A DATA-MINING TECHNIQUE FOR THE PLANNING AND ORGANIZATION OF TRUCK PLATOONS



Philipp MEISEN **Thomas SEIDL** **Klaus HENNING**
RWTH Aachen University
Aachen, Germany

Abstract
In this paper, we introduce the problem of mining frequent sub-routes for the use of platoon driving. The problem is related to the problem of mining frequent sequences. While the problem of mining frequent sequences normally focuses on patterns with a large support, the problem of mining frequent sub-routes has to deal with a small absolute support of two or more. We will present our algorithm to find truck platoon sequential pattern (TPSpan) which is based on pattern-growth methods and solves the problem of mining frequent sub-routes in an efficient way by using so called pruning parameters. We also present experimental results about the amount of possible truck platoons based on synthetic data and about the possibility of finding truck platoons with the use of pruning parameters in context of efficiency, flexibility and reliability.

Keywords: Truck platoon, Platoon driving, Road capacity, Data mining, Sequential pattern mining, Mining frequent sub-routes

Résumé
Dans cet article, nous introduisons le problème de recherche d'itinéraires secondaires pour l'utilisation de la conduite en peloton. Le problème est lié à la recherche de séquences fréquentes. Bien que ce problème soit normalement abordé du point de vue de la recherche sur un support large, dans notre cas, nous devons travailler avec un support restreint faible. Nous présentons notre algorithme pour trouver des séquences de poids lourds en peloton, fondé sur des méthodes d'augmentation de taille d'échantillon et qui permet de résoudre ce problème efficacement en utilisant des paramètres permettant l'élagage de certaines zones.
Nous présentons aussi des résultats expérimentaux sur le nombre éventuel de poids lourd en peloton à partir de données synthétisées et sur la possibilité de retrouver des pelotons avec l'utilisation de l'élagage paramétriques avec des contraintes d'efficacité, de flexibilité et de fiabilité.

Mots-clés: Pelotons de poids lourds, conduite en peloton, capacité d'un réseau, fouille de données, recherche de pattern, recherche de sous-groupes.

1. Introduction

Based on studies of the European Commission in 2006 for traffic development, road freight transport rose by 28 % between 1995 and 2005. Furthermore, until the year 2020 an increase in the sector of private transportation and an increase of 41 % in road transportation are expected. This development is dramatic in consideration of the fact that the road networks in most of Europe are already overloaded (EU-Commission, 2006).

One possibility to deal with this problem is to reduce the gap between vehicles on the road and achieve a better utilization of road networks. Because of safety reasons, advanced driver assistance systems have to support the drivers riding in their vehicles at very short distances from each other or allow automated driving. The concept of coupling trucks to “trains on road” with the aid of such assistance systems is a first step to achieve a better utilization using the idea of reduced gaps (Henning/Preuschoff, 2003). In the case of truck platoons, the trucks are electronically coupled keeping a distance from each other of approximately 10 meters. The steering data of the first truck and additional sensor data are used to realize an automated driving of the following trucks. In this case, there are more benefits expected than a better utilization of the road networks, e.g. an optimization of traffic flow and reduction of fuel consumption induced by slipstream driving (Schulze, 2001). Furthermore, a reduction of the driver’s workload and probability of lowered accident risks could be possible. For planning and organization of such truck platoons, as well as evaluation and simulation purposes, it is necessary to implement an algorithm which searches for economic truck platoons within the planned or current driven routes of all participating trucks. Such an algorithm is also needed for further research in the area of organized platoon driving. In any case of organized platoon driving, it is essential to search for and group possible participants. In principle, the “common route” is one general search criterion, whereas other criteria (e.g. profit, waiting-time, properties of the participating vehicles) can differ because of the focus.

In this paper, we present a data-mining technique to solve the aforementioned problem of searching for possible platoons which satisfy given specific criteria (in the case of truck platoons: economy). Before introducing our TPSpan-Algorithm and the possibility to use this algorithm as a solution for the problem of planning and organizing truck platoons in section 4, we will formalize the problem in section 2. In section 3, we will present related work to this topic within the field of data-mining. We conclude this paper in section 5 with experimental results.

2. Problem Description

The problem of planning and organizing truck platoons can be divided into different smaller problems. Before we describe those problems in detail, we have to define the basic terms: section, route and platoon. Another important term in the case of truck platoons is economy. Due to the fact that the system is generally used by carriers to transport goods as fast and cheap as possible, the economy is important for the acceptance of such a system. Without adequate economy, the system would not be used and thus useless.

2.1 Basic Terms and Definitions

The term route is used in everyday language as a list of waypoints starting at a given point and leading to a specific destination. Before we define the term route for this paper, we want to define the distance between two waypoints as a sector, whereas the route, and thus the list of waypoints, is generated by a route planner based on a digital map (i.e. a graph of a chosen infrastructure).

Definition Section
A *section* s is a non-divisible piece of a road, typically given by the underlying digital map. $K_A(s)$ identifies the start and $K_E(s)$ the end point of the section (e.g. as GPS-Coordinates). The distance of the section (e.g. in kilometers) is indicated by $Length(s)$. Furthermore, we define $Meet(s)$ as a meeting opportunity. $Meet(s)$ equals 1, if and only if there is a meeting point (e.g. rest area) within the section s, otherwise $Meet(s)$ is equal to 0.

Definition Route
A *route* r is a sequence of *sections*, denoted as $r = (s_1, \cdots, s_n)$, whereby for all $n > 0$ and $0 < i < n$ $K_E(s_i) = K_A(s_{i+1})$. A *part of a route* $r = (s_1, \cdots, s_n)$ is a cut-out of r, i.e. for $i, j > 0$ and $i, j \le n$ $Part(r)_i^j = \begin{cases} (s_i, \cdots, s_j) & i \le j \\ (\ ) & else \end{cases}$. A route $r' = (s'_1 \cdots s'_m)$ is called a *sub-route* of a route $r = (s_1 \cdots s_n)$, denoted as $r' \rhd r$, if and only if $i > 0$ exists, so that $r' = Part(r)_i^{i+m-1}$.
A route $r = (s_1, \cdots, s_n)$ is *leading*, if for all $i, j \le n$ and $i \ne j$ $s_i \ne s_j$ applies.

Definition Platoon
Let $R = \{r_1, \cdots, r_m\}$ be a set of routes as defined above. A platoon is a route $l = (s_1, \cdots, s_n)$ with $n > 0$, $Meet(s_1) = 1$ and $\exists w, v \in R, w \ne v$ $l \rhd w \wedge l \rhd v$. The *size of a platoon* l is defined as the number of possible platoon participants, i.e. $Size(l) = \left|\{w \in R \mid l \rhd w\}\right|$. The *length of a platoon* l is defined as the distance which can be driven together as a platoon, i.e. $Length(l) = \sum_{i=1}^{n} Length(s_i)$.

2.2 The Problem of Mining Frequent Sub-routes

The basic problem of planning and organizing truck platoons, is finding common routes among all participants. The idea of using data-mining techniques for the search of economic truck platoons is motivated by the huge amount of data which has to be stored in a database and evaluated on a central server. The database has to store every route of every participating vehicle. To do so, a record of the database consists of a vehicle-id, a section-id and a timestamp (Table 1). The vehicle-id is unique for each participant. The section-id identifies a section, as defined in 2.1, whereas the timestamp indicates the predicted time when the truck will pass this section. In this paper, we focus only on databases with leading routes.

Table 1 – Route-Database

Vehicle-id	Timestamp	Section-id
1	0	2
1	2	3
1	3	4
1	4	6
2	1	4
2	4	1

Vehicle-id	Timestamp	Section-id
2	6	5
3	0	2
3	1	3
3	2	4
3	3	1
3	5	7

A route can be read from the database as a timestamp-sorted sequence of section-ids. Figure 1 illustrates the route of vehicle 1 (vehicle_id = 1) stored in the database shown in Table 1.

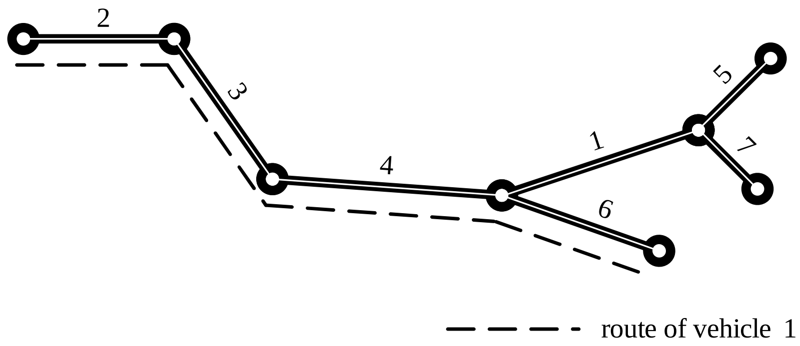

Figure 1 – Illustration of the Route of Vehicle 1 from Table 1

The problem of mining frequent sub-routes can be described as following: "We are given a database of leading routes. Each route consists of a set of records, grouped by the truck-id, with the fields mentioned above. We say that a route $r = (s_1, \cdots, s_n)$ *supports* a route $r' = (s'_1 \cdots s'_m)$ if the route r' is a sub-route of r. The $support_D(r)$ within a database D is the absolute amount of routes which supports r. The problem of mining frequent sub-routes is to find all sub-routes that have a support of at least two."

The following example illustrates this description: Consider the database shown in Table 1. The route $r_1 = (4, 6)$ has a support of 1, whereas the route $r_2 = (3, 4)$ has a support of 2. The reason for this is that r_1 is only supported by the route of vehicle 1 while r_2 is supported by vehicle 1 and 3.

By solving the problem of mining frequent sub-routes, we also solve the problem of finding possible platoons within a database. The reason for that is obvious: Every frequent sub-route with a meeting opportunity at the starting section is, as defined, a platoon, whereas frequent means, that a sub-route is driven by two or more vehicles. To find economic truck platoons, we have to define the factors which influence the economy of truck platoons.

2.3 Economy of Truck Platoons

The economy of a truck platoons depends on several criteria. One important criterion for the economy of truck platoons is the profit which can be achieved by driving as a platoon. This profit can be simply calculated by the proceeds of less fuel consumption which depends on the length of the frequent sub-route (i.e. in kilometers) and the number of participants minus the extra personnel costs which have to be paid while others wait for the rest of the platoon participants. As mentioned above, the profit is only one criterion of economic truck platoons. Another, more general criterion, would be the expected reduction of accidents and therefore a possible reduction of insurance contribution. When determining the economy of truck platoons only those criteria can be considered which are influenced by properties of the platoon, e.g. common distance, number and motorization of participants or weight-to-power ratio. All other criteria have to be regarded on a more global economical view. In this paper, we focus on the criteria shown in Table 2 which are assigned to different groups. *Exclusion criteria* are criteria which exclude a vehicle from a platoon, whereas *grouping criteria* are criteria which assign a vehicle to a specific group of other vehicles which can form a platoon. Finally, the *assessment criteria* are used to decide if a platoon is valid or not. In the case of truck platoons, assessment criteria are used to decide if the platoon is economic or not.

Table 2 – Criteria of Economic Truck Platoons

Exclusion Criteria	Grouping Criteria	Assessment Criteria
dangerous goods	motorization	common distance (length)
	weight-to-power ratio	participants (size)
	loading weight	fuel consumption
	brake power	waiting time
		wage rate
		gas price

By determining all frequent sub-routes of grouped trucks and calculating their profit with the given assessment criteria, we can find profitable and due to the reasons stated above, economic truck platoons.

Reduced Fuel Consumption

The fuel consumption which is induced by slipstream driving in a platoon can be theoretically calculated. Several simulations and measurements have shown that the theoretical results cannot be achieved in practice (Table 3).

Table 3 – Possible Fuel Consumption (Bonnet/Fritz, 2002)

	Theoretical	Simulation Daimler	Measurements Daimler
Fuel Consumption (1st vehicle)	2,17% (14t); 1,64% (28t);	2% (28t)	6% (14t)
Fuel Consumption (2nd vehicle)	38,06% (14t); 28,76% (28t);	19% (28t)	21% (28t)

On account of this and further simulations the expected fuel consumption for truck platoons is assumed as 2 % for the 1st truck, 11 % for the 2nd and 13 % for every other participant. Because of the reason that the profit of driving in a platoon on the 1st, 2nd or another position differs, it was furthermore assumed that every truck drives on average the same distance on each position of a platoon. Under those assumptions, the following equation (1) shows the fuel consumption of truck platoons with the size n.

$$\Delta B_{e,n} = \frac{2\% + 11\% + (n-2)\cdot 13\%}{n} \tag{1}$$

With equation (1) and the assessment criteria from Table 2, it is possible to calculate the fuel consumption of truck i achieved by driving with the platoon l with equation (2), whereas u_i indicates the fuel consumption of the truck on 100 km and k_{KS} the price of one liter of gas.

$$Proceeds_i(l) = u_i \cdot \frac{Length(l)}{100} \cdot \Delta B_{e,Size(l)} \cdot k_{KS} \tag{2}$$

Extra Personnel Costs

The extra personnel costs for participating in a truck platoon are caused by several criteria. First of all, the participants have to wait for other participants at an arranged meeting point or reduce speed until they got together. Another criterion for extra personnel costs is the platoon driving itself and the associated platoon maneuvers (e.g. a platoon has to be dissolved at a working area). It is also possible that a truck, which participates in a platoon, has to drive with reduced speed because of another, slower participant (e.g. uphill).

Due to the fact that the extra personnel costs caused by platoon maneuvers or speed reduction because of slow participants are unascertainable without accurate road and traffic information, therefore those influencing criteria will be ignored. The only criteria which take account of extra personnel costs will be the waiting time $Time_i(l)$ for each participating truck i and the wage rate k_{PK}.

$$Costs_i(l) = k_{PK} \cdot Time_i(l) \tag{3}$$

3. Related Work

Discovering patterns in sequences is an important knowledge discovery and data-mining research area. There are many different interesting fields like discovering rules in so called events (Mannila et al., 1997) or finding text subsequences (Califano/Rigoutsos, 1993) that match a regular expression. In 1995, Agrawal and Srikant introduced the problem of mining sequential patterns: “A database with customer transactions is given. Each transaction consists of the fields; customer-id, transaction-time and a set of items purchased in this transaction. Quantities of items purchased in a transaction are not considered. The problem of mining sequential patterns is to find all sequences that have a certain user-specified minimal support.”

This difficulty deals with the questions; “which items are bought in which order” or “which item will be bought next”. Agrawal and Srikant presented three algorithms for solving the sequential pattern problem and many other algorithms have been proposed to speed up the mining process (e.g. Zaki, 2001; Pei et al., 2004). In general, those algorithms are based on two different methods for mining sequential patterns: apriori-based methods and pattern-growth methods (Antunes/Oliveira, 2004).

The apriori-based methods are, as the name states, based on the so called apriori condition. This condition is used to generate larger candidates based on smaller frequent sequences in the so called join-step. Each generated candidate is then validated in the following so called prune-step. These two steps are repeated until there are no more frequent sequences found. The pattern-growth methods deduce from a smaller frequent sequence, a longer sequence. Instead of generating candidates, these methods directly look for frequent sequence.

In 2003, Yan et al. introduced the problem of mining closed sequential patterns which means that instead of finding all frequent sequences, only the “most supported” sequences are searched for. They presented a new algorithm to find closed sequential pattern, called CloSpan, which is based on a pattern-growth method introduced in 2001 by Pei et al. Figure 2 illustrates the main idea of the approach. The items are nodes in a lexicographical sequence tree. Each path of the tree is a frequent sequence found in the database. So called projected databases are stored for each node of the tree. If the projected databases of two different paths are equal to each other, the mining for more frequent sequences can be eventually stopped. In

Figure 2, the mining for further frequent sequences can be stopped at node B in the path B because the projected database of this node is equal to the projected database of node B in the path A, B.

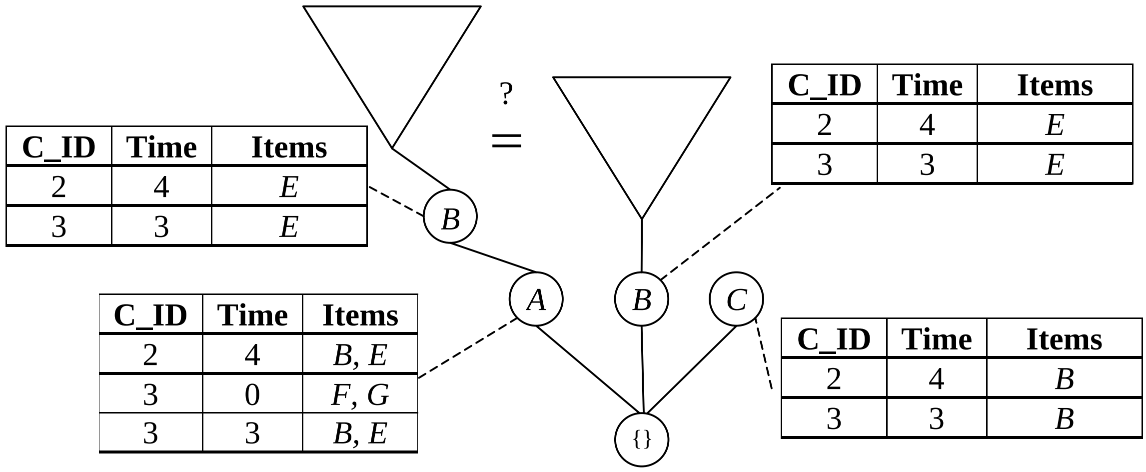

Figure 2 – Illustration of the Algorithm to Find Closed Sequential Pattern

The problem of mining frequent sub-routes or economic truck platoons cannot be solved easily by any of those algorithms. Due to the small support of two and the given sequences of hundreds of sections, those algorithms collapse. In the case of platoons the support has to be two, because two is the smallest number of participants to form a platoon. However, these algorithms are good approaches in finding possible solutions for solving this problem.

4. The Truck Platoon Sequential Pattern Algorithm

In this section, we will introduce our new algorithm which searches for possible platoons in a route database, especially optimized to find economic truck platoons. Our so called TPSpan-Algorithm (Mining Truck Platoon Sequential pattern) is split in four phases and illustrated in Figure 3.

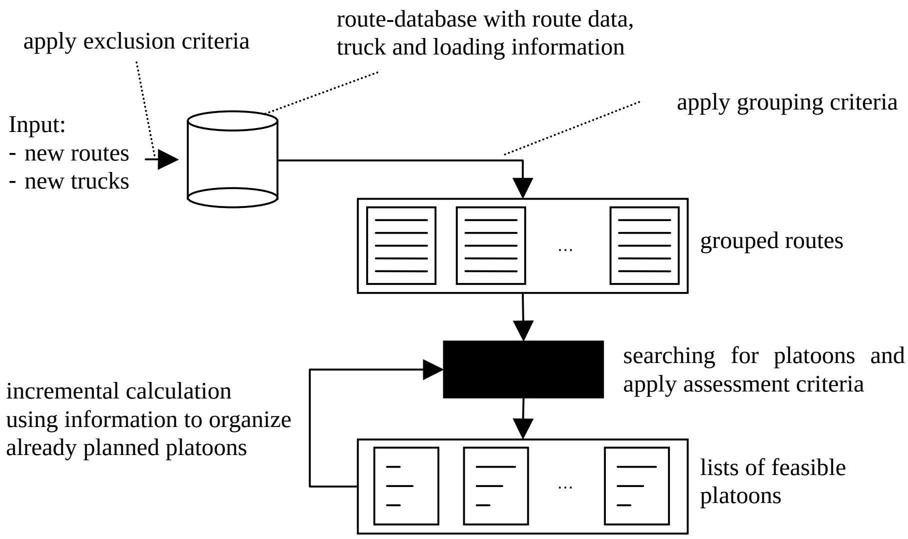

Figure 3 – Illustration of Our New Algorithm to Find Truck Platoon Sequential Pattern

In the first phase, the exclusion criteria are used to eliminate trucks or routes which do not meet those criteria. Within the second phase, the grouping criteria are applied to secure the forming of platoons within similar conditions. In the third phase, the data-mining technique checks each group for possible economic truck platoons regarding the assessment criteria. Finally, in the fourth and last phase, the economic truck platoons are grouped by truck-id. The grouped information is then transmitted to each truck (e.g. meeting points, profit).

The data-mining technique used in the third phase is based on the projected pattern-growth idea presented by Pei et al. (2001) and has been enhanced for solving the problem of mining frequent sub-routes. It generates a lexicographic tree as shown in Figure 2. Instead of starting with each frequent section, our tree starts with those frequent sections which have a meeting opportunity ($Meet(s_1) = 1$). This means that the first section has to be e.g. a rest area. Due to this, the width of the tree is reduced and the condition for a platoon (starting with a meeting opportunity) is ensured. Another important improvement is our so called node-compress-method which is used to reduce the validation of the assessment criteria. Instead of validating the criteria for every platoon, the common distance of a platoon will be increased as long as the number of participants does not change. It is ensured that the algorithm will still search for the most economic platoons: A platoon $l_1 = (s_1, \cdots, s_n)$ with $Size(l_1) = k$ is always more economical than a platoon $l_2 = (s_1, \cdots, s_n)$ with $Size(l_2) = k$, whereas $Length(l_1) > Length(l_2)$.

Table 4 shows the framework of the data mining technique. The algorithm is working from the root of the tree to the leaves. In each node, the algorithm calculates possible platoons (line 4) and determines new children (line 6). If the projected database of the node does not support any more platoons (line 1) the recursion terminates. The framework also shows the already mentioned node-compress-method (line 3) which returns a true value if the compression terminated with a split. If this is the case, further children are possible, otherwise the routes of the projected database end and no more children can be determined (line 5).

Table 4 – Framework of the Introduced Data-Mining Technique

Node::calculate()	
1.	if (Node.Depth == 1 && ProjRDB.Support < MinPlatoonSize)
2.	Return
3.	bSplit = ProjRDB.compress()
4.	ProjRDB.calculatePlatoons()
5.	if (bSplit)
6.	Node.createChildren()
7.	for each NodeChild in Node
8.	NodeChild.calculate

As shown in Figure 3 the algorithm also uses information that has already been calculated to organize planned platoons by calculating the tree incrementally. Comparing the differences between the trees enables the organization of platoons, e.g. delay of participants at the meeting point or cancelling of planned platoons. The main goal of the incremental calculation is to reach a comparison of root-nodes. If a root-node, i.e. a node on the first level of the tree, has been changed, the sub-tree is recalculated and existing changes are recorded, analyzed and submitted.

5. Experimental Evaluations

For evaluation and simulation purposes synthetic data were generated by calculating random routes using a routing planner. To do so, two zip codes were randomly chosen from all postal codes in Germany. The departure time was also randomly picked within one day. The calculated routes were saved in a database (Table 1). With this approach three different databases with 500, 2,000 and 5,000 routes where generated. Experimental evaluation has shown that our TPSpan-Algorithm finds profitable truck platoons within generated databases. The experiments have also shown that the number of economic truck platoons increase exponentially with the number of routes and that scaling parameters are necessary to prune the amount of possible truck platoons (i.e. when there are more than 2,000 routes). Within the experimental evaluations, we identified the following scaling parameters: minimal common distance in kilometers, minimal and maximal truck platoon size, maximal waiting time in seconds and minimal profit in €.

Figure 4 shows three diagrams about the efficiency, the flexibility and the reliability of truck platoons. The minimal and maximal truck platoon size was set to two and four. The maximal waiting time was 10,000 s, the minimal common distance was 1 km and the minimal profit was 1 €, if not chosen to be the value of the x-axis. We say that a platoon is efficient, if and only if the proceeds are above the costs. Furthermore, we say that a platoon is flexible, if alternatives to participate with a platoon are given. The reliability of truck platoons focus on the chance of finding a platoon to participate with. The experiments have shown that the best values for the known pruning parameters are a maximal waiting time of 3,000 s, a minimal common distance of 100 km and a minimal profit of 1 €. With those values for the pruning parameters, efficiency, flexibility and reliability are still assured (Friedrichs, 2008).

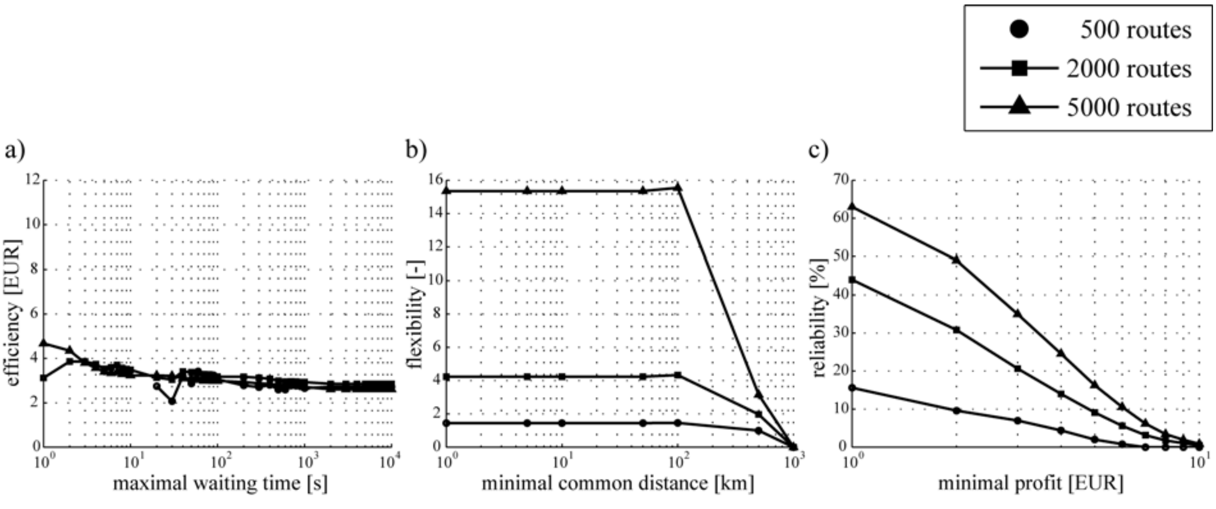

Figure 4 – a) Efficiency: Average Profit per Truck; b) Flexibility: Average Number of Possible Truck Platoons to Participate with; c) Reliability: Average Chance to Find a Platoon to Participate with (Friedrichs, 2008)

6. Conclusion

In this paper, we have introduced a data mining technique to plan and organize platoons. Furthermore, we have introduced and presented experimental results for an application area “truck platoons: trains on road”. For this case, we have examined that it is possible to find truck platoons and that the amount of platoons increase exponentially with the amount of routes (participating trucks). Due to this rise, pruning parameters are necessary. Further experiments have shown that it is possible to find truck platoons reliable, efficient and flexible, even if pruning parameters are used. We have given suggestions for these parameters to achieve the mentioned factors. Further work has to be done in the field of evaluation. One important evaluation topic is the validation of the results presented in this paper with experiments based on real data. Another area of further work is the field of organizations to insure savings through truck platoons. Truck platoons have to be planned and organized by a trustful, independent cross-shipping organization, whereas more research is necessary to identify further needs and requirements.

7. References

- Agrawal, R., Srikant, R. (1995), “Mining Sequential Patterns”, In Proc. of the 11th Int’l Conf. on Data Engineering, Taipei (Taiwan).
- Antunes C., Oliveira, A. L. (2004), “Sequential Pattern Mining Algorithms: Trade-offs between Speed and Memory”, In Proc. of the 2nd Int’l Workshop on Mining Graphs, Trees and Sequences (ECML/PKDD), Pisa (Italy).
- Bonnet,C., Fritz, H. (2002), “Fuel Consumption Reduction Experienced By Two PROMOTE-CHAUFFEUR Trucks In Electronic Towbar Operation”, Brussels (Belgium).
- Califano, A. and Rigoutsos I. (1993), “Flash: A fast look-up algorithm for string homology”, In Proc. of the 1st Int’l Conf. on Intelligent Systems for Molecular Biology, Bethesda (USA).
- Commission of the European Communities (2006), “Keep Europe moving - Sustainable mobility for our continent”, Mid-term review of the European Commission’s 2001 Transport White Paper.
- Friedrichs, A. (2008), “A Driver Information System for Truck Platoons”, Doctoral thesis at RWTH Aachen University (in press), VDI Verlag, Düsseldorf (Germany).
- Henning, K., Preuschoff, E. (2003), “Einsatzszenarien für Fahrerassistenzsysteme im Strassengüterverkehr und deren Bewertung”, VDI Fortschritt-Berichte, Reihe 12, Nr. 531. Düsseldorf (Germany).
- Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M-C. (2001), “PrefixSpan: Mining Sequential Patterns Efficiently by PrefixProjected Pattern Growth”, In Proc. of the Int’l. Conf. Data Engineering (ICDE'01), Heidelberg (Germany).
- Schulze, M. (2001), “Die elektronische Deichsel – ein zukunftweisendes Fahrerassistenzsystem”, 4. Internationales Stuttgarter Symposium, Kraftfahrtwesen und Verbrennungsmotoren – Automative and Engine Technology, Stuttgart (Germany).
- Mannila, H., Toivonen H., Inkeri Verkamo A. (1997), “Discovery of frequent Episodes in Event Sequences. In Data Mining and Knowledge Discovery, Volume 1, pp. 259 – 289.
- Yan, X., Han, J., Afshar, R. (2003), “CloSpan: Mining Closed Sequential Patterns in Large Datasets”, In SDM'03, pp. 166 – 177, San Fransisco (USA).
- Zaki, M. J. (2001), “SPADE: An Efficient Algorithm for Mining Frequent Sequences”, In Proc. of Machine Learning Journal, Special issue on Unsupervised Learning (Doug Fisher, ed.), Vol. 42 Nos. 1/2, pp 31 – 60.