PREPARATION OF BIG DATA SETS FOR OPTIMIZATION OF INTERCITY FREIGHT DRIVER SCHEDULES

Olexandr Melnychenko[1], Myroslav Oliskevych[2,*], Nazar Khomyn[1]

[1]National Transport University, Faculty of Automotive Engineering, Department of Production, Repair and Materials Science, Kyiv, Ukraine
[2]Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies, Faculty of Mechanics, Energy and Information Technologies, Department of Automobiles and Tractors, Dubliany of Lviv Region, Ukraine

*E-mail of corresponding author: oliskevychm@gmail.com

Olexandr Melnychenko 0000-0001-7236-5891, Myroslav Oliskevych 0000-0001-6237-0785, Nazar Khomyn 0009-0008-1906-3024

Resume

The problem of using big data for optimal planning of work and rest of truck drivers in intercity freight transportation is considered. The problem is solved for a team of drivers in compliance with the EU regulation 561/2006. The proposed data preparation can reduce the practical computational complexity and find the optimal driver work schedule. The clustering of the initial data was substantiated for optimizing schedules based on the criterion of compatibility of transport orders. The compatibility of orders follows from the analysis of similarity of two consecutive transportation processes performed by one driver during his/her work shift. Simulation modelling was performed. The influence of the scale of the territory on the quality of the prepared initial data was also investigated. A rational range of the compatibility coefficient for the high-quality preparation of a large data set was substantiated.

Article info

Received 12 January 2026
Accepted 31 March 2026
Online 15 April 2026

Keywords:

road transportation
driver scheduling
big data
clustering
order compatibility

Available online: https://doi.org/10.26552/com.C.2026.022

ISSN 1335-4205 (print version)
ISSN 2585-7878 (online version)

1 Introduction

Due to the alarming trend of a shortage of commercial vehicle drivers worldwide, studies of the efficiency of transport systems are increasingly conducted based on the criteria of increasing the labor productivity and reducing the number of personnel required as a result [1]. One of the most effective ways to address this problem is currently the use of improved logistics in the form of more detailed and accurate driver schedules [2]. It is also recognized that promising communication technologies and vehicle autonomy will help to solve these difficulties. However, the problem is also complicated by the fact that the duration of work, driving and rest of drivers is regulated by the European Agreement 561/2006 with the motivation of occupational safety and health [3]. It is known to be a very common phenomenon that majority of road freight carriers assign drivers to trucks to increase their responsibility for the technical condition of the equipment. This is also accompanied by the problem of downtime and inefficient use of vehicles. However, both the use of a shift method of the driver team work and the use of modern communications are insufficient for the full use of the driver's working time fund. This requires drivers to pick up and drop off trucks to their partners when their maximum allowed shift time begins and ends. Such a process has already been modeled previously using shift method operation of a carrier. A reduction of the required number of driver teams with a constant volume of transportation by 32% was achieved [4]. However, the carrier had a rather small fleet, so this achievement is insignificant in relation to the real need for hiring drivers. The volume of vacancies for truck drivers currently reaches 30% in the European Union. Therefore, the shift method and other logistics improvements do not justify real expectations on a small scale.

It is necessary to draw up a tight schedule for a sufficiently large team of drivers to ensure their coordinated work and freight transportation with minimal time losses (idle runs and downtime). Cloud computing has become the most effective among known scheduling algorithms, as it can solve multiple problems simultaneously and use big data [5]. However, despite a fairly large number of known studies, heuristic and metaheuristic methods have been successfully used to solve multi-objective scheduling optimization problems. Traditionally, heuristic and meta-heuristic solution methods have been successfully used for such problems to some extent. In recent years, the emphasis has shifted to hybrid methods that offer more efficient solutions. Choosing an effective optimization algorithm is also important in modern information systems of transport enterprises. Processing large amounts of data and making decisions in real time becomes a routine process. The results of the comparative analysis showed that none of the known algorithms is universal for all types of tasks. The choice of the optimal algorithm depends on the specific task, the requirements for the speed and accuracy of decision-making, the availability of resources, etc. The use of hybrid approaches that combine the advantages of several algorithms can provide the best results in many cases.

Thus, the conditions for application of promising cloud computing and hybrid algorithms are the availability of big data, which, on the other hand, creates difficulties in finding guaranteed optimal solutions in an acceptable time. Modern methods of schedule optimization work effectively under the condition of structuring the initial data flows, their preliminary analysis, filtering and segmentation.

The main efforts of these studies were aimed to prepare initial data on freight transport parameters that relate to a set of interested freight carriers connected to the planning network. The purpose of the research was to develop such a criterion for structuring the data flow, according to which a large amount of data would enable the use of cloud computing, on one hand, and will provide an opportunity to successfully apply meta-heuristic methods for solving various problems of transport process planning, primarily a scheduling for drivers of various carriers connected to the cloud computing network. To achieve this goal, the following research tasks were set: (a) to substantiate the criterion for processing initial data such that it ensures compliance with the conditions and limitations of the work of several driver teams and their interaction and makes it possible to process large volumes of initial data; (b) to investigate the influence of the scale of input data on their adaptability to the use of heuristic scheduling optimization algorithms.

2 Literature review

The number of planning algorithms related to driver and vehicle scheduling research is quite large, increasing annually by an average of 16.62% [6]. At the same time, the problems of optimal scheduling are spreading to areas of activity where they were not previously applied [7]. Now, schedule optimization is also relevant for the road freight transportation, multimodal transportation, terminal technologies, etc. The main goal of such research is the most rational distribution of available transport resources. Most of the published works are concerned with the urban passenger transportation and the organization of work of bus drivers [1] and taxis [8]. However, research results on improving freight logistics are increasingly appearing in modern publications [9].

Improving the logistics of personnel work is seriously complicated by the diversity of potential that individuals with different qualifications and performing different functions are endowed with [10]. This is also illustrated by the example of air transport crew planning [11]. A similar problem can also be identified in freight road transportation. However, solving it separately for each category of employees is impractical, since that disrupts the coordination of individual transport groups and destroys the integrity of the transport system as a whole [1].

Since the complexity of complex scheduling tasks is significantly increasing, some researchers have proposed to narrow the scale of planning from tactical to operational. There has been a trend in formulating complex road transportation planning tasks. In particular the tasks of drivers team scheduling assigned to several depots [12], the tasks of planning individual parts of the transport network, with connected cooperative vehicles, were combined into one were formulated and solved [13]. There has been a trend in formulating complex road transportation planning tasks of improving the work schedules of drivers and vehicles are now also solved comprehensively [14].

Formulating the complex search conditions yields the better-quality solutions. However, the complexity of the initial conditions increases the scale of the required initial data. Therefore, the algorithms used to process such a volume of data must be qualitatively different from traditional ones. Such algorithms must be equipped with modern information technologies, artificial intelligence. By analyzing large datasets, historical trends, and real-time data, AI-based algorithms can determine the most efficient routes and schedules for transportation and operators for a variety of transportation and logistics needs [15].

Big data operations have the well-known benefits, but their application presents certain challenges for implementation. The advantages include: (I) the possibility of using AI algorithms and deeper learning; (II) large volumes of data are low-cost; previously, collecting such data required human effort, which was expensive, time-consuming, or impossible; (III) large volumes of data are characterized by greater reliability and accuracy of the information generated, since a large number of sources allows for the verification of data streams; (IV) large data flows allow for predictive analytics in transportation, in particular, estimating arrival times or warning of potential incidents [16].

Challenges: (I) data flows coming from different sources often have a mutually inconsistent structure, requiring complex pre-processing to convert the data into formats that facilitate analysis; (II) the complexity of distribution tasks is increasing, namely the assignment of vehicles to known tasks, the distribution of tasks to drivers, taking into account work and rest time regulations, as well as chronic staff shortages; (III) a higher level of synchronization of data streams is required due to time delays; (IV) problems with data collection and storage; (V) problems with data standards agreement [17].

In addition, in the work [18] is stated that there are significant gaps in the knowledge of researchers and practitioners regarding the right information and tools for analyzing large databases. If a large amount of data is generated by different sources, processed by different applications, and used by different means, a programmable framework for distributed computing that uses a "divide and conquer" approach to processing big data is needed. Future challenges include developing batch big data processing tools for distributed data indexing, replication, load balancing, automated disaster recovery, and data restoration [19].

A key area of development for big data processing is artificial intelligence, which fundamentally changes the way companies approach complex routing problems and solve them in a reliable and simple way [20].

To solve the Nondeterministic Polynomial (*NP*) hard schedule planning problem with *NP*-hard characteristics, a large number of heuristic algorithms have been developed: density-based clustering [9], cloud computing [21], recursion [1], etc. However, the above methods have certain limitations in the process of solving transport planning problems, since they require more solution time and at the same time significantly increase the computational complexity of the planning system [22].

A very important task in organizing the work schedule of truck drivers is the task of predicting precisely the moments: the arrival of trucks for loading, delivery of cargo, etc. These and other moments are critical for the schedule as a whole. However, with large volumes of data and the random nature of processes, traditional heuristic methods do not provide accurate solutions in a reasonable time [23].

The crucial importance in building efficient supply chains based on clustering is ensuring continuous freight flows through the full integration of the involved participants and logistics infrastructure facilities in the network, based on clustering and a single partner space. Analysis of the methodology and the results obtained from the application of the more common K-means clustering algorithm in the context of container terminals provide the grounds for applying a similar approach to cargo transportation. Specific data regarding container transportation in multimodal systems is a reduction in the total number of required transport trips by 31.58% and provides decision makers with a compromise between the total truck turnover time and the deviation from the desired time window for cargo pickup by transport companies [24]. It is obvious that bringing the content of the clustering criterion closer to the essence of the objects will improve the quality and responsiveness of the obtained solutions.

3 Problem formulation

Next a finite set of logistics processes $J = \{J_1, \dots , J_x, \dots , J_n\}$ is considered. Here, a process means a sequence of logistics operations for fulfilling one order for goods transportation by road vehicles on an intercity transport network. Orders are independent, so they can be fulfilled in any sequence. However, each order may have a time window to limit the maximum duration of its fulfilment. The processes of fulfilling all J orders represent the project horizon H, within which it is necessary to fulfil the maximum number of orders from the set J with the minimum number of drivers. So, k identical trucks are used for this aim. The average duration of the trucks runs between any points of the network, the duration of loading /unloading of trucks are the values predicted. The trucks are driven by a team of drivers who are able to replace each other then and at those points of the transport network where it will be necessary to do so due to the restrictions 561/2006. However, the objects of study in this work are intercity processes with a horizon of approximately 2-3 days, so weekly/fortnightly restrictions according to Regulation 561/2006 were not considered for this reason.

Thus, the drivers must be coordinated, and this is possible if a clear schedule of their work is given. The schedule of the team of drivers is considered to be unambiguously given if for each driver there is a specified start time and end time for the execution of the order / sequence of orders assigned to him. Order fulfilment operations may be interrupted by drivers due to time constraints. If the shift driver work method is used, the schedule must also indicate the truck number that the driver will drive when fulfilling the order. It also indicates the points in the transport network where the driver must accept the truck, and deliver the goods, and hand over the truck to his partner. Unlike the known methodology [24] an unambiguous schedule consists of sets of values:

- $t_1^b, t_2^b, \dots, t_i^b, \dots, t_n^b$, where t_i^b is the earliest time to start an order J_i execution;
- $t_1^e, t_2^e, \dots, t_i^e, \dots, t_n^e$, where t_i^e is the latest time to finish an order J_i completion.

In the known method, the schedule of the entire process is considered to be uniquely determined if the exact start or end time of all operations is known. However, the schedule of the process at deterministic moments is difficult to reconcile in practice with the random factors that are always present in the logistics of cargo delivery. Therefore, one of the differences of this algorithm is the use of time points with tolerances.

The quality of the decomposition is characterized by the minimum value of the non-decreasing function in each argument:

$$H(x_1, x_2, \ldots x_n), x_i = t_i^e - t_i^b \tag{1}$$

The schedule is considered as optimal if it corresponds to the smallest value of the specified Equation (1), while observing the condition of full implementation of the project with the involvement of a given number of drivers, as well as compliance with the regulations of their work and rest.

4 Methodology

A previously developed method for optimizing the work schedule of a team of truck drivers working in a limited area is used, and applying a shift work method, using the criterion of the minimum total duration of the project. The method is based on the ordering of mixed (disjunctive) graphs $G(J,V,U)$, where J is a set of vertices $J = \{j_1, \ldots j_n\}$, V, U are sets, respectively, of edges and arcs, each is given with a weight $a_{i,j}$, $i,j = 1 \ldots n$ [24]. Such graphs may contain contours of positive weight, so it is impossible to construct a unique schedule after them. To overcome this obstacle, it is necessary to remove all edges V, or replace them with arcs. The minimum guaranteed duration of the entire project depends on operation done with each edge $v_{i,j}$ (removed or replaced with an arc $u_{i,j}$ or $u_{j,i}$). This value is the length in time scale of the critical chain in a graph without loops and contours. However, even in a graph without edges there may be contours, or isolated vertices, which make it impossible to find the critical path from the fictitious vertex S, which symbolizes the formal beginning, to the vertex F, which symbolizes the formal end of the project. A well-known graph ordering algorithm was improved in [24].

If there was a contour of positive weight in the graph $G(J,U)$, the search for the conflicting arc $u_{k,i,j}$ would be performed. Removing a conflicting arc from a graph G will not only make the cycle-free graph, but also improve the constructed schedule. In fact, there may be a finite set of such conflicting arcs, so the task of finding and removing them is a multivariate problem; the complexity of the problem is polynomial [4]. An empirical algorithm was applied, in which an indirect optimality criterion for the graph $G(J,U)$, which does not contain circuits, is used. A feature of the improved method is that it gives a satisfactory estimate of optimality $\inf(H_{\min})$, where $H_{\min}$ is the minimum project duration with the dimension of the problem $J \leq 40$ vertices:

$$H_{\min} = \min\{h_{i,j}, h_{j,i}\}, \tag{2}$$

where:

$$h_{i,j} = t_i^e(J,U) + x_i + \nu_y(J,U) - \nu(J,U), \tag{3}$$

v_j means the maximum value of the chain in graph G, which starts at the vertex j_y; $\nu(G, U)$ is the maximum G graph chain. That simplified the task of finding an active optimal order execution schedule after graph $G(J,U,V)$, using criterion in Equation (2), by choosing the most conflicting edge. The algorithm that was improved allows for one to find an approximate solution faster than previously possible, taking into account that the desired schedule consists of approximate operation execution times. The complexity of the algorithm does not exceed $O(2^J+V)$ elementary operations. However, the algorithm still works slowly with large amounts of initial data. In addition, the error in estimating the quality of the schedule also increases with increasing problem dimensionality. When the number of vertices increases by 10, the accuracy of the optimum estimation decreases by more than 7% and this error increases exponentially with the increase in the number of vertices J [14]. When the number of vertices arises up to 90, the solution of the problem in the permissible time becomes unattainable. Thus, the algorithm based on the ordering of the mixed graph is effective with a small volume of vertices, as well as conflicting edges or arcs. On the other hand, if the initial graph G has no edges or contours, then it uniquely determines the work schedule of each driver. Such a result is trivial and does not contain the subject of optimization. These facts set limits on the size of the initial data for the problem of constructing the schedule of a driver team when using a heuristic algorithm on conjunctive graphs. An additional criterion is required to prepare the initial data, which will lead to an increase in the efficiency of the schedule.

5 Criterion for assessing the closeness of organizational ties

The duration of driving, work and rest periods for drivers is limited by EU Regulation 561/2006 with all its amendments, successively introduced into the regulations [3]. A driver's shift consists of driving time, the continuous duration of which cannot exceed 4.5 hours. The total driving time during a shift cannot exceed 9 hours, or as an exception 10 hours, but only twice a week. A driver must have a rest period during a shift for at least 45 minutes. A driver must have a rest period between shifts for at least 11 hours. It is allowed to divide the duration of rest period, and rest period between shifts. Driver's duties also include loading and unloading operations, preparatory operations (fastening, checking the integrity of the load), guarding the truck at parking lots, technical inspection and maintenance of the vehicle. All these works, together with driving, are included in the 13-hour work cycle (Figure 1).

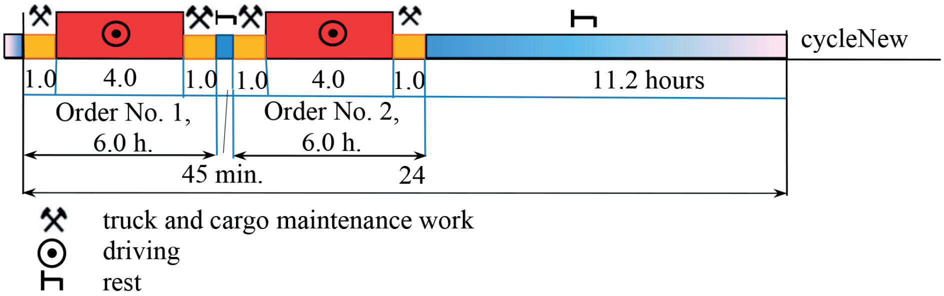

Figure 1 *Idealized scheme of driver work and rest modes when performing two consecutive orders #1, 2 within one shift*

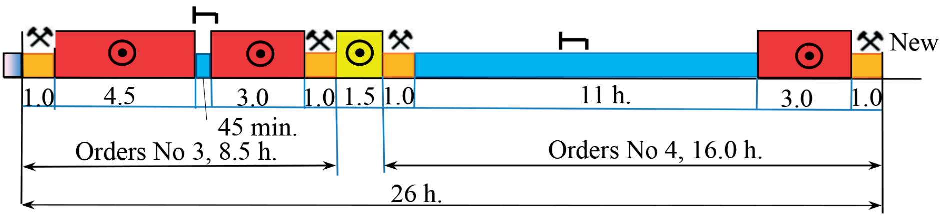

Figure 2 *Displaying sample of partially compatible orders*

Figure 1 shows the execution of some orders #1, 2, the driving duration of each is 4 hours, and the total duration of each order is 6 hours. Orders represent the transportation of goods between pairs of transport points. In this case, the final point of order #1 is the initial point of order #2. Therefore, the sequence of execution of orders by one driver on one truck will result in the absence of idle runs and useless downtime in such a transport cycle. The required variable rest of the driver is a standard value of 45 minutes. The duration of continuous driving, the total duration of driving, and the duration of the work shift do not differ from the permitted standard. Therefore, such a scheme can be considered as idealized. The scheme shows the zones: cargo work in orange bars; truck driving in red bars, driver rest in blue bars. If there is a need for idling, this is indicated by yellow bars.

The schedule of logistics operations can be implemented using an idealized scheme, which is shown in Figure 1 due to the fact that orders #1, 2 are fully compatible in terms of organizational and technological parameters. Such compatibility can be estimated as 100%, or using some compatibility coefficient $R_c = 1.0$.

Figure 2 shows two orders #3, 4 that do not have similar compatibility.

Since the truck's driving time during order #3 exceeds 4.5 hours, it is interrupted for 45 minutes for the driver's necessary rest. Order #4 starts at a different time and at a different transport point, when order #3 ends, so there is a need for idling. The driver needs a long inter-shift rest of 11 hours after loading the cargo, when executing order #4. Ultimately, order #4 is executed with a long delay of 16 hours, although the "net duration" of its execution is only 5 hours. Orders that vary in duration of characteristic periods and differ from idealized compatibility are here called partially compatible.

Thus, for the fully compatible orders that are sequentially executed by one driver throughout the day, the duration of driving a loaded truck must be at least 8 hours, the duration of auxiliary operations must not exceed 4 hours, the driver's rest period must not be less than 45 minutes, and idling is excluded. Such a shift must alternate with an 11-hour inter-shift rest. The considered orders #3, 4, which are performed in the sequence 3→4, are called partially compatible. The time for performing these orders takes more than a day. At the same time, the duration of a run with a cargo is 10.5 hours, which is 1.6 times less than the two-day norm. An idle run is required to perform the orders. Inter-shift rest interrupts the execution of order #4. The degree of their compatibility is determined by the coefficient, which has the meaning of comparing the real compatibility to the ideal one:

$$R_c = \begin{cases} \dfrac{cargo}{cargo + idle + rest + loading}, \\ if\ (cargo + idle) \geq 4.5 \\ \dfrac{4.5 - (cargo + idle)}{4.5}, \\ if\ (cargo + idle) \leq 4.5 \end{cases}, \tag{4}$$

where: *cargo* is total duration of loaded runs; *idle* is the total idle time; *rest* is actual duration of the driver's rest; *loading* is duration of cargo operations during the cycle; all values in Equation (4) are measured over two consecutive transport cycles, hours.

The assessment of the two tasks compatibility depends on the comparison base. If the driving duration of the cycle is less than the allowed without rest duration (4.5 hours) of the driving, then the comparison base is only the time spent on driving (with cargo and without cargo), and idle times and delays are not taken into account. If the driving cycle is longer than the duration of the maximum allowed driver's trip without a break, then the comparison base is the maximum duration of a shift (12 hours), or several shifts of the driver, including cargo operations, truck maintenance and minimum shift' rest. Thus, the compatibility coefficient of orders #3, 4 for the example shown in Figure 2, R_c = 10.5/26 = 0.4.

The reverse order execution sequence is generally independent of the forward order execution sequence. The coefficients R_c may differ for the forward and reverse order sequences.

In this paper is also considered the influence of the time windows execution of each order on their compatibility. Here, the time window is understood as the time tolerance ($t_i^{e.b}$, $t_i^{l.e}$), where $t_i^{e.b}$, $t_i^{l.e}$ are the most probable early beginning, and late ending of the i-th order fulfilment. If, for a sequence of two orders, $i \rightarrow j$, inequalities are satisfied

$$t_i^b \, t_j^{e.b} < t_i^{l.e} \text{ and } t_j^{l.e} \leq t_i^{e.b}, \tag{5}$$

then such orders are considered completely incompatible in the specified sequence, for them the coefficient R_c=0, and the specified sequence is excluded from consideration.

If for the same sequence $i \rightarrow j$ one has:

$$t_i^{e.b} \leq t_j^{e.b} \leq t_i^{l.e} \text{ and } t_j^{l.e} \geq t_i^{l.e}, \tag{6}$$

then such orders are considered partially compatible. The compatibility coefficient for these orders by time windows is determined by:

$$R_c = \frac{1}{t_j^{e.b} - t_i^{l.e}}, t_i^{l.e} > t_j^{e.b}, \tag{7}$$

where the order execution time frames are measured in integer units of time.

If the compatibility coefficients R_c for the same pair of orders calculated by Equations (4) and (7) are different, then a smaller value is chosen for further clustering operations.

To form clusters of similar orders, it is necessary to combine orders based on organizational compatibility, that is, by the coefficient R_c. However, the organizational and technological connection between a pair of orders can be of different quality, since 0 ≤ R_c < 1.0. Therefore, the level of quality of the connection requires justification, that is, it is necessary to establish at which value of R_c the connection between a pair of orders should be considered close enough to include them in same a group. For this purpose, simulation was planned and conducted under different conditions of execution of transport orders, and a study was conducted of the influence of the adopted value of R_c on the quality of data preparation for building an optimal schedule.

6 Simulations

To adhere to the principle of generality (universality) of the scientific problem and avoid the subjectivity of any particular case, it was used the modelling of the processes of formation and fulfillment of freight orders based on random input data. Such cases significantly depend on commercial, legal, geographical and even political conditions. The specifics of the practical activities of carriers can serve as examples of the practical significance of the developed methodology rather than as a means for its development.

The purpose of the simulation was to generate random data on the availability, location, and time parameters of orders with a large amount of data in order to perform their clustering and prepare for construction of several independent optimal schedules in terms of the number of drivers involved.

Orders are received by the carrier as a need for transportation of consolidated cargo in an intercity route. The simulation generates random values of the geographical coordinates of the points that relate to all orders known during the modelling period, as well as other points that are not included in any route. Random coordinates of transport points in the territory under consideration were determined using the program generator of random numbers (PGRN) with uniform distribution. The algorithm and the initial number of the generator were written by the authors in the Delphi programming language. The coordinates of q_i points of the territory are received:

$$x_i := random(0; X_{\max}), \, y_i := random(0; Y_{\max}), \tag{8}$$

where X_{max}, Y_{max} are the maximum possible relative coordinates in the serviced area, km.

The starting and the ending points of each cargo delivery route were also randomly assigned.

So, the indices of the points of departure of the goods: i := $random(1; N_l)$; indices of delivery points: j := $random(1; N_l)$, and any departure points cannot be a delivery points simultaneously for the same cargo.

The assumption was made that on the transport network, where the freight transportation is performed, there is a road connection between any pair of points. To determine the distance between any transport points g_i, g_j, the following formula was used:

$$l_{i,j} = k_c \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \text{ km}, \tag{9}$$

where: k_c is a coefficient of curvature of the path.

The path length calculated by Equation (9) approximately reflects the distance between the two points that would lie between real points.

The road curvature coefficient takes into account two factors that distinguish a segment from a real road: (a) the curvature of the road in plan; (b) the absence of a direct connection between points q_i, q_j, but through

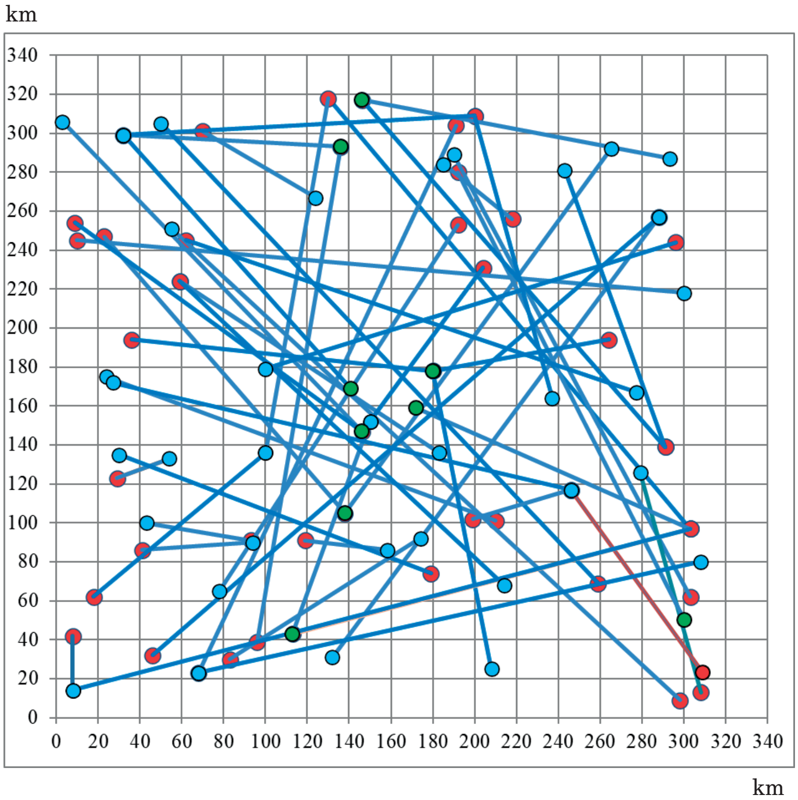

Figure 3 *An example of generating a network of random orders: red markers are departure points; blue markers are reception points; green markers are points that are both senders and receivers of goods; segments are vectors of goods movement without taking into account the curvature coefficient; population density is 1.35 settlements per km²; number of orders - 50; area of the territory - 115.6 km²*

a transit point q_t. Given the random nature of the simulated transport network model, this method of determining the length of the paths can be considered quite satisfactory. After all, not only the coordinates of the points are random, but the length of the paths between them and the network configuration, as well.

To specify the configuration of the transport network, the population density indicator was used. Actually, the population density of the territory is an indicator that characterizes the location of settlements per unit area of the geographical territory where the transport network is located. The population density is determined by:

$$\rho = \frac{N_l}{A}\text{, settlements / 100 km}^2, \tag{10}$$

where: N_l is the number of settlements located in the territory under consideration;

A is the area of the territory under consideration.

Statistical sources provide numerical values of this indicator for different regions of Europe. The population density can reach a value of 9 settlements/100 km² for some regions. The minimum density value was noted as 1.8 settlements/100 km². However, to set and implement this indicator in the simulation model of this work, a different approach was applied. For simulation modelling, it is necessary to set such a random location of transport points, which is characterized by the desired density coefficient. It is necessary to divide the territory with a given area into equal squares that can simultaneously accommodate several settlements. In this work, such settlements were taken into account that are considered cities, that is, have a population of more than 3500 people, and have the appropriate infrastructure for performing large-scale freight transportation between cities. Cities in East Europe, with suburban areas and satellite cities, occupy an area with a fairly wide range: from 20 to 900 km². Areas of more than 700 km² are occupied by megacities, which are administrative centres of regions. There are no publicly available statistics that would provide an accurate value of the average area for all cities. However, if one considers the regional structure (region), the average area of a city that is not an administrative centre ranges from 80 km² to 3.7 km². Therefore, the size of the square of the territory to be considered should be no less than 100 km². In this case, one square cell can contain no less than 3 settlements due to random distribution.

The following initial data were used for simulation modelling of the conditions for performing intercity transportation: (A) the total number of settlements P; (B) maximum size of the territory in latitude X_{max} and longitude Y_{max} km; (C) order execution period T, hours; (D) the set of orders $Z=\{z_1, \dots z_n \dots, z_m\}$, time windows of which are included in the execution period; (E) order time windows $t_i^{e.b}$, $t_i^{l.e.i}$; (F) population density coefficient € (G) road curvature coefficient k_c; (H) average technical speed of trucks on the network V_t.

The random number generator of the simulation model produces the following data: (I) coordinates of each settlement x_i, y_i; (J) number of the initial and final route point *i, j*, which concerns each known order; (K) duration of movement with cargo for each order cargo; (L) duration of downtime associated with cargo operations, and truck maintenance, loading. An example of random orders (coordinates of the points of departure and receipt of cargo are given in Figure 3.

All orders are compared in pairs after their formation. The ratios were obtained, which reflect the consequences (time spent on driving the truck, driver rest, idling) of the sequential execution of orders by the same driver, on the same truck, in strict order. The results of consecutive execution of two orders #*i, j* were evaluated according to the following rules, taking into account Equation (4), as well as Regulations 561/2006 EU.

1. First, build the simplest cargo transportation cycle to fulfill the first order #*i*. The cycle must fully comply with EU Regulation 561/2006 and contain all the necessary elements for the implementation of the process (loading, movement with cargo, unloading, driver rest, idling, etc.).
2. Build a complex transportation cycle to fulfill two consecutive orders #z_i, #z_j by the same vehicle, and by one driver, or by team of drivers, who can work using a shift work method. It is necessary to take into account the possible need for idling and vehicle downtime, while waiting for the start of loading / unloading. There may be a need for drivers to rest between shifts, as well.
3. Using Equation (4), determine the compatibility coefficient of two consecutive orders that will be performed only in a given sequence #$z_i \rightarrow$ #z_j.
4. Determine the time tolerances for the execution of logistics operations, and the compatibility coefficient of orders using Equation (7). The smaller numerical value of the compatibility coefficient $R_{c.i.j}$, which is calculated by Equation (4), or (7), is taken as the correct value for further analysis of the entire data set on existing orders.

Based on the random data, a square matrix of values $R_{c.i.j}$ of order compatibility coefficients is constructed. Matrix ($R_{c.i.j}$) gives the ability to assess the closeness of the connections between each pair of transport cycles, and to select the most significant ones for clustering a large amount of initial data. A passing value of the compatibility coefficient is indicated for a known compatibility matrix, i.e., one below which the two orders cannot be performed sequentially by one driver. This threshold value is subject to practical justification. To determine the most favourable conditions for combining orders into clusters, simulation modelling was carried out with variable input data. The variables were such values:

1) the maximum area of the service area (102.4, 78.4, 57.6, 32.4, 19.6 thousand km^2);
2) the number of settlements in a given area (200, 140, 120, 100, 50);
3) the number of orders on the planning horizon (150, 100, 50);
4) the curvature coefficient of roads in the plan (1.35, 1.75, 2.5).

Each combination of data gives the resulting set of graphs with a different number of connections and isolated vertices. If the coefficient R_c has a too low value, i.e., $R_c \rightarrow 0$, then many insignificant connections in the graph *G* will not be discarded. The graph will be strongly connected. Clustering will not actually occur, and there will be a large number of edges *V* between the vertices *J* of the graph *G*. It means that the resulting graph is not very suitable for finding a fast and guaranteed optimal solution of the driver work and rest schedule. If the numerical value of the coefficient R_c is too large, i.e., R_c€1.0, then the vast majority of connections in the graph *G* will not be preserved. This applies to both the arcs *U* and the edges *V*. Additionally, there would be a large number of isolated vertices among the initial data. These vertices are the orders that cannot be combined with any of the other orders. This means that there will be no division into clusters in such a model, since most of the vertices will be isolated. Finding the optimal schedule in such initial data will be a trivial task, since the schedule will be a set of start times for almost all orders: $t_i^s = 0$, for $I = 1 \ldots Z$. In this regard, the desired value of R_c was justified after analysing all the previously obtained data, which were clustered for all possible $0 \leq R_c \leq 1$. If the number of clusters is greater than 0, and if the number of contours in the graphs of the initial data is greater than 1, but not more than the number of arcs, then the coefficient R_c was considered significant.

Figure 4 shows the result of dividing the initial data into 5 subgraphs $G_1 \ldots G_5$ from the general graph *G*, which were obtained for $R_c = 0.3$ and 12 isolated vertices. Subgraphs have edges, respectively: V_1 - 3; V_2 - 2; V_3 - 1; V_4, V_5 - 0 items.

The simulation results for other modified conditions are shown Table 1.

If the existing edges are removed from the subgraph G_1, or converted into arcs, then this subgraph would still contain 3 contours, which must be removed by analyzing all the available arcs for criticality [14]. Thus, the subgraph G_1 is one of the most complexes of all the clusters, but the task of finding the optimal location on it can give a qualitative result due to a larger field of variables. Subgraph G_2 is less complex, since its variable field contains only two edges and no other contours. Subgraph G_3 is even less complex, but it leads to only two possible solutions. Finally, subgraphs G_4 and G_5 give a unique orders fulfilment schedules with an admissible compatibility coefficient $R_c = 0.3$.

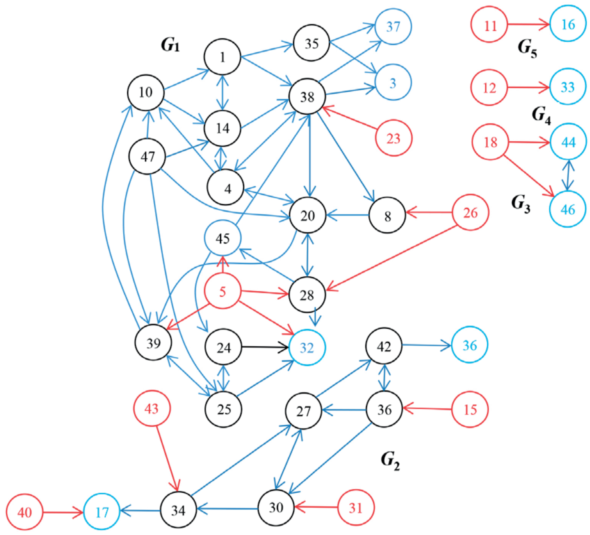

Figure 4 *An example of mixed graph with parameters: threshold value of the order compatibility coefficient $R_c \geq 0.3$; number of vertices (orders) J = 50; number of vertices connected by links: 38; number of isolated vertices: 12; number of settlements P = 120; area of the territory A = 19.6 thousand km²*

Table 1 *Simulation results*

Version	Number of orders	Area of the territory, thousand km²	Number of settlements	Settlement coefficient	Road curvature coefficient	Number of orders without a group	Compatibility coefficient threshold	Number of groups	Number of contours
1	50	19.60	50	1.42	1.35	21	0.50	1	5
2	50	19.60	100	1.42	1.35	18	0.40	2	8
3	50	19.60	120	1.42	1.35	12	0.30	5	9
4	50	32.40	50	1.42	1.35	28	0.60	1	10
5	50	32.40	100	1.42	1.35	20	0.50	1	18
6	50	32.40	120	1.42	1.35	16	0.40	2	21
7	50	32.40	140	1.42	1.35	3	0.30	1	32
8	50	57.60	50	1.42	1.35	18	0.60	1	28
9	50	57.60	100	1.42	1.35	8	0.50	4	12
10	50	57.60	120	1.75	1.35	3	0.40	5	6
11	50	57.60	140	1.75	1.35	0	0.30	3	4
12	50	102.40	50	1.75	1.35	4	0.60	2	14
13	50	102.40	100	1.75	1.35	2	0.50	6	6
14	50	102.40	120	1.75	1.35	2	0.40	4	18
15	100	19.60	100	1.75	1.35	24	0.60	2	8
16	100	19.60	120	1.75	1.35	22	0.50	4	21
17	100	19.60	140	1.75	1.35	18	0.40	3	12
18	100	32.40	100	2.15	1.35	20	0.60	3	16
19	100	32.40	120	2.15	1.75	20	0.50	2	9
20	100	32.40	140	2.15	1.75	16	0.40	2	9
21	100	57.60	100	2.15	1.75	18	0.60	5	16
22	100	57.60	120	2.15	1.75	18	0.50	6	27
23	100	57.60	140	2.15	1.75	16	0.40	1	34
24	100	102.40	120	2.15	1.75	18	0.60	2	22
25	100	102.40	140	2.15	1.75	16	0.50	2	35
26	100	102.40	200	2.15	1.75	12	0.40	3	21
27	150	32.40	120	2.15	1.75	28	0.60	3	26
28	150	32.40	140	2.15	2.50	24	0.50	7	19
29	150	32.40	200	2.15	2.50	24	0.40	2	11
30	150	57.60	120	2.50	2.50	24	0.60	2	16
31	150	57.60	140	2.50	2.50	24	0.50	3	22
32	150	57.60	200	2.50	2.50	20	0.40	4	10
33	150	102.40	120	2.50	2.50	32	0.60	2	12
34	150	102.40	140	2.50	2.50	28	0.50	4	25
35	150	102.40	200	2.50	2.50	14	0.40	5	10

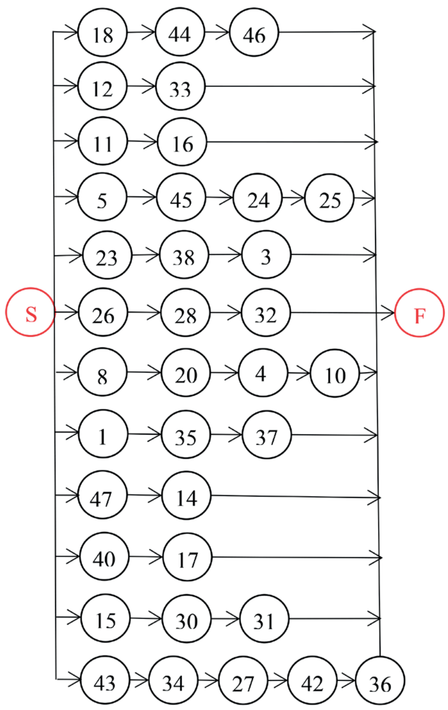

***Figure 5** Example of an ordered graph*

```
Predicted time - 20.0 hours
Transport task of driver #1: order 18-44-
46; by truck #1
Truck #1 route: points  2-4-3-1
Order #18 start 0.0 end 3.5, truck #1 from
point 2 to point 4, driver #1
Order #44 start 4.0 end 5.5, truck #1 from
point 4 to point 3 driver #1 Order #46
start 6.2 end 9.5, truck #1 from point 3
to point 1
Minimum duration of this project = 11.0
hrs. Team of drivers - 12 persons
```

***Figure 6** Fragment of optimal schedule*

***Table 2** Results of schedules optimization of truck drives based on big data*

Version	Number of drivers	Project total duration, h	Version	Number of drivers	Project total duration, h
18	28	7.3	27	52	18.6
19	24	7.1	28	51	14.2
20	26	7.8	29	69	26.5
21	33	10.6	30	82	24.2
22	27	8.4	31	75	19.3
23	44	13.8	32	79	22.8
24	67	14.5	33	115	28.4
25	72	13.5	34	92	26.0
26	71	12.8	35	102	22.8

Only the data obtained for $R_c \leq 0.6$ are shown, since for larger values of the coefficient no cluster was obtained (all vertices were isolated). If $R_c \leq 0.3$, then the clustering also does not occur, since only one graph with a significant number of edges, almost twice the number of arcs was obtained. The data obtained as a result of modelling also make it possible to assess the impact of the scale of freight transportation planning and driver schedules on the quality and availability of optimal plans. Figure 5 shows an example of an ordered graph $B(Q,U)$, which is constructed from subgraphs G_1 to G_5. Graph B does not contain contours of positive weight. Additionally, a fictitious vertex S is introduced into the graph, which symbolizes the formal start of the whole project. From the vertex S, only arcs exit to the vertices, which are shown in Figure 4 by red circles, which are the initial orders of the project. The fictitious vertex F represents the fictitious completion of the project. There are only incoming arcs to the vertex from the vertices, which are shown in Figure 4 by blue circles.

When applying the schedule construction method and the corresponding algorithm and program [25] to the formed graph G, a work schedule for 12 drivers was obtained the content of which is recorded in Figure 5.

A graph without loops and contours, an example of which is shown in Figure 5, makes it possible to establish a unique minimum work schedule for truck drivers and an order fulfilment schedule taking into account the specified maximum predicted time (Figure 6).

The modified algorithm was applied to all randomly generated initial data. The schedule optimization algorithm was applied to each order cluster separately. After that, the individual partial optimal schedules were combined into one horizon and a set of resulting schedules was obtained for each version of the initial data. Table 2 presents the parameters of those schedules for which the clustering gives a significant effect in optimizing the order execution sequence.

Thus, the first 17 versions for the initial data presented in Table 1 do not lead to a significant effect of data grouping. They are excluded from the following analysis. Other versions 18-35 allow to identify the following consequences of clustering. An increase in the average duration of one order leads to an increase in the required number of drivers, regardless of the method of preliminary data preparation. The greatest effect of data preparation is achieved with large-scale planning. Thus, with the number of orders of 150 and the average duration of one order of 3.5 hours, it is possible to achieve a reduction in the duration of the project by 5.6 hours, and reduce the number of drivers by 23 persons. Thus, it can be argued that the most effective way to prepare the initial data is to maintain a large number of positive weight contours with a sufficiently large number of clusters. Such an effect is observed over a large service area if the compatibility coefficient is kept at 0.5.

Thus, with an increase in the transportation area without an increase in the number of planned orders, the pairwise organizational compatibility of orders weakens, which is associated with an increase in travel distances and, accordingly, the time spent by drivers on driving a truck both with cargo and idle. Therefore, the expansion of the service area should be accompanied by an increase in the density of orders. At the same time, the number of settlements does not significantly affect the possibility of obtaining a high-quality schedule.

The number and total length of roads in Europe do not increase proportionally, practically, with the increase in the size of the service area. After all, the population density coefficient, especially in the western regions, is quite high (reaches 8.8 points per 100 km^2), while the length of local roads, which accounts for 1 km of highway, is small. Therefore, the expansion of the service area is not a leading factor in the growth of truck fleet productivity.

The developed algorithm for assessing the compatibility of a set of the available orders was used for the computer program in the Delphi language. The program was tested at several international enterprises which are the large road carriers in Eastern Europe. One of such large enterprises is the Trans-service-1 Company [26]. The enterprise uses more than 150 road trains with tilting semi-trailers which are suitable for transporting the necessary types of cargo. Most carriers such as Transservice-1 have a network of their regular customers, and are also used by random orders from intermediaries. However, the Transservice-1 enterprise, as well as others like it, do not choose specialized software or any other methodological tools to distribute orders to drivers and to assign vehicles to known transportation tasks, and do not contain schedules for their implementation. That is why the planning of drivers' work schedules was carried out at these enterprises first.

The logistics department of the carrier receives an average of 200 orders from intermediaries and about 70 orders from the regular customers, which geographically extend along almost all highways throughout Europe. The company rejects some received orders because the company's trucks are spread out quite haphazardly, and otherwise would have to suffer a lot of downtime and downtime and idling. Currently, the company experiences a major shortage of drivers and a significant impact on the cost of transportation of high fuel prices. The carrier fulfills approximately 30-50 orders, primarily these are orders from regular customers, other orders are rejected. The 5 typical daily transportation plans of the company during the periods of the highest productivity of the fleet are reviewed and compared with the existing demand in the same periods, as well as with the actual capabilities of the fleet. Software are developed for more thorough planning of work schedules for road trains operated by teams consisting of 1-2 drivers, and the possibility of using a shift work method for drivers was also provided. The criterion of maximum performance of the volume of transportation (mileage with cargo) of the fleet is applied while limiting the total duration of the process. The obtained analysis results are presented in Table 3.

When planning the transportation according to the initial data taken from incoming orders at the specified date, optimization was performed with a preliminary division of orders into clusters according to the order compatibility criterion $R_c \geq 0.5$, since the criterion for improving the plan was the maximum volume of

***Table 3** Improving the results of freight transportation plans in a trucking company*

		Known plan					Optimized plan according to the developed methodology				
Plan number / date	Number of orders received	Number of the orders accepted	Total mileage with cargo, thousand km	Number of drivers / trucks involved	Idle mileage, thousand km	Total project time*, hours	Number of orders accepted with clustering	Total mileage with cargo, thousand km	Number of drivers/ trucks involved	Idle mileage, thousand km	Total time/ duration of the project**, hours
1/25.08.11	280	32	14.9	45/32	1.1	378	115	89.9	100/56	12.3	2338/45
2/25.09.02	230	25	13.6	38/22	3.6	455	96	78.8	90/48	9.2	2095/43
3/25.09.03	265	24	9.2	38/22	2.7	313	103	73.4	100/50	4.6	1857/37
4/25.10.13	320	44	18.5	68/53	7.2	580	124	95.8	100/82	22.3	2811/44
5/25.11.20	189	17	24	25/22	2.1	212	62	47.4	60/30	3.0	1200/42

* the cumulative time spent by all truck crews on the routes. This time contains the downtime associated with customs procedures, so it was impossible to isolate the duration of logistics operations from the primary data;

** excluding downtime at customs.

transportation performed. As a result, some planning indicators deteriorated. However, the new transportation plans are generally more productive for the existing fleet of trucks, which under the old plans were idle, and drivers were left without work. Table 3 shows the idle mileage of trucks on international routes, which is increased from 3 to 27%. However, the number of orders, which Transservice-1 could fulfill if it worked according to optimized plans, would increase by 2.3 times. This could have been achieved not only by attracting reserve vehicles and drivers (their number barely doubled), but by improving the structure of transport cycles, as well. Thanks to the new planning, the average daily mileage of trucks also increased by approximately 20% and their downtime is reduced.

7 Conclusions and further work

The use of the compatibility index of freight orders is justified when planning the schedules of a team of drivers. Thanks to the use of such a criterion, some of the a priori irrational ways of finding the optimal solution are discarded. Organizational and technological compatibility makes it possible to divide the total volume of big data into clusters such that the search for the optimal schedule in the obtained sub-graphs gives a guaranteed result. The number of vertices in such subgraphs reduces the complexity of the scheduling calculations, on one hand, and provides a sufficient field of variables for optimizing the resulting schedule, on the other. A reasonable threshold value of the coefficient lies within $0.3 \le Rc \le 0.6$, i.e., within its domain of definition. The choice of such values leads to the fact that successful clustering is ensured, and the volume of big data can be processed by an optimization algorithm based on meta-heuristics. However, there are other criteria for improving logistics plans, depending on the business situation. Each optimization criterion requires a reasonable choice of the numerical value of R_c. For example, a clustering criterion value close to its upper limit should be used when there is a need to increase the productivity of fleet vehicles. Values close to the lower limit create the prerequisites for obtaining tighter schedules and more efficient use of available resources in the shortest time.

Scaling the driver scheduling problem by increasing the territory, that is, by extensive transportation planning, leads to worsening results. More successful is the intensive scaling: by increasing the area of the service area and the number of planned orders simultaneously. The settlement factor does not play a significant role in the quality of the preparation of initial data for optimization, if it is not considered together with the density characteristic of the transport network. In this work, less attention was paid to the parameters of the transport network, which may be a promising task for the future research.

Acknowledgment

The authors received no financial support for the research, authorship and/or publication of this article.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference

[1] REDMER, A., KISIELEWSKI, P., OBLAZA, J. Optimal driver scheduling with evolutionary approach in urban public transport. *Archives of Transport* [online]. 2025, **74**(2), p. 65-80 [accessed 2025-12-08]. ISSN 0866-9546, eISSN 2300-8830. Available from: https://doi.org/10.61089/aot2025.919ryp91

[2] LIM, C. H., MOON, S. K. A two-phase iterative mathematical programming-based heuristic for a flexible job shop scheduling problem with transportation. *Applied Sciences* [online]. 2023, **13**(8), 5215 [accessed 2024-04-30]. ISSN 2076-3417. Available from: https://doi.org/10.3390/app13085215

[3] BENUS, J., DEMIRCI, E. Regulation (EC) No 561/2006-review of the adopted changes on 15 July 2020. *Automotive Archives* [online]. 2020, **90**(4), p. 59-73 [accessed 2025-12-25]. ISSN 1234-754X. Available from: https:// doi.org/10.14669/AM.VOL90.ART5

[4] SAUKENOVA, I., OLISKEVYCH, M., TARANIC, I., TOKTAMYSSOVA, A., ALIAKBARKYZY, D., PELO, R. Optimization of schedules for early garbage collection and disposal in the megapolis. *Eastern-European Journal of Enterprise Technologies* [online]. 2022, **1**(3), p. 12-23 [accessed 2022-02-28]. ISSN 1729-3774, eISSN 1729-3774. Available from: https://doi.org/10.15587/1729-4061.2022.251082

[5] SHANG, R. Efficient task scheduling for large-scale graph data processing in cloud computing: A particle swarm optimization approach. *Journal of Combinatorial Mathematics and Combinatorial Computing* [online]. 2024, **1**(122), p. 135-148 [accessed 2025-05-17]. ISSN 0835-3026, eISSN 2817-576X. Available from: https://doi.org/10.61091/jcmcc122-11

[6] EBRAHIMI, H., REZAEIAN‐MARJANI, S., GALVANI, S., TALAVAT, V. Probabilistic optimal planning in active distribution networks considering non‐linear loads based on data clustering method. *IET Generation, Transmission & Distribution* [online]. 2022, **16**(4), 686-702 [accessed 2025-12-31]. ISSN 1751-8687, eISSN 1751-8695. Available from: https://doi.org/10.1049/gtd2.12320

[7] OMOTEHINWA, T. Examining the developments in scheduling algorithms research: A bibliometric approach. *Heliyon* [online]. 2022, **8**(5), e09510 [accessed 2025-09-12]. ISSN 2405-8440. Available from: https://doi.org/10.1016/j.heliyon.2022.e09510

[8] SUN, Y., DUAN, S., YEN, J., XIONG, W., JIN, M., WANG, Y. Cluster Hopper: Cross-region order dispatching optimization for ride-hailing drivers. *Expert Systems with Applications* [online]. 2025, **289**, 127878 [accessed 2025-05-28]. ISSN 0957-4174. Available from: https://doi.org/10.1016/j.eswa.2025.127878

[9] SHOMAN, W., YEH, S., SPREI, F., KOHLER, J., PLOTZ, P., TODOROV, Y., RANTALA, S., SPETH, D. A review of big data in road freight transport modeling: gaps and potentials. *Data Science for Transportation*, [online]. 2023, **5**(1), 2 [accessed 2025-08-14]. ISSN 2948-135X. Available from: https://doi.org/10.1007/s42421-023-00065-y

[10] BRANKE, J., MATTFELD, D. Anticipation and flexibility in dynamic scheduling. *International Journal of Production Research* [online]. 2005, **43**(15), p. 3103-3129 [accessed 2024-06-16]. ISSN 0020-7543, eISSN 1366-588X. Available from: https://doi.org/10.1080/00207540500077140

[11] WEN, X., SUN, X., SUN, Y., YUE, X. Airline crew scheduling: Models and algorithms. *Transportation Research Part E: Logistics and Transportation Review* [online]. 2021, **149**, 102304 [accessed 2025-12-25]. ISSN 1366-5545, eISSN 1878-5794. Available from: https://doi.org/10.1016/j.tre.2021.102304

[12] SARTORI, C., SMET, P., VANDEN BERGHE, G. Scheduling truck drivers with interdependent routes under European Union regulations. *European Journal of Operational Research* [online]. 2022, **298**(1), p. 76-88 [accessed 2025-12-10]. ISSN 0377-2217. Available from: https://doi.org/10.1016/j.ejor.2021.06.019

[13] WANG, B., HAN, Y., WANG, S., TIAN, D., CAI, M., LIU, M., WANG, L. A review of intelligent connected vehicle cooperative driving development. *Mathematics* [online]. 2022, **10**(19), 3635 [accessed 2025-05-16]. ISSN 2227-7390. Available from: https://doi.org/10.3390/math10193635

[14] FLAMINI, M., NICOSIA, G. A generalized disjunctive graph model for a complex production problem. In: 2023 International Conference on Industry Sciences and Computer Science Innovation: proceedings. Vol. 237. Procedia Computer Science. 2024. ISSN 1877-0509, p. 289-296.

[15] VAKA, D. From complexity to simplicity: AI's route optimization in supply chain management. *Journal of Artificial Intelligence, Machine Learning and Data Science* [online]. 2024, **2**(1), p. 386-389 [accessed 2025-06-20]. ISSN 2583-9888. Available from: https://doi.org/10.51219/JAIMLD/dilip-kumar-vaka/100

[16] MAHDI, M., HOSNY, KHALID M., ELHENAWY, I. Scalable clustering algorithms for big data: a review. *IEEE Access* [online]. 2021, **9**, p. 80015-80027 [accessed 2026-03-12]. ISSN 2169-3536, eISSN 2169-3536. Available from: https://doi.org/10.1109/ACCESS.2021.3084057

[17] KLUMPP, M., SEVERIN, B., LECHTE, H., MENCK, J. H. D., KEIL, M., STRAUB, S. M., HESENIUS, M. Driving big data-integration and synchronization of data sources for artificial intelligence applications with the example of truck driver work stress and strain analysis. In: International Journal of Technology and Human Interaction Conference: ICIS 2022: proceedings. Vol. 20, Iss. 1. 2024. ISBN 978-171389361-5, p. 125-130.

[18] IKEGWU, A., NWEKE, H., ANIKWE, C. ALO, U., OKONKWO, O. Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing* [online]. 2022, **25**(5), p. 3343-3387 [accessed 2025-07-25]. ISSN 1386-7857. Available from: https://doi.org/10.1007/s10586-022-03568-5

[19] ABDALLA, H. A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data* [online]. 2022, **9**(1), 107 [accessed 2024-05-19]. ISSN 2196-1115, Available from: https://doi.org/10.1186/s40537-022-00659-3

[20] TALMALE, G., SHRAWANKAR, U. Tasks scheduling using dynamic cluster-based hierarchical real-time scheduler for autonomous car. *Ambient Science* [online]. 2021, **8**(2), p. 01-06 [accessed 2025-04-16]. ISSN 2348-5191, eISSN 2348-8980. Available from: https://doi.org/10.21276/ambi.2021.08.2.ga01

[21] AL-JUMAILI, A., MUNIYANDI, R., HASAN, M., PAW, J., SINGH, M. Big data analytics using cloud computing based frameworks for power management systems: status, constraints, and future recommendations. *Sensors* [online]. 2023, **23**(6), 2952 [accessed 2025-04-24]. eISSN 1424-8220. Available from: https://doi.org/10.3390/s23062952

[22] NOVAK, A., SUCHA, P., NOVOTNY, M., STEC, R., HANZALEK, Z. Scheduling jobs with normally distributed processing times on parallel machines. *European Journal of Operational Research* [online]. 2022, **297**(2), p. 422-441 [accessed 2025-12-19]. ISSN 0377-2217. Available from: https://doi.org/10.1016/j.ejor.2021.05.011

[23] LI, N., SHENG, H., WANG, P., JIA, Y., YANG, Z., JIN, Z. Modeling categorized truck arrivals at ports: big data for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems* [online]. 2022, **24**(3), p. 2772-2788 [accessed 2025-10-08]. ISSN 1524-9050. Available from: https://doi.org/10.1109/TITS.2022.3219882

[24] TALAAT, A., IIJIMA, J., GHEITH, M., ELTAWIL, A. An integrated k-means clustering and bi-objective optimization approach for external trucks scheduling in container terminals. In: *Industrial Engineering and Applications*. *Vol. 35*. *Advances in Transdisciplinary Engineering*. Phuket: IOS Press, 2023. ISBN 978-164368408-6, p. 805-814.

[25] KHOMYN, N., OLISKEVYCH, M., TARAN, I., MURATBEKOVA, G. Algorithm for scheduling drivers on intercity road routes: case study of the shift method. *National Mining University. Scientific Bulletin* [online]. 2025, **4**, p. 185-194 [accessed 2025-12-06]. ISSN 2071-2227, eISSN 2223-2362. Available from: https://doi.org/10.33271/nvngu/2025-4/185

[26] Cargo transportation in Ukraine and Europe - Trans-Service-1 [online] [accessed 2026-01-16]. Available from: https://www.trans-service-1.com.ua/en/home/