Fleet Sizing of Dynamically Routed In-plant Milk-run Vehicles Based on a Genetic Algorithm

Dimensionierung von Routenzugsystemen mit dynamischen Routen auf Basis eines Genetischen Algorithmus

Fabian Hormes  
Amin Siala  
Christian Lieb  
Johannes Fottner

Chair of Materials Handling, Material Flow, Logistics  
Department of Mechanical Engineering  
Technical University of Munich (TUM)

In-plant milk-run (MR) systems enable efficient supply of assembly lines in small lot sizes. One major challenge for MR systems are demand fluctuations and short-term changes within schedules. Dynamic control strategies aim at increasing flexibility and efficiency of MR systems in volatile environments. This paper presents an application-oriented approach for determining the fleet size of an MR system with dynamically controlled routes based on a genetic algorithm. The approach is evaluated and discussed using a case study from a commercial vehicle manufacturer. The results show that the approach enables effective analytical dimensioning of MR systems with dynamic routes. In addition, the case study indicates that the implementation of dynamic routes can lead to a reduction in fleet size.

[Keywords: In-plant milk-run, control strategies, routing problems, genetic algorithms]


[Schlüsselwörter: Routenzugsysteme, Steuerungsstrategien, Routing-Probleme, Genetische Algorithmen]

1 INTRODUCTION

In-plant milk-runs (MR) enable efficient supply of assembly lines in small lot sizes [Tak06]. They deliver material on defined routes throughout the factory following predetermined schedules [Har03]. Routes and schedules are defined in the planning process of an MR system based on average material consumption and layout restrictions [Bru12]. Each route is assigned a certain number of operators and vehicles. The routes start and end at an MR station (depot) close to a warehouse or supermarket. During each MR cycle (tour), a certain number of parts are delivered to the line-side points of use (POU). A tour includes three or more stops at different POUs. However, the number of parts delivered on a tour is limited by the capacity of the MR vehicle [VDI5586a]. Although MR systems are widely used in the industry due to efficiency advantages over single or unit-load transport, challenges arise regarding fluctuating transport demands or short-term changes within schedules [Klu13, Lie18]. The dynamic control of MR systems is a concept to increase flexibility of in-plant MR systems and to ensure efficiency despite fluctuations [Kle15].

Existing methodologies for MR system dimensioning focus on MR systems with static, previously defined routes. In this paper we present a fleet sizing approach for MR systems with flexible, dynamically controlled routes. The approach extends the standard procedure defined in VDI 5586 by a genetic algorithm that generates dynamic routes and enables practitioners to determine the required number of MR vehicles under given constraints. The paper is structured as follows: in section 2, we describe the theoretical background regarding MR system planning including the standard fleet sizing approach and control strategies. In section 3, the development of the fleet sizing approach for MR systems with dynamic routes and the genetic algorithm are outlined. In section 4, we evaluate and discuss the approach using a case study example of a commercial vehicle manufacturer. Section 5 contains a summary and an outlook on future work.
2 IN-PLANT MILK-RUN (MR) SYSTEMS DESIGN

2.1 PLANNING OF MR SYSTEMS

The basic planning approach for MR systems is defined in the VDI 5586 standard. The planning framework consists of four major planning steps [VDI5586b]. Step 1 addresses the collection of relevant data for the planning task. Relevant data includes, for example, the different types of parts, the required throughput and the layout restrictions. In step 2, planning alternatives are generated. A planning alternative is created by combining different processes and technology variants, e.g. with regard to the loading and unloading of MR vehicles [Keu18]. Step 3 includes the dimensioning and determination of the required resources, e.g. the number of MR vehicles (fleet size) or operators, for each planning alternative generated. Finally, in step 4, all planning alternatives are evaluated based on performance indicators and economic factors.

2.2 FLEET SIZING ACCORDING TO VDI 5586

The average MR cycle time ($t_{cyc}$) can be calculated using the average travel time ($t_T$), the time for acceleration and deceleration at the stops ($t_S$), the average train loading time at the depot ($t_L$), and the average unloading and handling time at a stop ($t_U$):

$$t_{cyc} = t_T + t_S + t_L + t_U$$

Given the required throughput per part k to be delivered at stop m ($T_{k,m}$), the throughput of the total route ($T_R$) with the total number of stops ($n_m$) and parts ($n_k$) can be calculated as follows:

$$T_R = \sum_{n_m} \sum_{n_k} T_{k,m}$$

Based on the total route throughput, the tour start interval ($t_{FS}$) can be calculated using the train capacity ($C_{max}$) and its degree of utilization ($\eta$):

$$t_{FS} = \eta \cdot \frac{C_{max}}{T_R}$$

The required number of MR vehicles ($n_{MR}$) to deliver the average throughput $T_R$ is then calculated as follows:

$$n_{MR} = \frac{t_{cyc}}{t_{FS}}$$

The presented fleet sizing approach for MR systems enables simple and effective system dimensioning in the early planning stages. However, as stated by the authors, the approach only refers to single-route MR systems with previously determined and static routes [VDI5586b].

2.3 CONTROL STRATEGIES FOR MR SYSTEMS

One major challenge for traditional MR systems with static routes and schedules are demand fluctuations. Demand fluctuations arise, for example, due to a high number of product variants, changes within the production program or deviations in process times [Lie18].

One strategy for dealing with demand fluctuations is the dynamic control of MR systems [Emd12, Aln15, Hor17, Kle19]. Typical decision problems in in-plant vehicle control include i) the selection and assignment of orders to vehicles (dispatching), ii) the determination of routes between start and end points (routing), and iii) the definition of departure and arrival times (scheduling) [Co91, Kle19].

A transport control system allows for dynamic control of MR vehicles based on actual demand. Therefore, no fixed routes or schedules are required. The control system calculates routes and schedules in real-time in order to optimize MR transports under given constraints [Hor17].

3 FLEET SIZING APPROACH FOR DYNAMICALLY Routed MR SYSTEMS

3.1 ROUTING PROBLEMS

Due to dynamic control, routes and travel times ($t_T$) in dynamic MR systems vary between each MR cycle. However, average stop times ($t_S$), loading times ($t_L$), and unloading times ($t_U$) can be assumed to be constant and deterministic. In order to determine the number of MR vehicles ($n_{MR}$) in a dynamic MR system, an approach is therefore needed that allows the calculation of average travel times assuming dynamic routes for each MR cycle.

To determine an average travel time Urru et al. propose an exact approach that determines all possible combinations of routes between the depot and the customers of the system under consideration. However, due to the complexity of the decision problem, the exact approach is only applicable to a very limited number of POUs [Urr18].

The underlying decision problem corresponds to the well-known Vehicle Routing Problem (VRP) [Dan59]. The VRP aims to minimize transport costs associated with transporting goods between a depot and several customers using a fleet of vehicles with limited capacity [Cla64]. Since the complexity of the VRP is to be regarded as NP-hard, i.e. cannot be solved optimally for arbitrary instances in polynomial time, heuristics or metaheuristics in particular are used in addition to exact procedures in order to find solutions for the problem instance [Kil13].

Genetic algorithms (GA) are a form of metaheuristics based on biologically inspired methods that are used in optimization problems [Kra17]. GA are none other than the equivalent of the biological evolution in algorithmic form.
GA have been successful at finding decent solutions to complex routing problems [Faz15]. In the following we present a genetic algorithm that allows the determination of dynamic routes and average travel times ($t_{T,d}$) assuming dynamic routes. Based on the average travel times ($t_{T,d}$), the number of MR vehicles can be calculated following the approach for static routes according to VDI 5586.

3.2 DEVELOPMENT OF THE GENETIC ALGORITHM

3.2.1 INTRODUCTION

GA are composed of a set of candidate solutions which individually provide a solution to our optimization problem. The set of candidate solutions is called “population.” In our case, each solution represents a vector containing a set of stops and tours. At the beginning, an initial population is randomly generated, which allows to cover a large solution space at an early stage of the optimization (Figure 1).

![Figure 1. General flowchart for the genetic algorithm](image)

Similar to nature, GA mix the genetic material of two selected parent solutions to create child solutions (crossover). The one-point crossover is the most popular crossover approach [Kra17]. It involves splitting up two parent solutions at an arbitrary point and reassembling them to get two child solutions (Figure 2). Another nature-inspired approach is mutation. It makes selective changes within the child solutions. A possible mutation is illustrated in Figure 2. The mutation operator randomly swaps two indices within the child solutions. The aim of mutation is to search a larger solution space and to escape local optima.

The example in Figure 2 includes five parts (1, 2, 3, 4, 5). The required throughput of part 1 is two units. The throughput of parts 2, 3, 4, and 5 is one unit. Therefore, the total throughput $T_r$ adds up to six units. The delivery vector $d$, for parent solution $s$ contains the delivery order of all units ($d_1 = [3, 2, 1, 4, 5, 1]$). The crossover point is at index two of the delivery vector. Child solutions are created by combining parent solutions 1 and 2. The tour vector $t_r, s$ is generated after mutation and contains the units delivered on tour $t$ for child solution $s$ ($t_{1,1} = [4, 2, 5], t_{1,2} = [3, 1, 1], t_{2,1} = [2, 3, 1], t_{2,2} = [1, 5, 4]$). The number of parts to be delivered during a tour depends on the capacity of the MR vehicle. In our example, the maximum capacity $C_{max}$ is three units.

![Figure 2. Crossover and mutation](image)

After creating a new population of solutions using crossover and mutation, it is necessary to evaluate the solutions. This is done with the help of a fitness function. Fitness functions assess the quality of the solutions according to the optimization problems. There are several ways to create a fitness function. Depending on the problem, it can impose costs or penalties for solutions that are not feasible and do not meet the constraints and requirements. In the case of multiple-objective optimization, where the solution is optimized to meet two or more objectives, the fitness function is the weighted sum of the individual fitness functions of the optimization. The fitness-based evaluation can be used in several steps during the algorithm, e.g. for the selection of parent solutions or the selection of child solutions (Figure 1).

Finally, a criterion for stopping the algorithm must be determined. One approach is to stop after a certain number of iterations (generations). Another approach is to stop when there are no more significant improvements in the fitness function. These types of approaches, however, carry the risk of being terminated at a local optimum. Restart strategies are therefore required.

3.2.2 FITNESS FUNCTION AND CONSTRAINTS

The fitness function allows to evaluate the generated solutions of a population $P_n$. The number of solutions of the population $n_P$ is set at the beginning. A higher $n_P$ allows a higher diversity of solutions. However, this also has a negative impact on the computation time. Each solution $s$ is
described by the solution vector \( P_s \in \mathbb{P} \). Each solution vector \( P_t \) contains the delivery vector \( d_t \), the tour vector \( t \), and the fitness value \( F_s \). The fitness value can be calculated according to the following equation:

\[
max F = \left( 1 - a \right) \sum_t \sum_i \sum_j x_{tij} d_{ij} + a \sum_t \sum_k y_{tk} c_{tk} \bar{d}
\]

\[
x_{tij} = \begin{cases} 
1 & \text{, if tour } t \text{ contains arc } (i,j) \\
0 & \text{, otherwise}
\end{cases}
\]

\[
y_{tk} = \begin{cases} 
1 & \text{, } c_{tk} > 0 \\
0 & \text{, otherwise}
\end{cases}
\]

\[
c_{tk} = \left( \sum_m z_{tkm} \right) \cdot \frac{T_{kT}}{n_T}
\]

\[
z_{tkm} = \begin{cases} 
1 & \text{, if part } k \text{ is delivered to stop } m \text{ of tour } t \\
0 & \text{, otherwise}
\end{cases}
\]

As stated in 3.2.1, the algorithm aims at minimizing the distance travelled and the travel time by the MR vehicles to serve all stops. Therefore, the first part of the fitness function covers the calculation of the total distance travelled of the corresponding solution. It adds up all distances \( d_{ij} \) between all nodes \( N \) that are part of a tour \( t \) for the total number of nodes \( n_N \) and tours \( n_T \) of the solution. The formulation of the first part of the equation is based on existing VRP formulations [Vai99]. The higher the total distance travelled, the lower the fitness value \( F_s \) of solution \( s \).

The second part of the equation serves as a penalty function and ensures that the deliveries of a number of units of part \( k \) are spread over multiple tours (split delivery routing). The function ensures a continuous supply of parts and low line-side inventories. The variable \( z_{tkm} \) counts the number of deliveries of part \( k \) on tour \( t \) for all parts \( K \), stops \( M \) and tours \( T \). Under the assumption of continuous supply, the optimal number of deliveries of part \( k \) for the set of tours \( T \) can be calculated by dividing the total throughput of part \( k \) \( (T_{kT}) \) by the number of tours \( n_T \). The penalty costs \( c_{tk} \) only occur if the actual number of deliveries of part \( k \) on tour \( t \) exceeds the optimal number of deliveries of part \( k \) per tour \( (c_{tk} > 0) \). In this case, \( y_{tk} = 1 \); in all other cases, \( y_{tk} = 0 \) (no penalty costs occur).

Following our example from 3.2.1, the actual number of deliveries per tour of part 1 of child solution 2 is one unit. The optimal number of deliveries of part 1 per tour is also one unit \((T_{1T} = 2 \text{ divided by } n_T = 2)\). As both units of part 1 are delivered on two different tours, no penalty costs occur \((c_{11}, y_{11} = 0)\). The same applies to parts 2, 3, 4, and 5 of child solution 2. However, for child solution 1 (after mutation) both units of part 1 are delivered on tour 2. Therefore, in this solution, a penalty for tour 2 and part 1 occurs \((c_{21} = 1)\). The penalty costs are weighted by the average distance \( \bar{d} \) between two nodes of the MR system. The penalty is added to the total distance travelled and reduces the total fitness value of the solution. The factor \( a \) is used to weight the influence of both parts of the fitness function.

The generation of solutions must comply with the following set of constraints:

- All tours \( t \in T \) start and end at a single depot
- The number of units delivered of part \( k \in K \) must be equal to the required throughput per part \( T_{kT} \)
- The total number of units delivered must be equal to the total required throughput \( T_T \)
- The deliveries of part \( k \in K \) to stop \( m \in M \) can be split between several tours
- The number of deliveries per tour is limited by the maximum capacity \( C_{\text{max}} \) of the MR vehicle

### 3.2.3 Modeling of the Genetic Algorithm

In the first step, the algorithm (Figure 3) uses the data specified by the user to create \( n_P \) random solutions for the initial population \( P \). As stated in 3.2.2, the generation of the first population of solutions must meet the predefined set of constraints. Once the first generation of solutions is generated, the quality of these solutions is assessed. The fitness function evaluates the generated solutions according to the total distance travelled and the distribution of the deliveries of the same type of parts over the different tours. The shorter the distance driven and the more spread and varied the part deliveries are, the better the fitness value will be.

After the generation of the initial population \( P_t \), the optimization process starts (Figure 3). In each iteration, several parent solutions \( P_{\text{parent}} \) are selected from the previous generation and build the basis for the next generation. The number of selected solutions depends on the parent quotient \( p \), which determines the number of parent solutions \( n_P \) from the number of total solutions in the population \( n_P \). Which solutions are selected as parent solutions, depends on their fitness value \( F_s \), compared to the total fitness of the population \( F_P \). Which determines the parent selection probability \( p_P \). Therefore, solutions with a higher fitness have a higher probability of being selected as parents.

In the next step, the crossover operator is used to form child solutions. First of all, two parent solutions are selected from the set of parents \( P_{\text{parent}} \). The selection is based on \( p_P \), which means that solutions with a higher fitness are selected more often to form a child solution. The parent solutions are combined so that the child solutions are still valid and respect all constraints. Therefore, a modified order crossover operator is used. The algorithm is displayed in Figure 4.
After selecting two parents, a random crossover index \( c \) is generated. The crossover index refers to the delivery vectors \( d_{\text{parent},1} \) and \( d_{\text{parent},2} \) of the selected parent solutions, which contain all deliveries in a specific order. Up to this index, the first part of the delivery vector \( d_{\text{child},1} \) of the first child solution is built from the first parent and the rest from the second parent. The same process is repeated for the delivery vector \( d_{\text{child},2} \) of the second child, with the only difference being that the first part is given by the second parent and the second part by the first parent (Figure 2).

To diversify the solution space and to avoid getting stuck in local optima, a mutation algorithm is used (Figure 4). The mutation function decides with a defined probability \( (p_m) \) whether a child solution is modified or not. If a child solution is selected for modification, the operator swaps two deliveries of the delivery vector. The operator is executed for both delivery vectors of the two child solutions. After mutation of the delivery vectors \( d_{\text{child},1} \) and \( d_{\text{child},2} \), the tour vectors \( t_{\text{child},1} \) and \( t_{\text{child},2} \) are generated based on the maximum capacity \( C_{\text{max}} \) of the MR vehicles. Finally, the operator calculates the fitness values for both child solutions and selects the child solution with the highest fitness \( F_{\text{child,fit}} \). The solution is added to the child population \( P_{\text{child}} \).

The crossover and mutation operators are repeated until the number of parent \( n_p \) and child solutions \( n_c \) equals the total number of solutions of a population \( n_p \). When the new population is complete, the best solution of the population \( P_{\text{fit}} \) is determined based on the highest fitness value. This entire process is repeated for \( i \) iterations (generations). After \( n_i \) iterations, the algorithm returns the best solution found \( P_{\text{fit}} \), which contains the best delivery and tour vector found in terms of the total distance travelled and the distribution of parts on different tours. It should be noted that \( P_{\text{fit}} \) is not necessarily the optimal solution of the optimization problem.

### 4 Case Study

#### 4.1 Description of the MR System

The case study was conducted in the final assembly of a commercial vehicle manufacturer. The investigated MR system consists of three predefined routes serving a total of 17 POUs with 17 different parts. One manually-operated MR vehicle is assigned to each route. The maximum capacity \( C_{\text{max}} \) of each MR is three units. The average travel speed is 2 m/s. The routes are illustrated in Figure 6. Table 1 displays the critical fleet sizing parameters for the MR system with static routes. For the calculation of the cycle time, the following times for \( t_s \), \( t_l \) and \( t_u \) are used:

- Stop time \( t_s = 40 \) s
- Loading time (depot) \( t_l = 60 \) s
- Unloading time (stop) \( t_u = 300 \) s
As displayed in Table 1, the throughput $T_R$, the number of tours per hour $n_T$ and the route length $L_R$ differ between the different routes. Route 3 is the longest route leading to the longest travel time $t_r$ per tour and the longest cycle time $t_{cyc}$. As route 3 has to deliver 12 parts per hour and the capacity is limited to three parts per tour ($C_{max} = 3$), the number of tours per hour $n_T$ is 4. Therefore, the required tour start interval $t_{TS}$ is 15 minutes (= 900 s) and the total distance travelled $D_R$ is 2876 meters. However, the total cycle time of route 3 is 760 s leading to an exact required number of 0.84 MR vehicles ($n_{MR}$) for route 3. As the assignment of vehicles to routes is fixed, the MR system with static routes requires a total number of three MR vehicles. The average utilization rate $\eta_T$ of the MR system is 0.76.

### 4.2 Fleet Sizing for Dynamic Routes

Table 2 displays the results generated by the developed GA. For the execution of the algorithm, we used the following parameters:
- population size $n_T = 1000$
- number of iterations $n_i = 100$
- parent quotient $p = 0.25$
- mutation probability $p_m = 0.7$
- fitness weight $\alpha = 0.85$

The results include a set of 12 routes with an average route length $L_R$ of 375.4 meters. The average travel time $t_{r,d}$ for dynamic routes is 187.7 seconds, resulting in an average cycle time $t_{cyc}$ of 587.7 seconds based on the constant stop, loading and unloading times. Table 3 summarizes the fleet size calculation for the MR system with dynamic routes. The exact number of required MR vehicles $n_{MR}$ is 1.96.

### 4.3 Discussion of Results

Figure 5 displays the optimization results of the GA. The algorithm was implemented in a Microsoft Visual Basic for Applications (VBA) programming environment to create an application-oriented tool that is easy to use by practitioners in logistics planning. The optimization was executed on an Intel i7 CPU (2.6 GHz, 16 GB RAM). The required CPU time was 56 seconds. After 50 iterations no significant improvements in distance and fitness were observed. In the case study described, the total distance travelled $D$ per hour is reduced from 6779 to 4505 meters when comparing static to dynamic routes. Given the optimized distance travelled, the fleet size can be reduced from 3 to 2 MR vehicles.

![Figure 5. Optimization results of the genetic algorithm](image-url)
The developed approach enables fleet sizing for MR systems with dynamic routes. In contrast to MR systems with static routes, the route length $L_R$ and the travel time $t_T$ vary between each tour. Therefore, the existing approach of the VDI 5586 standard cannot be applied. The genetic algorithm finds a set of tours that fulfills the constraints and optimizes route lengths and travel times. Due to the heuristic approach, the algorithm solves even larger problems in a reasonable time. The implementation in VBA ensures easy use of the fleet sizing tool. However, an implementation in another programming environment (e.g. C++) could further improve computation times [Erd17].

![Figure 6. Layout of the MR system](image)

### 5 SUMMARY AND FUTURE RESEARCH

This paper presents a metaheuristic-based fleet sizing approach for MR systems with dynamic routes. The developed GA optimizes the total distance travelled under given constraints and determines the average travel time $t_{T,d}$ per MR cycle for dynamic routes. Based on $t_{T,d}$, the average cycle time and fleet size can be determined according to existing dimensioning approaches for MR systems.

In conclusion, the approach enables a dimensioning of MR systems with dynamic routes in early planning stages and supports practitioners in the decision-making process whether to implement an MR system with dynamic routes. As the case study of the commercial vehicle manufacturer shows, an MR system with dynamic routes is more flexible and can offer efficiency advantages by reducing unnecessary travel times and the number of vehicles. However, the implementation of an MR system with dynamic routes is usually more complex and requires additional information and communication technology and availability of data.

Future work will focus on the inclusion of further constraints in the algorithm, such as variable capacity consumptions for different parts and throughput variations, to cover even more scenarios. Furthermore, the approach will be applied to various industry use cases to further evaluate the concept of dynamic routes in MR systems.

### LITERATURE


Fabian Hormes, M. Sc., Research assistant at the Chair of Materials Handling Material Flow Logistics at the Technical University of Munich.

Amin Siala, B. Sc., Student assistant at the Chair of Materials Handling Material Flow Logistics at the Technical University of Munich.

Christian Lieb, M. Sc., Research assistant at the Chair of Materials Handling Material Flow Logistics at the Technical University of Munich.

Prof. Dr.-Ing. Johannes Fottner, Professor and head of the Chair of Materials Handling Material Flow Logistics at the Technical University of Munich.

Address: Technische Universität München, Lehrstuhl für Fördertechnik Materialfluss Logistik (fml), Boltzmannstr. 15, 85748 Garching, Germany, Phone: +49 89 289 15931, E-Mail: fabian.hormes@tum.de