perspective transportation using an improved Artificial Bee Colony optimisation algorithm. Int J Prod Res 53(14):4380–4404
31. Li W et al (2015) Subtask scheduling for distributed robots in cloud manufacturing. IEEE Syst J 11(2):1–10
32. Lin C-C, Lin Y-J, Lin D-Y (2003) The economic effects of center-to-center directs on hub-and-spoke networks for air express common carriers. J Air Transp Manag 9(4):255–265
33. Liu B & Zhang Z (2016) QoS-aware service composition for cloud manufacturing based on the optimal construction of synergistic elementary service groups. Int J Adv Manuf Technol, Available at: https://doi.org/10.1007/s00170-016-8992-7
34. Liu J, Li C-L, Chan C-Y (2003a) Mixed truck delivery systems with both hub-and-spoke and direct shipment. Transp Res Part E: Logist Transp Rev 39(4):325–339
35. Liu L-L et al (2003b) Self-Organization Manufacturing Grid and its task scheduling algorithm. Comput Integr Manuf Syst 9(6):449–455
36. Liu N, Li X, Shen W (2014) Multi-granularity resource virtualization and sharing strategies in cloud manufacturing. J Netw Comput Appl 46:72–82
37. Liu W et al (2013) Study on multi-task oriented services composition and optimisation with the "Multi-Composition for Each Task" pattern in cloud manufacturing systems. Int J Comput Integr Manuf 26(8):786–805
38. Liu, Y. et al., 2016. Workload-based multi-task scheduling in cloud manufacturing. Robotics and Computer-Integrated Manufacturing, Available at: https://linkinghub.elsevier.com/retrieve/pii/S0736584516303210
39. Lu Y, Xu X, Xu J (2014) Development of a hybrid manufacturing cloud. J Manuf Syst 33(4):551–566
40. Magnusson M, Pasche M (2014) A contingency-based approach to the use of product platforms and modules in new product development. J Prod Innov Manag 31(3):434–450
41. Mahmutoğulları Aİ, Kara BY (2015) Hub location problem with allowed routing between nonhub nodes. Geogr Anal 47(4):410–430
42. Mai J et al (2016) Customized production based on distributed 3D printing services in cloud manufacturing. Int J Adv Manuf Technol 84(1–4):71–83
43. Manne AS (1960) On the job-shop scheduling problem. Oper Res 8(2):219–223
44. Masehian E, Akbaripour H, Mohabbati-Kalejahi N (2013) Landscape analysis and efficient metaheuristics for solving the n-queens problem. Comput Optim Appl 56(3):735–764
45. Nikolopoulou AI et al (2017) Moving products between location pairs: cross-docking versus direct-shipping. Eur J Oper Res 256(3): 803–819
46. O'kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. Eur J Oper Res 32(3):393–404
47. Ren L et al (2015) Cloud manufacturing: from concept to practice. Enterp Inf Syst 9:37–41
48. Ren L et al. (2014) Cloud manufacturing: key characteristics and applications. Int J Comput Integr Manuf, Available at: https://www.scopus.com/inward/record.url?eid=2-s2.0-84897355143&partnerID=tZOtx3y1
49. Seghir, F. & Khababa, A., 2016. A hybrid approach using genetic and fruit fly optimization algorithms for QoS-aware cloud service composition. J Intell Manuf, pp1–20
50. Shen Y, Yang X (2011) A self-optimizing QoS-aware service composition approach in a context sensitive environment. J Zhejiang Univ Sci C 12(3):221–238
51. Silva APD, Stam A (1997) A mixed integer programming algorithm for minimizing the training sample misclassification cost in two-group classification. Ann Oper Res 74:129–157
52. Song T et al (2014) Common engines of cloud manufacturing service platform for SMES. Int J Adv Manuf Technol 73(1–4):557–569
53. Tao F et al (2017) SDMSim: a manufacturing service supply–demand matching simulator under cloud environment. Robot Comput Integr Manuf 45:34–46
54. Tao F et al. (2011) Cloud manufacturing: a computing and service-oriented manufacturing model. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 225, pp.1969–1976
55. Tao F et al (2010) Correlation-aware resource service composition and optimal-selection in manufacturing grid. Eur J Oper Res 201(1):129–143
56. Tao F et al (2013) FC-PACO-RM: a parallel method for service composition optimal-selection in cloud manufacturing system. IEEE Trans Ind Inform 9(4):2023–2033
57. Tao F et al (2008) Resource service composition and its optimal-selection based on particle swarm optimization in manufacturing grid system. IEEE Trans ind Inform 4(4):315–327
58. Tian S et al. (2013) A discrete hybrid bees algorithm for service aggregation optimal selection in cloud manufacturing. *Lecture Notes in Computer Science*, 8206 LNCS, pp.110–117
59. Topcuoglu H et al (2005) Solving the uncapacitated hub location problem using genetic algorithms. Comput Oper Res 32(4):967–984
60. Valilai OF, Houshmand M (2013) A collaborative and integrated platform to support distributed manufacturing system using a service-oriented approach based on cloud computing paradigm. Robot Comput Integr Manuf 29(1):110–127
61. Wang SQ et al (2014) Research on the selection problem of machine tools in cloud manufacturing environment. Mater Sci Forum 800–801:649–653
62. Wei X, Liu H (2015) A cloud manufacturing resource allocation model based on ant colony optimization algorithm. Int J Grid Distrib Comput 8(1):55–66
63. Wu D et al (2015) Cloud-based design and manufacturing: a new paradigm in digital manufacturing and design innovation. Comput Aided Des 59:1–14

64. Wu D et al (2013) Cloud manufacturing: strategic vision and state-of-the-art. J Manuf Syst 32(4):564–579
65. Xiang F et al (2014) QoS and energy consumption aware service composition and optimal-selection based on Pareto group leader algorithm in cloud manufacturing system. CEJOR 22(4):663–685
66. Xiang F et al. (2016) The case-library method for service composition and optimal selection of big manufacturing data in cloud manufacturing system. Int J Adv Manuf Technol, pp.59–70
67. Xu X (2012) From cloud computing to cloud manufacturing. Robot Comput Integr Manuf 28(1):75–86
68. Xue X, Wang S, Lu B (2016) Manufacturing service composition method based on networked collaboration mode. J Netw Comput Appl 59(1):28–38
69. Yan-hua Z, Lei F, Zhi Y (2011) Optimization of cloud database route scheduling based on combination of genetic algorithm and ant colony algorithm. Procedia Eng 15:3341–3345
70. Yuan M et al. (2016) Multi-objective optimal scheduling of reconfigurable assembly line for cloud manufacturing. Optimization Methods & Software, Available at: https://doi.org/10556788.2016.1230210
71. Zhang L et al (2014) Cloud manufacturing: a new manufacturing paradigm. Enterp Inf Syst 8(2):167–187
72. Zheng H, Feng Y, Tan J (2016) A fuzzy QoS-aware resource service selection considering design preference in cloud manufacturing system. Int J Adv Manuf Technol 84(1–4):371–379
73. Zhou J & Yao X (2016a) A hybrid artificial bee colony algorithm for optimal selection of QoS-based cloud manufacturing service composition. Int J Adv Manuf Technol. Available at: https://doi.org/10.1007/s00170-016-9034-1
74. Zhou J & Yao X (2016b) DE-caABC: differential evolution enhanced context-aware artificial bee colony algorithm for service composition and optimal selection in cloud manufacturing. Int J Adv Manuf Technol, Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84975226461&partnerID=40&md5=5c0b4595703304b2dedcff7878ef034f
75. Zhu L, Li Q, He L (2012) Study on cloud computing resource scheduling strategy based on the ant colony optimization algorithm. IJCSI Int J Comput Sci Issues 9(5):54–58