Mutual manufacturing service selection and routing problem considering customer clustering in Cloud manufacturing

Mozhgan Assari, Jalal Delaram and Omid Fatahi Valilai

Industrial engineering department, Sharif University of Technology, Tehran, Iran (the Islamic Republic of)

ABSTRACT
Considering the globalization concept, all companies around the world are able to communicate with each other, share their technologies, and eliminate their production incapability issues. To achieve this goal, Cloud manufacturing system has been proposed. One of the most important issues in Cloud manufacturing is service composition, in which composition of cloud services is determined to meet customer orders. This paper considers a single-product supply network over one period in Cloud manufacturing system with the goal of service composition optimization. The network is designed in a layered structure considering transportation and production services and customers can be classified in specific clusters. A new algorithm is proposed to assign the services to the network layers, and the problem is formulated as a mixed integer model with the objective of maximizing the profit of the network. Also as the problem has an NP-hard origin a meta-heuristic algorithm is exploited to solve it.

ARTICLE HISTORY
Received 16 August 2018
Accepted 24 August 2018

KEYWORDS
Cloud manufacturing; services composition; routing optimization; customer clustering

1. Introduction

Very recently, the globalization paradigm has intensified the needs for interoperability and integration among processes, operations, companies, and factories. The globalization paradigm necessitates lower production cost and time (Johnson, 2006; Valilai & Houshmand, 2013, 2014b, Delaram & Valilai, 2018a). Product development and implementation of operations with lower cost help manufacturers decrease costs without any compromise in quality of their products (Delaram & Valilai, 2018b). In addition, manufacturers can adjust their operations with appropriate scale as soon as fluctuations in customer demand occur. Globalization has changed the structure of supply chains and the way they work (Wu et al. 2013a, Delaram & Valilai, 2016, 2017). This change is mainly empowered by new manufacturing models especially Cloud manufacturing which decrease the cost of initial investment and increase the production speed by creating virtual production lines. Figure 1 shows how Cloud manufacturing has affected the supply chains. In current manufacturing supply chain requirements flow down from product level to suppliers, who can then engage sub-tier suppliers to assist with the product development process. Although these relationships can be fruitful for all parties involved, the opportunity to enhance the consumer needs (reduce costs, improve...
quality, etc.) are severely limited because of fixed suppliers. Furthermore, when traditional supplier relationships is undesirable, it is costly to dissolve and change them. In contrast, in the Cloud manufacturing environment, manufacturing supply chain relationships are customer-based, and the cloud-based application layer can be used to generate numerous options for the users based upon their specifications, so users are linked to resource providers who can fulfill their needs (Tao, Hu, & Zhang, 2010; Houshmand & Valilai, 2012, 2013).

Cloud manufacturing is customer-oriented manufacturing model which allows access to a shared collection of distributed and diverse production resources (Valilai & Houshmand, 2010 and 2014a). This model creates temporary production lines and leads to more productivity, lower cost of product lifecycle and optimal use of resources in respond to customer demand (Wu et al. 2013b, Valilai and Houshmand 2014c). Cloud manufacturing is an adaptation of cloud computing which uses physical resources instead of computational resources (Tao et al., 2010). Services are classified based on regulations and special algorithms, and the customers use services in different product lifecycle disciplines such as product design, product manufacturing and other areas of product life cycle according to their demands (Xu, 2012). One of the goals of Cloud manufacturing is the access to value added services so that available services are manipulated to implement complicated manufacturing tasks which is called the service composition problem.

Service composition is used when it is impossible to implement a complicated task service by only one service (Hashemian & Mavaddat, 2006). This technique was used and explained in cloud computing by Kofler et al. in 2009 and Zeng et al. in 2009. The purpose of this paper is to investigate the service composition in Cloud manufacturing. One of the

Figure 1. Comparison of cloud manufacturing and traditional manufacturing.
contributions of this paper is considering the routing problem between among service providers that have not been investigated up to now in the literature. Another innovation is the fulfillment of different production operations in different places while they were accomplished in centralized mode in related studies. Since the paper aims to have the production routing problem (PRP) in Cloud manufacturing, the researches related to PRP and VRP in traditional supply chain are used as a base for the new problem.

This paper is organized as follows. In Section 2, the related literatures are reviewed to provide a general framework about the service composition problem. In Section 3, a new algorithm is presented to allocate the services to supply network agents, and a mathematical formulation of the problem is introduced. In Section 4, a genetic algorithm is presented to solve the problem considering the nature of the proposed mathematical model. In Section 5, computational experiments are presented. Sensitivity analysis is accomplished in Section 6. Finally, conclusions and future works are provided in Section 7.

2. Background

2.1. Vehicle routing problem

Vehicle routing problem (VRP) is one of the most challenging problems in the field of transportation and supply chain. It is a generalization of the travelling salesman problem (TSP) and addresses the delivering customers’ goods using a fleet of vehicles. A central station is available, the goal is to find a set of routes with the lowest cost in a way each customer will meet exactly once, and each vehicle will start its route from the initial station and end to it.

Several variations and specializations of the VRP exist (Braekers, Ramaekers, & Van Nieuwenhuyse, 2016):

1. Vehicle Routing Problem with Time Windows (VRPTW): The delivery locations have time windows within which the deliveries (or visits) should be made.
2. Capacitated Vehicle Routing Problem (CVRP or CVRPTW): The vehicles have limited capacity for carrying the goods should be delivered.
3. Vehicle Routing Problem with Multiple Trips (VRPMT): The vehicles can do more than one route.
4. Open Vehicle Routing Problem (OVRP): Vehicles are not required to return to the depot.

Among the variants of VRP, the OVRP model can help this paper to present a new model as in this paper problem routing is started from one point and finished in another point, and there is not any close rout constraint. Therefore, review of the researches related to OVRP is presented. For the first time in 2000, a heuristic solution based on a minimum spanning tree with penalties procedure for OVRP was proposed (Sariklis & Powell, 2000). Since 2000, researchers published a large number of heuristic and meta-heuristic methods such as tabu search (TS) and branch and cut algorithms for this problem (Repoussis, Tarantilis, Bräysy, & Ioannou, 2010).
In overall, all articles include capacity constraints for the vehicles; additionally, more than half of the articles include distance (or time) constraints. All articles, except for one (Li, Leung, & Tian, 2012), assume a homogeneous fleet of vehicles and all but one (Cao & Lai, 2010), assume deterministic demands. All proposed solution methods are metaheuristics, except for one (Salari, Toth, & Tramontani, 2010), and all but one (Yu, Ding, & Zhu, 2011) are tested on benchmark instances, either gathered from literature or generated by the authors. Table 1 shows the articles which (VRP) is as their main research focus.

The results of the review of VRP researches show that there are a number of differences among VRP and the considered problem in this paper. First, Routing in VRPs is among customers of a plant or distribution center, while in this paper problem routing is among different manufacturers. The second difference is that transportation among customers is accomplished by the fleet of factory in VRPs, while in this paper, transportation is outsourced and provided by the service provider. The last difference is that the production problem as an important part of this paper problem is not considered in these articles, so due to these differences, articles related to PRP are also investigated as follows.

### 2.2. Production routing problem

PRP is one of the inventory management approaches that supplier controls the inventory of retailers and specify their policy of inventories. The goal of PRP is optimization of production, inventory, distribution and routing decisions.

Lie et al. in 2006 investigated the production-routing-inventory problem in order to meet the requirements of customers with the lowest cost. In this case, vehicles are considered homogenous and each distribution center has a specific capacity. Since solving such a problem is not easy due to its combined nature and the routing problem, a two-step method is proposed for solving this model. Boudia et al. in 2007 considered the production distribution problem for one product. The purpose of this problem was the minimization of distribution, inventory and production costs. In this paper, a linear integer model is proposed which is not appropriate for big scale problems, so a metaheuristic method is proposed instead of classic two-step approach. Boudia et al. in 2008 studied the problem of production and distribution of a single product in a time

### Table 1. Summary of OVRP articles.

| Author(s) | Variant of OVRP | Vehicle | Demand | Distance constraint |
|-----------|-----------------|---------|--------|---------------------|
| MirHassani and Abolghasemi (2011) | OVRP | Hom. | Deterministic | No |
| Yu et al. (2011) | OVRP | Hom. | Deterministic | No |
| Zachariadis and Kiranoudis (2010) | OVRP | Hom. | Deterministic | No |
| Fleszar, Osman, and Hindi (2009) | OVRP | Hom. | Deterministic | Yes |
| Repoussis et al. (2010) | OVRP | Hom. | Deterministic | Yes |
| Salari et al. (2010) | OVRP | Hom. | Deterministic | Yes |
| Liu and Jiang (2012) | Close–open mixed VRP | Hom. | Deterministic | Yes |
| Li et al. (2012) | Heterogeneous Fixed fleet open VRP | Het. | Deterministic | Yes |
| Cao and Lai (2010) | Open VRP with fuzzy demand | Hom. | Fuzzy | Yes |

Hom., homogeneous; Het., heterogeneous.
horizon. The scenario presented in their article is a plant that produces a product capable of storing at the place of production or delivery location. This product can be distributed among customers with specific demands by limited capacity vehicles. The purpose is to determine the amount of production and delivery paths in each period, in such a way that the cost of production and distribution is minimized throughout the planning horizon. An integer-programming model is presented and two innovative methods and local search are considered for solving this model. Table 2 gives an overview of the PRP articles.

According to results of the reviewed articles, classical PRP and this paper problem are different because in classical PRP, production operations are accomplished in one place while in this paper problem operations are accomplished in different places. Since this paper problem is different from VRP and more similar to PRP, a new model based on PRP models is proposed in next session. Table 3 represents the similarity and differences of this paper problem with VRPs and PRPs.

2.3. Problem description

In Cloud manufacturing, a number of manufacturing resources are presented by the companies and plants who are registered in the cloud platform. Then, the resources are virtualized by cloud operator, and the fundamental for interoperability is created, so customers can easily submit their orders. Customers’ submitted orders and information then are divided into the subtasks for further manipulation and production. In fact, cloud platform is capable to produce complicated product parts by virtual production lines that none of the plants can make individually. The architecture of the Cloud manufacturing system is a hierarchical structure as illustrated in Figure 2. It consists of the following 10 layers, in which seven functional layers of the architecture are facilitated by the three layers of knowledge, cloud security, and a network such as the internet (Wu et al. 2013b).

Table 2. Summary of PRP articles.

| Author(s) | Production | Distribution |
|-----------|------------|--------------|
|           | N. plant   | Product      | Capacity | Fleet | N. vehicles | Capacity |
| Lie et al. (2006) | Multiple | Single | ✓ | Het. | Limited | ✓ |
| Boudia et al. (2007) | Single | Single | ✓ | Hom. | Limited | ✓ |
| Boudia et al. (2008) | Single | Single | ✓ | Hom. | Limited | ✓ |
| Bard and Nananukul (2009) | Single | Single | ✓ | Hom. | Limited | ✓ |
| Archti et al. (2011) | Single | Single | ✓ | Hom. | Single | ✓ |
| Adulyasak, Cordeau, and Jans (2013) | Single | Single | ✓ | Hom. | Multiple | ✓ |
| Adulyasak, Cordeau, and Jans (2015) | Single | Single | ✓ | Hom. | Multiple | ✓ |
| Absi, Archetti, Dauzere-Péres, and Feillet (2014) | Single | Single | ✓ | Het. | Multiple | ✓ |

Hom., homogeneous; Het., heterogeneous.

Table 3. Summary of comparison between paper problem, PRP, and VRP.

| Problem | Routing | Production operations |
|---------|---------|-----------------------|
| VRP     | Between customers of a plant or distribution center | -- |
| PRP     | Between customers of a plant or distribution center | By one plant |
| Paper problem | Outsourced and between different manufacturers | By different manufacturers |
1. **Resource layer.** provides the manufacturing resources and abilities involved in the whole lifecycle of a product

2. **Perception layer.** is able to sense the physical resources and abilities, process the related data and information, and connect them to the wider internet.

3. **Resource virtualization layer.** is responsible for the virtualization of manufacturing resources and abilities, and encapsulating them into the manufacturing cloud service (MCS).

4. **Cloud service layer.** primarily provides two categories of service, the MCS and the Cloud manufacturing core services (CMfg). MCS is the result of service encapsulation of manufacturing resources and ability, which can be invoked by end users. CMfg are the main services provided by the platform for the three category users (i.e. the provider, operator, and consumer) to manage, access and invoke MCSs, including match and search, scheduling, evaluation, and so forth.

5. **Application layer.** develops a special manufacturing application system according to specific demands, such as cooperative supply chain management system, Cloud manufacturing-based ERP, etc.

6. **Portal layer.** provides various man–machine interaction interfaces for user access.

7. **Enterprise cooperation application layer.** including commerce cooperation, business cooperation, collaborative design, and manufacturing cooperation.
8. **Knowledge layer.** provides various knowledge needed in the other layers such as manufacturing domain knowledge, process knowledge, model knowledge, etc.

9. **Cloud security layer.** provides different security architecture, mechanisms, and strategies for the Cloud manufacturing system.

10. **Wider internet layer.** provides the communication environment to all resources, services, users, and operations in the Cloud manufacturing system.

The important issue in this model is that cloud operator is responsible for deciding based on a predefined objective function to determine which of the tasks/subtasks should be carried out by which of the plants. In this paper, composition of production and transportation services in cloud supply network is considered, and it is assumed that service providers who provide different production services are located in the different geographic locations. This network is considered for only one product and one period. Further assumptions and explanations are presented in the following.

The network has several layers in which each layer includes certain production service providers. In addition, a number of specified production services can be performed in each of the layers, and each service provider is capable of performing certain production services with regard to its capacity. There are two fleet sets for transportation, first for the transportation among service providers and terminal with unlimited capacity and the second for transportation between two layers with limited capacity. It is possible to move from each layer only to its next layer by the defined fleet between the two layers, because the operations should be accomplished with a specific sequence. It is assumed that operation process chart (OPC) in this network is unique that means there are specific steps to produce the product that cannot been changed. Finally, there are customers with different certain demands that can be clustered in the last layer. It means depending on the final output, it is determined which customers are covered so that maximum customer demands are fulfilled and minimum cost for transporting the products to customers is resulted.

Manufacturing service providers are plants with various manufacturing capabilities, and transportation service providers are companies with a fleet of transportation. In order to solve such a problem, at first necessary operations should be identified and presented in the form of OPC. Finally, after allocating the services to the layers, the optimal combination of services is determined based on a mathematical model and it becomes clear that each service should be performed and transported by which plants and logistic service providers, respectively.

This model can be used in production of expensive goods especially automobile industry as the execution of manufacturing services in different places leads to transportation cost among resources. Therefore, the outcome of the product should be considerable in order to cover such costs. As shown in the literature review section, as the problem origins form the VRP and PRP which both of them are NP-hard in nature, this problem is also NP-hard and by increasing the problem size, the number of parameters and constraints are increased in a large amount, and solving the problem in a reasonable time will not be possible.
3. Proposed method

In this section, at first a supply network is designed to produce a specific product by using of production and routing services. The goal of such a network is to satisfy customers’ needs by appropriate optimal composition of different services available in the network. In the next step, a mathematical model is proposed to plan and generate routing and production in the cloud platform. This model is based on the assumptions in Section 2 and will be described the detail in the following.

According to Figure 3, designed network includes some layers, nodes, and terminals. Terminals are considered as a temporary place to collect output of each factory and transfer them to the next layer terminal. In each layer, there are specified nodes as service providers. In addition, in each layer specified production services can be carried out, and each node is capable of doing certain production services with limited capacity in a specific time duration.

The number of layers is \( n + 1 \) which \( n \) layer is assigned to various production services and the last layer is assigned to customers. The transportation is outsourced and performed by service provider of logistic, and it is from one layer to its next layer and between nodes and corresponding terminal. The fleet between two layers is dedicated only to the transportation between those two layers and cannot be used for transportation between other layers.

In order to propose a mathematical model, it is first necessary to determine the number of layers of the supply network, and services should be performed on each layer. This will be done using the algorithm proposed below.

3.1. Algorithm of service allocation to layers

Proposed algorithm uses OPC to assign services to layers. OPC is graphical figure of production operations and shows production steps of a special product. Each operation

![Figure 3. Cloud supply network.](image)
is showed as a node and the connection between nodes is according to sequence of execution of them.

The algorithm is performed in reciprocal form consists of downward movement and upward movement. The number of layers is gained from downward movement and in the return movement, the assignment of operations to the layers is done. To simplify the problem in downward movement, start counting from the node that has the biggest distance (greatest number of nodes with the final product node) and move vertically to down.

Initially, in the first node, the number of layer \((i)\) is 1, then by moving down and seeing composition of an operation with another operation, the number of layers are increased by one. This is repeated until reaching to last node. Finally, the number obtained in the final node is the number of layers required for all operations to be accomplished. In other words, the number of layers needed to design the network is the number of slices for the composition in downward movement.

In upward movement for node whose layer number is unknown, the number layer of the node in which they are combined with another operation is reduced to one and the resulting number is attributed to that operation. This is also repeated for other operations, and finally the layer number of each operation is specified and operations are assigned to layers. The OPC is interpreted by the proposed algorithm and operations are assigned to layers so that prerequisite operations of one operation is performed in previous layer. Finally, a matrix is obtained which determines each operation should be done in which layer.

As shown in Figure 4, the implementation of this algorithm is shown on a sample of OPC. In this diagram, five operations are required to produce a product and must be executed in a special sequence according to the shape. The proposed algorithm is applied to this graph and the number of layers required for the supply network is 3. Also, the layer in which each operation should be performed is also specified and written besides the node.

![Figure 4. Algorithm of allocation of services to layers.](image-url)
After allocation of services to the layers of supply network, according to Figure 5 optimum service composition can be obtained through the following steps:

1. Selection of necessary resources in each layer to perform the services considering limited capacity of each resource and its production cost.
2. Collecting result of each operation in each node and transferring them to the terminal to transfer to next layer.
3. Transferring materials and components to the next layer terminal and distributing between selected resources for subsequent operations.
4. Repetition the previous steps to the nth layer.

Finally, after completing the above steps for the n layer and selecting the final services in the nth layer, the customers can be satisfied should be selected.

3.2. Mathematical formulation

The indexes, parameters and decision variables are introduced in the following. Then the objective function and constraints of the model are described. The model presented in this section is a nonlinear hybrid model.

3.2.1. Model indexes

The problem indexes are considered as follows:
   i: shows the layer number
   J_i: shows the nodes in layer i

3.2.2. Model parameters

Problem parameters are supposed as follows:

![Service composition in cloud supply network.](image)
3.2.3. Decision variables

The decision variables for the problem are as follows:

- $J_i$: start time related to the execution of the operation $J_i$ in node $J_i$
- $J_i$: the unit number of operation $J_i$ performed in node $J_i$
- $J_i$: if operation $J_i$ executed in node $J_i$, the value is 1 otherwise zero.
- $J_i$: if at least one operation executed in node $J_i$ the value is 1 and otherwise zero.
- $J_i$: if cluster $J_i$ visited, the value is 1 and otherwise zero.

3.2.4. Mathematical model

Max $\left[ \left( \sum_{c_a} X_{C_a} D_{C_a} \right) R \right]$

$$- \left[ \sum_{k} \sum_{h} \sum_{i} P_{O_k, J_i} X_{O_k, J_i} V_{O_k, J_i} + \sum_{i} T_{C_i} + \sum_{h} \sum_{i} T_{C_h} Y_{J_i} + \left( \sum_{k} \sum_{h} T_{O_k, J_h} \right) W_c + \sum_{h} X_{C_h} D_{C_h} \right]$$

(1)

$$X_{O_k, J_i} \leq FOP_{O_k, J_i}, \ \forall \ O_k, J_i$$

(2)

$$X_{O_k, J_i} \leq Y_{J_i}, \ \forall \ O_k, J_i$$

(3)

$$\sum_{J_i} X_{O_k, J_i} V_{O_k, J_i} = ROP_{O_k, \ I}; \ \forall \ O_k$$

(4)

$$\sum_{O_k} X_{O_k, J_i} V_{O_k, J_i} \leq P_{J_i}, \ \forall J_i$$

(5)
\[
\sum_{O_k} \sum_{J_i} X_{O_k, J_i} V_{O_k, J_i} \leq C_{Vi}, \quad \forall i \tag{6}
\]

\[
T_{O_k, J_i} \leq M X_{O_k, J_i}, \quad \forall O_k, J_i \tag{7}
\]

\[
V_{O_k, J_i} \leq M X_{O_k, J_i}, \quad \forall O_k, J_i \tag{8}
\]

\[
T_{O_{k'}, J_{i+1}} \geq (T_{O_k, J_i} + D_{O_k, J_i} + T_i + T_{J_i+1}) X_{O_k, J_{i+1}} X_{O_k, J_i} Pr_{O_{k'}}, \quad \forall O_{k'}, O_k, J_i, J_{i+1} \tag{9}
\]

\[
\sum_{C_n} X_{C_n} D_{C_n} \leq I \tag{10}
\]

\[
T_{O_k, J_i} \geq 0, \quad V_{O_k, J_i} \geq 0, \quad X_{O_k, J_i} \in \{0, 1\}, \quad Y_{J_i} \in \{0, 1\}, \quad X_{C_n} \in \{0, 1\} \tag{11}
\]

The objective function (1) expresses the maximization of profit, which is equal to the difference between the sale of the product and the production, transportation and waiting time (Time duration to produce and deliver the product) costs associated with it. Constraints (2) guarantee that an operation is executed in a node if \( F_{op} = 1 \). Constraints (3) refer that if at least one production operation is performed in a node, then the \( J_i \) variable of that node is 1. Such a restriction ensures that, if multiple operations are performed in a node, the transportation cost is calculated once for the model. Constraints (4) ensure that in each layer production operations should be executed proportional with final demand. Constraints (5) are related to the production capacity of each node and constraints (6) represent fleet capacity that is available between two layers. Constraints (7) indicate the relationship between decision variables so that start time of an operation take a value if \( J_i = 1 \). Constraints (8) are the relationship between decision variables so that value of operation in each node set if \( J_i = 1 \). Constraints (9) are related to the start time of each operation, so that the start time of the operation in a node should be after the prerequisite operations in the nodes of the previous layer considering the transportation times. Constraints (10) guarantee that total demand of customers should be less than or equal to specified I. Constraint (11) defines the type of variables and the limitation of them which \( J_i \) and \( J_i \) are continuous variables and the others are integer variables.

### 4. Genetic algorithm

Genetic algorithm was proposed by John Holland and his fellow colleagues in Michigan University. This algorithm is one of the most important meta-heuristic algorithms, which is recognized as one of the optimization random methods. Genetic algorithms are search methods based on principles of natural selection and genetics which solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, stochastic, or highly nonlinear. The genetic algorithm can address problems of mixed integer programming, where some components are restricted to be integer-valued (Holland, 1992).
The algorithm works in an iterative process and starts from a population of randomly generated individuals. In each iteration, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. Then the more fit individuals are stochastically selected from the current population, and each individual is changed by crossover and mutation operators to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. The function of the algorithm and its parameters will be described in detail in the following (Kumar, Husian, Upreti, & Gupta, 2010):

- **Encoding.** This step may be the most difficult step in the algorithm. The genetic algorithm deals with encoded form instead of working on the parameters or problem variables. In fact, the encrypted form is the probable answer of the problem. In this case, each chromosome is a solution of the problem that can be feasible or unfeasible. As shown in Figure 6, the selected method is encoding based on the value. In this method, the chromosomes are represented as a string of natural numbers, each of which is related to a specific variable in the problem. The ordering of these numbers is important and the length of the string is exactly equal to the number of variables defined in the problem.

- **Evaluation.** In optimization problems, the objective function is applied as fitness function and the suitability of the answer is measured by the amount obtained from the fitness function. In this study, the objective function in mathematical problem is selected as fitness function.

- **Selection.** Design of the selection method is such that it is possible to select members with more accommodation. The selection should be in such a way that each new generation has better accommodation average than previous generations. The selection method in the proposed algorithm is selection based on roulette wheel.

- **Crossover.** The most important operator in genetic algorithm is crossover. This operator remove the diversity of the population because it allows good chromosome find each other. The method chosen in this study is a multi-point crossover. In multipoint crossover, \( m \) position for replacement is selected randomly from \( J_i \) which \( J_i \) is the point, \( l \) is length of chromosome and \( m \) is considered two. Then to create new offspring bits between specified points in the parents are changed with each other.

![Figure 6. Encoding of the problem.](image-url)
• **Mutation.** Mutation is another operator, which creates other feasible solution. In this study, selected method is changing the ordering of the genes. This method is used in algorithms that encoding is based on value and value of the two genes in the chromosome are exchanged.

• **Decoding.** Decoding is inverse of coding operator. In this step, it is necessary to decode the solutions after the algorithm find the best solution.

The problem studied in this research is a mixed integer programming and a NP-hard problem, as well as the time to solve the samples with a large number of variables is high. For this reason, population-based algorithm (genetics) is used as one of the meta-innovative algorithms for problem solving. As regards efficiency of genetic algorithm depends on the value of its parameters, selected suitable value for parameters is very important.

Naturally, GAs have many parameters that need to be tuned in order to find the best performance for any optimization problem. The main parameters are the population size, the crossover probability (Pc), and the mutation probability (Pm) whose values depend on the type of the problem. In order to identify suitable values for these parameters to solve the problem, the performance of the GA evaluated for several different sets of these parameters, and finally the optimum values are obtained as follows:

1. Population size: It is equal to 100.
2. Crossover possibility: Since composition has a significant role in the growth of the average of population accommodation, Pc is better to be between 0.7 and 0.8. In this study it is considered 0.8.
3. Mutation possibility: The mutation should not occur too much, because then the genetic algorithm will turn into a completely randomized search which is considered 0.1 in this paper.
4. Termination condition: Specific number of iteration is selected as termination condition, which is considered 50 repetitions for small instance and 100 repetitions in large instance.

5. **Computational results**

In this section by solving six distinct test-cases of small and large instances, the efficiency of the proposed genetic algorithm in the previous section is investigated. Since no test-cases are publicly available for the problem, these instances are randomly generated.

Table 4 highlights test results for instances with up to 10 layers and 50 factories. In this table second column to fifth demonstrate the number of layers (N), the number of operations (M), the number of factories (F), and the number of customer clusters (C). Additionally, $J_i$ and $j_i$ are the best objective function values returned by GAMS and genetic algorithm, respectively, and $j_i$ is the runtime of problem solving in GAMS. Moreover, the results of the exact solution approach and the genetic algorithm are compared and reported as gap in last column which is computed as follows:

$$GAP = \left[ \frac{\sigma_{\text{GAMS}} - \sigma_{\text{Genetic}}}{\sigma_{\text{Genetic}}} \right] \times 100\%$$

(12)

As can be seen, there is a gap between GAMS solution and genetic solution in instance 1 to 4. It is due to the fact that genetic algorithm is a meta-heuristic algorithm.
finding a solution close to the exact solution. In addition, the largest instance (problem no. 4) that can be solved optimality by GAMS contained 10 layers and 20 factories, and therefore, in instances 5 and 6, GAMS cannot find an optimum solution, and the problem is solved in a reasonable time by proposed genetic algorithm.

### 6. Discussion and conclusion

#### 6.1. Sensitivity analysis

For evaluation the effect of objective function coefficients on the optimum solution, a sensitivity analysis has been conducted through modification of production and transportation costs. The effect of this modifications are presented in the following:

1. **Sensitivity analysis of production costs.** In order to analyze the sensitivity of production costs, an instance with four layers, four operations, seven plants, and five customer clusters is considered (sample 1). Then, by increasing the production costs of factories by 10% for four times and keep other parameters constant, the total profit of the network is calculated for sample 2 to sample 5. As can be seen in Figure 7, increasing the production cost by 40% in sample 5 causes that the total profit is decreased by almost 6.5% in comparison with sample 1.

2. **Sensitivity analysis of transportation costs.** Sensitivity analysis of transportation costs between each factory and the corresponding terminal is carried out as same

![Figure 7. Sensitivity analysis of production cost.](image-url)
as the production costs. The results of this experiment are displayed in Figure 8, and it is observed that by increasing transportation cost by 40% in sample 5 total profit is decreased by almost 0.69% as compared to sample 1. Therefore, the effect of changing transportation costs is less than production costs on the profit of the network because production cost is considered for unit of operation, which is multiplied by the amount of performed operation to calculate total production costs in objective function.

6.2. Conclusion

Cloud manufacturing is the future model of manufacturing systems, and many of manufacturing companies will upgrade their systems into cloud-based manufacturing models. The high conformity and adaptability of cloud-based manufacturing systems are the keys that enable this model for future manufacturing environments and post-Industry 4.0 era. In a cloud-based manufacturing model the complexity of planning for high personalized products can be alleviated by distributed planning agents and algorithms; also the scarcity of some resources that commonly causes high production costs or inability of production can be solved by exploiting extremely distributed planning and outsourcing. The other competitive advantages of such models that in near future will revolutionize the production systems is the potential of exploitation from a crowd of manufacturing resources in order to create virtual companies who produce products that are monopolized by a number of exclusive companies.

The authors used prior models in production and routing problems to model a piece of such capabilities of cloud-based manufacturing models. Then the paper presents a brief discourse about these models and then exploits them for Cloud manufacturing environments. Also in this paper, the concept of Cloud manufacturing and some of the related articles have been investigated and then the notion of service composition in Cloud manufacturing is presented. Service composition is the kernel of cloud-based manufacturing models and systems which attracted a lot of attention in recent years. This paper using a different look tried to address the service composition concerns by exploiting VRP and PRP models.
Since we want to have PRP in Cloud manufacturing, the articles related to PRP and VRP in traditional supply chain are reviewed. The important contribution of this study is considering transportation concept in service composition problem and the execution of production operations in different places.

So, at first a supply network with production and transportation services in layer structured is considered, and then a new algorithm to allocate the services to layers of network and a mixed integer model of the problem are presented. The model seeks to maximize the profit by minimizing production costs, transportation costs and waiting time costs, and realizing the maximum demand of customers.

Moreover, in order to solve the model for large instances a genetic algorithm is exploited; then the efficiency of the proposed algorithm is evaluated through random six test-cases. The results show that the proposed algorithm is efficient to solve the small instances by a little gap to exact solution and it can find the best solution for large instances in a reasonable time. Finally, a sensitive analysis of transportation and production cost is carried out which prove that production costs have most important role in total cost of the network.

6.3. Future works

Based on experience gained in this research, the suggestion for this research are as follows:

1. Providing an algorithm for customer clustering while in this research customer clusters are initially specified and presented as problem assumptions.
2. Sharing transportation fleet among the factories of the layers.
3. In the future, terminals in each layer can be omitted in the supply network and the transportation can be done directly from each factory to other factory for subsequent production operations.
4. Product storage and its cost can be considered in the subsequent research.
5. The multi-period and multi product version of the problem is also an interesting perspective.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Omid Fatahi Valilai http://orcid.org/0000-0001-7087-6946

References

Absi, N., Archetti, C., Dauzere-Péres, S., & Feillet, D. (2014). A two-phase iterative heuristic approach for the production routing problem. Transportation Science, 49(4), 784–795.
Adulyasak, Y., Cordeau, J.-F., & Jans, R. (2013). Formulations and branch-and-cut algorithms for multivehicle production and inventory routing problems. *INFORMS Journal on Computing, 26*(1), 103–120.

Adulyasak, Y., Cordeau, J.-F., & Jans, R. (2015). Benders decomposition for production routing under demand uncertainty. *Operations Research, 63*(4), 851–867.

Archetti, C., Bertazzi, L., Paletta, G., & Speranza, M. G. (2011). Analysis of the maximum level policy in a production-distribution system. *Computers & Operations Research, 38*(12), 1731–1746.

Bard, J. F., & Nananukul, N. (2009). The integrated production–Inventory–Distribution–Routing problem. *Journal of Scheduling, 12*(3), 257–280.

Boudia, M., Louly, M. A. O., & Prins, C. (2007). A reactive GRASP and path relinking for a combined production–Distribution problem. *Computers & Operations Research, 34*(11), 3402–3419.

Boudia, M., Louly, M. A. O., & Prins, C. (2008). Fast heuristics for a combined production planning and vehicle routing problem. *Production Planning and Control, 19*(2), 85–96.

Braekers, K., Ramaekers, K., & Van Niewenhuyse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering, 99*, 300–313.

Cao, E., & Lai, M. (2010). The open vehicle routing problem with fuzzy demands. *Expert Systems with Applications, 37*(3), 2405–2411.

Delaram, J., & Valilai, O. F. (2016). Development of a novel solution to enable integration and interoperability for cloud manufacturing. *Procedia CIRP, 52*, 6–11.

Delaram, J., & Valilai, O. F. (2017). A novel solution for manufacturing interoperability fulfillment using interoperability service providers. *Procedia CIRP, 63*, 774–779.

Delaram, J., & Valilai, O. F. (2018a). An architectural view to computer integrated manufacturing systems based on axiomatic design theory. *Computers in Industry, 100*, 96–114.

Delaram, J., & Valilai, O. F. (2018b). An architectural solution for virtual computer integrated manufacturing systems using ISO standards. *Scienica Iranica*. doi:10.24200/SCI.22018.20799

Fleszar, K., Osman, I. H., & Hindi, K. S. (2009). A variable neighbourhood search algorithm for the open vehicle routing problem. *European Journal of Operational Research, 195*(3), 803–809.

Hashemian, S. V., & Mavaddat, F. (2006). A graph-based framework for composition of stateless web services. *Web Services, 2006. ECOWS’06. 4th European Conference on, IEEE*.

Holland, J. H. (1992). Genetic algorithms. *Scientific American, 267*(1), 66–73.

Houshmand, M., & Valilai, O. F. (2012). LAYMOD: A layered and modular platform for CAx product data integration based on the modular architecture of the standard for exchange of product data. *International Journal of Computer Integrated Manufacturing, 25*(6), 473–487.

Houshmand, M., & Valilai, O. F. (2013). A layered and modular platform to enable distributed CAx collaboration and support product data integration based on STEP standard. *International Journal of Computer Integrated Manufacturing, 26*(8), 731–750.

Johnson, M. E. (2006). Supply chain management: Technology, globalization, and policy at a crossroads. *Interfaces, 36*(3), 191–193.

Kofler, K., Ul Haq, I., & Schikuta, E. (2009). A parallel branch and bound algorithm for workflow QoS optimization. Parallel processing, 2009. *ICPP’09. International Conference on, IEEE*.

Kumar, M., Husian, M., Upreti, N., & Gupta, D. (2010). Genetic algorithm: Review and application. *International Journal of Information Technology and Knowledge Management, 2*(2), 451–454.

Lei, L., Liu, S., Ruszczynski, A., & Park, S. (2006). On the integrated production, inventory, and distribution routing problem. *IIE Transactions, 38*(11), 955–970.

Li, X., Leung, S. C., & Tian, F. (2012). A multistart adaptive memory-based tabu search algorithm for the heterogeneous fixed fleet open vehicle routing problem. *Expert Systems with Applications, 39*(1), 365–374.

Liu, R., & Jiang, Z. (2012). The close–open mixed vehicle routing problem. *European Journal of Operational Research, 220*(2), 349–360.

MirHassani, S., & Abolghasemi, N. (2011). A particle swarm optimization algorithm for open vehicle routing problem. *Expert Systems with Applications, 38*(9), 11547–11551.

Repouissis, P. P., Tarantilis, C. D., Bräysy, O., & Ioannou, G. (2010). A hybrid evolution strategy for the open vehicle routing problem. *Computers & Operations Research, 37*(3), 443–455.
Salari, M., Toth, P., & Tramontani, A. (2010). An ILP improvement procedure for the open vehicle routing problem. *Computers & Operations Research, 37*(12), 2106–2120.

Sariklis, D., & Powell, S. (2000). A heuristic method for the open vehicle routing problem. *Journal of the Operational Research Society, 51*, 564–573.

Tao, F., Hu, Y., & Zhang, L. (2010). Theory and practice: Optimal resource service allocation in manufacturing grid. *China Machine Press. Ed, 1*, 1–18.

Valilai, O. F., & Houshmand, M. (2010). INFELT STEP: An integrated and interoperable platform for collaborative CAD/CAPP/CAM/CNC machining systems based on STEP standard. *International Journal of Computer Integrated Manufacturing, 23*(12), 1095–1117.

Valilai, O. F., & Houshmand, M. (2013). A collaborative and integrated platform to support distributed manufacturing system using a service-oriented approach based on cloud computing paradigm. *Robotics and Computer-Integrated Manufacturing, 29*(1), 110–127.

Valilai, O. F., & Houshmand, M. (2014a). Depicting additive manufacturing from a global perspective; using cloud manufacturing paradigm for integration and collaboration. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 229*(12), 2216–2237.

Valilai, O. F., & Houshmand, M. (2014b). A manufacturing ontology model to enable data integration services in cloud manufacturing using axiomatic design theory. In D. Schaefer (Ed.), *Cloud-Based Design and Manufacturing (CBDM)* (pp. 179–206). London: Springer. doi:10.1007/978-1-003-1319-07398-07399_07397

Valilai, O. F., & Houshmand, M. (2014c). A platform for optimisation in distributed manufacturing enterprises based on cloud manufacturing paradigm. *International Journal of Computer Integrated Manufacturing, 27*(11), 1031–1054.

Wu, D., Greer, M. J., Rosen, D. W., & Schaefer, D. (2013a). Cloud manufacturing: Drivers, current status, and future trends. *ASME 2013 International Manufacturing Science and Engineering Conference collocated with the 41st North American Manufacturing Research Conference, American Society of Mechanical Engineers.*

Wu, D., Greer, M. J., Rosen, D. W., & Schaefer, D. (2013b). Cloud manufacturing: Strategic vision and state-of-the-art. *Journal of Manufacturing Systems, 32*(4), 564–579.

Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing, 28*(1), 75–86.

Yu, S., Ding, C., & Zhu, K. (2011). A hybrid GA–TS algorithm for open vehicle routing optimization of coal mines material. *Expert Systems with Applications, 38*(8), 10568–10573.

Zachariadis, E. E., & Kiranoudis, C. T. (2010). A strategy for reducing the computational complexity of local search-based methods for the vehicle routing problem. *Computers & Operations Research, 37*(12), 2089–2105.

Zeng, C., Guo, X., Ou, W., & Han, D. (2009). Cloud computing service composition and search based on semantic. *IEEE International Conference on Cloud Computing, Springer.*