Proceedings of the 33rd International Conference on
Information Technologies (InfoTech-2019)
19-20 September 2019, Bulgaria

PREDICTION OF TRAVEL TIME USING HISTORICAL DATA

Full paper

Raymond Sondak and Leon Rothkrantz

Czech Technical University in Prague; Delft University of Technology
e-mail: LJM.Rothkrantz@gmail.com
Czech Republic, The Netherlands

Abstract: To plan a route along the highways, car drivers have to take care of the regular jams especially in the rush hours. The delay in travel time caused by regular jams can be predicted by historical data composed of loggings of travel time on specific road segments. Information on travel time can be assessed by sensors along the roads or via smart phones. Usual these data are not public domain available but can be extracted from websites. These data must be averaged over similar days with respect to the following parameters: time of the day, day of the week, seasonal effects and weather conditions. In this paper we describe the process of how to store these data in a database.
Key words: historical travel data, route planner, polling website, shortest path algorithm

1. INTRODUCTION

Routing devices installed in a car or on a smart phone enable a car driver to find the shortest path from start to destination. In many cases a shortest path algorithm has been used. But many car drivers want to select the shortest path in time which can be different from the shortest path in distance. Especially in the rush hours or during incidents on the roads the regular travel time can be delayed and maybe there are alternatives to reach the destination in a shorter time.

Most route planners have real-time information about the traffic density on the roads. Given all this information the route planners select the best path from source to destination. This is a way of instant based planning. Future developments are difficult to predict and most planners don’t take care of future trends or expectations. Our assumption is that regular traffic jams appear in similar ways on similar days. It should be possible to match the current day with a similar day in the past and take the average congestion development as the most probable congestion for that day. We assumed that the congestion depends of four parameters the day of the week, time of the day, weather, and season and take care of holidays and special days.

We considered the network of highways in the Netherlands. Roads are segmented in parts of the roads from junction to junction. Only at junctions, cars can leave or join the highway. We logged the speed of cars on the highway from the past year and labelled the data with the parameters as defined before. To get homogeneous data we put road segments with traffic incidents or accidents in a different category. We used these data to model traffic jams caused by incidents or accidents. The speed of the cars was sensed by sensors in the rod surface, and by tracking smart phones. Using the speed of the cars, the average traveling time between the junctions was computed.

A dynamic route planner was developed based on the shortest path algorithm of Dijkstra. The travel times along the roads were selected from a designed historical database after selection of the best matching parameters for that day.

2. RELATED WORK

In 2001 [1] we started our research on dynamic route planners. First, we explored the use of knowledge-based systems to find the shortest path from start to destination. Car drivers traveling daily in the rush hours were tracked to discover alternative routes. The knowledge of car drivers about alternative routes was implemented in an expert system. A critical point was if many car drivers using the alternative routes, congestion appears also on the alternatives. A routing advice based on real time traffic information was requested.

For many years we researched congestion on the highways using Ant Based Control Algorithm (ABC) [2], [3]. Car drivers were tracked on the highways and the traveling time was used to design a real time control algorithm using the ABC-algorithm. The ABC algorithm is based on the food forage behaviour of ants. On their way to find food sources from the nest, ants mark their routes using a chemical component, called pheromone. Pheromones from different ants along the same route are added up, but pheromone evaporates in time. Ants choose always the strongest pheromone marked path, which corresponds to the shortest path. The ABC algorithm was used to find the shortest path in the network of highways avoiding congestions.

An interesting application was the use of smart phones to guide car drivers using a smart routing algorithm [4]. Car drivers were tracked using the GPS system on their smart phones and this information was send to a central server. Based on the tracking data the shortest path was computed using the dynamic Dijkstra algorithm and send to an individual car driver. In the current paper we take also into account future developments using historical data.

In [6] Ritzinger et al. surveyed many papers stressing the probabilistic nature of travel time estimations. Ehmke et al. [5] proposed to use historic data to estimate current delay. This approach is used in the paper.

In [7] Solomon et al. present an overview of the fast-growing body of research focussed on vehicle routing and scheduling. The authors discuss the traveling salesman problem, the shortest path algorithms, minimum spanning tree problem, and the pickup and delivery problem. In their paper they also present future developments.

At this moment there is much interest in the design of automated guided vehicles (AGV). These AGV need a special routing algorithm. In [8], Qiu et al. provides a survey paper on routing algorithms for AGV's.

In his survey paper [9], Psaraftis et al. reported about an explosion of research on routing algorithms after 2000. They considered three decades of publications and ordered the research papers in 11 categories. Their focus was dynamic vehicle routing, on-line vehicle routing and stochastic vehicle routing. Papers taking future developments into account were missing, the focus is on real time routing systems.

3. MODELLING REAL TIME TRAFFIC INFORMATION

3.1. Different sources of traffic information

Real time traffic information is provided by different sources:

- Routing planners on smart phones. Routing planners compute the shortest path from start to destination. Most routing planners take care of momentary congestion caused by increased traffic density on the roads or by blocking road segments caused by accidents or incidents.
- Traffic news broadcast via radio. The radio news is focussed on special accidents and incidents and provides advices how to reroute. Related with the weather forecast stations special warnings are broadcasted
- Special information sites on WWW provide a visualisation of traffic jams on the road map and additional explanation of causalities.

The traffic information provided by different news sources is not always consistent. News site are not synchronised and use different sources to get their information. In figure 1 we display some information providers. The MONICA system is composed of sensors in the road surfaces of highways on special locations. Via the sensors the speed of passing cars can be assessed and transmitted to central processing stations. Different models are developed to measure the average travel time on road segments. One of the problems is that there are different categories of cars such as lorries, busses, passenger cars driving with different speeds. Cars with routing devices on board use their GPS system to transmit location and time to central servers where the average speed can be computed. Camera systems installed along the road are able to measure the local speed of cars, but in combination with license plate recognition also the average speed over some trajectory.

Cameras installed along the highway are also used to monitor traffic streams. In case of an incident, or starting congestion the situation is observed by human operators in traffic centres. These operators give an interpretation of accidents and incidents and make guesses about the time it takes before the problems have been solved.

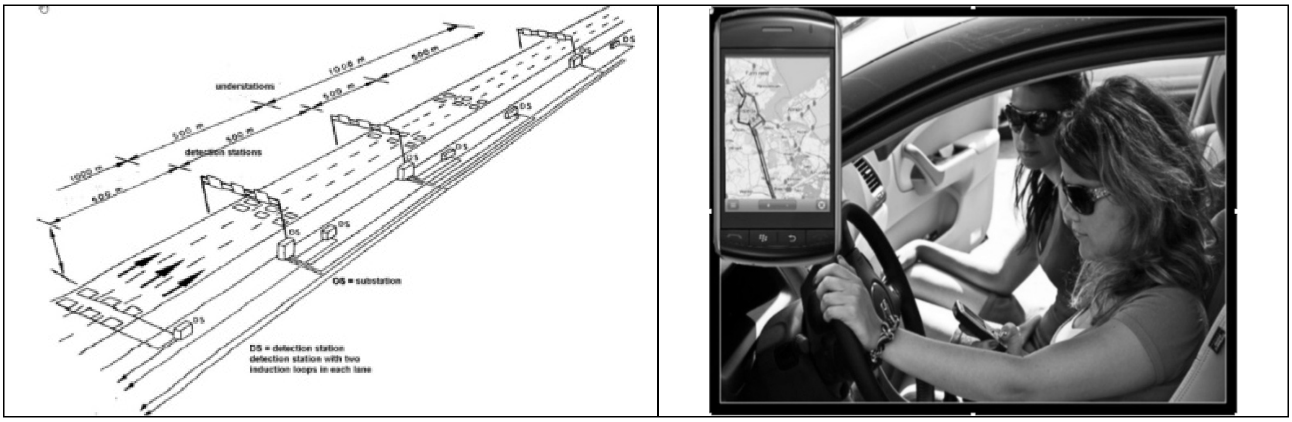


Fig. 1. Sensors on the road, cameras and smart phones as sources of travel information

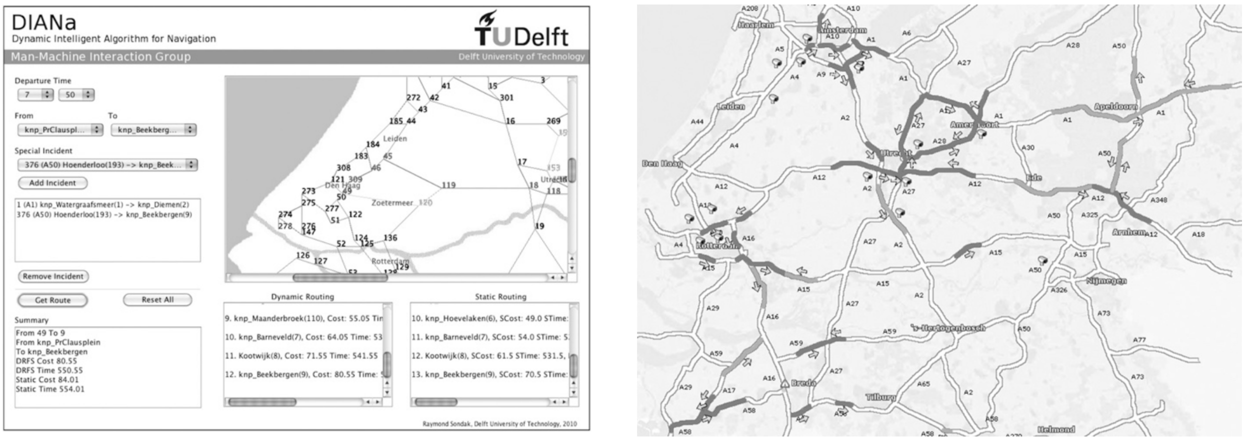


Fig. 2. Websites providing travel information

3.2. Conversion real time traffic info to historical database

From different data sources real time traffic information is presented. From the given data we computed the travel time from junction to junction. In this section we show how to extract travel information from public domain websites. We label this information with different parameters as time of the day, day of the week, and seasonal info as winter or summertime, weather information from the weather forecast. Important to know if delay is a regular to be expected one or caused by special incidents. In the traffic reports is usually reported if the cause of delay was an accident, an incident by road maintenance etc. This info is based on the interpretation of the operators in the traffic centre. Using their experience, the operators are able to predict how long the delay will be. We log the data every two minutes, day after day, so predictions of the traffic delay are not needed in our historical database.

In the next steps we describe how we processed extracted traffic data from a website.

- *Polling data from the website and clean-up*

A script was designed to extract traffic 24/7 data from the ANWB-website every two minutes during one year. A second script was designed to clean up the data, filter out all the duplications and unnecessary data.

- *Delay from junction to junction (point A to point B)*

To calculate the delay from junction to junction we look up the name of the nodes of the junction and the delay time between the nodes and then add the delay time accordingly to the database.

Fig. 3. Traffic information from junction to junction

- *From exit to exit between 2 adjacent junctions (Point A and B)*

When the traffic information is mentioned from exit to exit between two adjacent junctions we follow the same procedure as before.

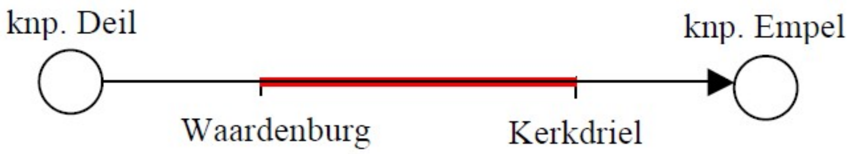


Fig. 4. Traffic information between two junctions

- *Several exits between two junctions*

In this case we simply add up the delay times from the successive exit to exit and the following delay time from exit to exit.

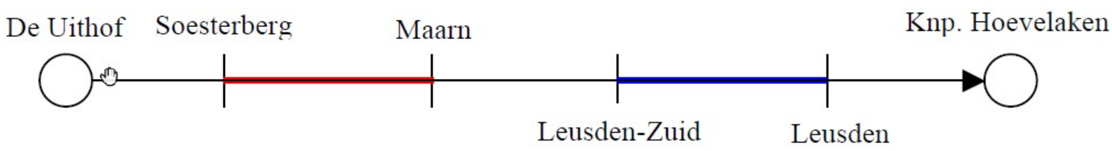


Fig. 5. Traffic information between two junctions

- *From exit/junction to exit/junction where there is/are another junction(s) in between (point A to C).*

Usually the traffic information is given between point A and point C. We allocate the time delay proportionally according to the distance from point A to B and from Point B to point C.

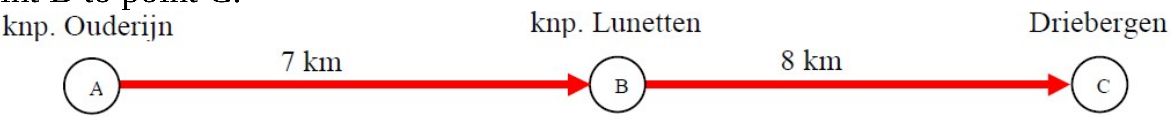


Fig. 6. Traffic information between several junctions

- *From two traffic information where the exits/junctions overlap each other*

In that case we delate the overlapping time, so that the time delay between the junctions will not be doubled.

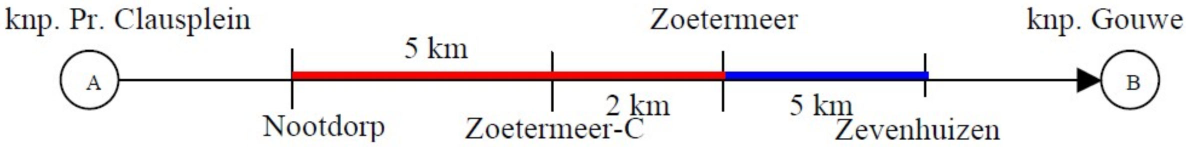


Fig. 7. Overlapping traffic information

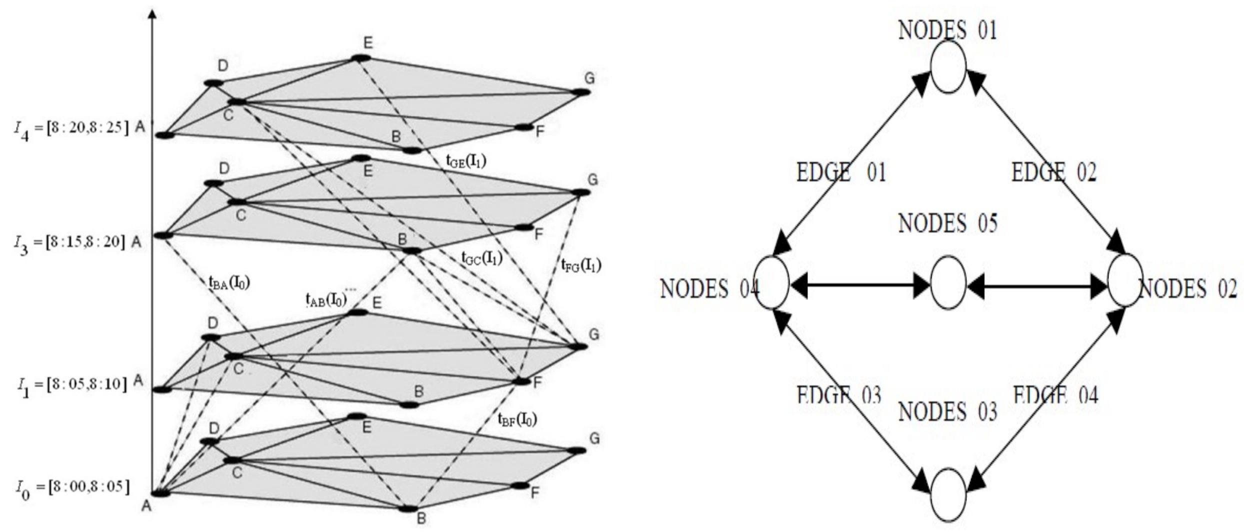
- *Traffic information with length and no delay information*

Traffic information with length but without delay duration is regarded as missing information. Next, we apply linear interpolation between data before and after the missing value to fill in the missing data.

We introduced some parameters as time of the day, day of the week, season and weather conditions. For all parameters and road segments, we averaged all available traffic data and store this value in the data base.

3.3. Database with historical traffic data

In order to build a traffic database, we start with a directed graph composed of junctions on the highways in the Netherlands. We consider only junctions that connect one highway to another, from highway to regional road and vice versa or regional road to regional road. For very edge and given time of the day we attach the travel time between the nodes (see figure 8).



(a) (b)

Fig. 8. (a) dimensional directed graph and (b) directed graph of highways in the Netherlands

Table 1. Traffic database

RD_ ID	RD_ NAME	ND_NAME_FR	ND_ID_ FR	ND_NAME_TO	ND_ID_ TO	DUR	…	17:10	…
2	A1	knp_Watergrafsmeer	1	knp_Diemen	1.80	…	…	9.30	…
3	A1	knp_Diemen	2	knp_Muiderberg	4.20	…	…	11.70	…
4	A1	knp_Muiderberg	3	Laren	6.60	…	…	6.60	…
…	…	…	…	…	…	…	…	…	…

The nodes in the directed graph, displayed in figure 9b are junctions and the edges that connect pair of nodes are the roads directions between junctions. Every node and edge will have a name and unique identifier (ID). Each road has length and maximum speed taken from public domain database. The traveling time is calculated using road length and maximum speed. If there is a delay due to congestion, the extra traveling time is added in one of the columns. The historical data is a matrix with dimension of

808x145, which consists of 808 road connections, road id column and 144 timestamp data.

4. DYNAMIC DIJKSTRA ALGORITHM

In figure 8a we display a 3-dimensional graph which is composed 2D directed graphs in horizontal planes ordered long a vertical time axis. At every time moment $t=t_n$ we have a graph of city-nodes and connecting street-links. The changing travel times for different moments are data from the historical database. The dynamic Dijkstra algorithm runs as follows. Let us assume that we want to travel from A to G on a given day. We look up the most similar day in our historical database with respect to the parameters day of the week, time of the day, season, and weather conditions and add the corresponding travel times to the links. At start $t=t_1$ we compute the shortest path from A-G using the historical travel time along the links on time $t=t_1$. On time $t=t_2$ we arrive at node B where we have the option to change our route. After arrival we update the travel times along the links and apply again static Dijkstra algorithm [10] to compute the shortest route from B-G. In this way we finally arrive at G along the shortest path with adapted travel times from the historical database along the links. In case of a long journey we may adapt the matching day or in case of an incident use special incident models.

5. TEST

To test our dynamic routing algorithm, we performed the following experiment. We start with 21 biggest cities in the Netherlands and consider the traveling time of travelers between these cities. As starting and endpoint we took the rings surrounding the cities. We took all Thursdays and computed the traveling time every 20 minutes between 15:20h and 18:20h, including the evening rush hours. (t_0=15.20, t_1=15.40, t_2=16.00, t_3=16.20, t_4=16.40, t_5=17.00, t_6=17.20, t_7=17.40, t_8=18, t_9=18.20). We computed a matrix of size 21x21. The rows and columns correspond with the 21 selected cities. In the entries of the matrix we indicated the difference in static and dynamic travel time between corresponding cities by grey values. We compared the static routing time and dynamic routing by using grey-colored matrices (see Fig. 9)

The grey-colored matrices show the comparison of the static version of Dijkstra algorithm and dynamic version using historic travel data. A day ends with light colored grey but when the evening rush starts, we observe dark colors of grey which means that the difference in static and dynamic routing time is increasing. The total saving of travel time was about 15 minutes.

We also observed a difference between cities. Big cities have a bigger traffic density than smaller cities and traffic jams start earlier and stay longer because of limited road capacity. Accidents and incidents were not considered in this experiment. Special days were considered as outliers and excluded from the experiments.

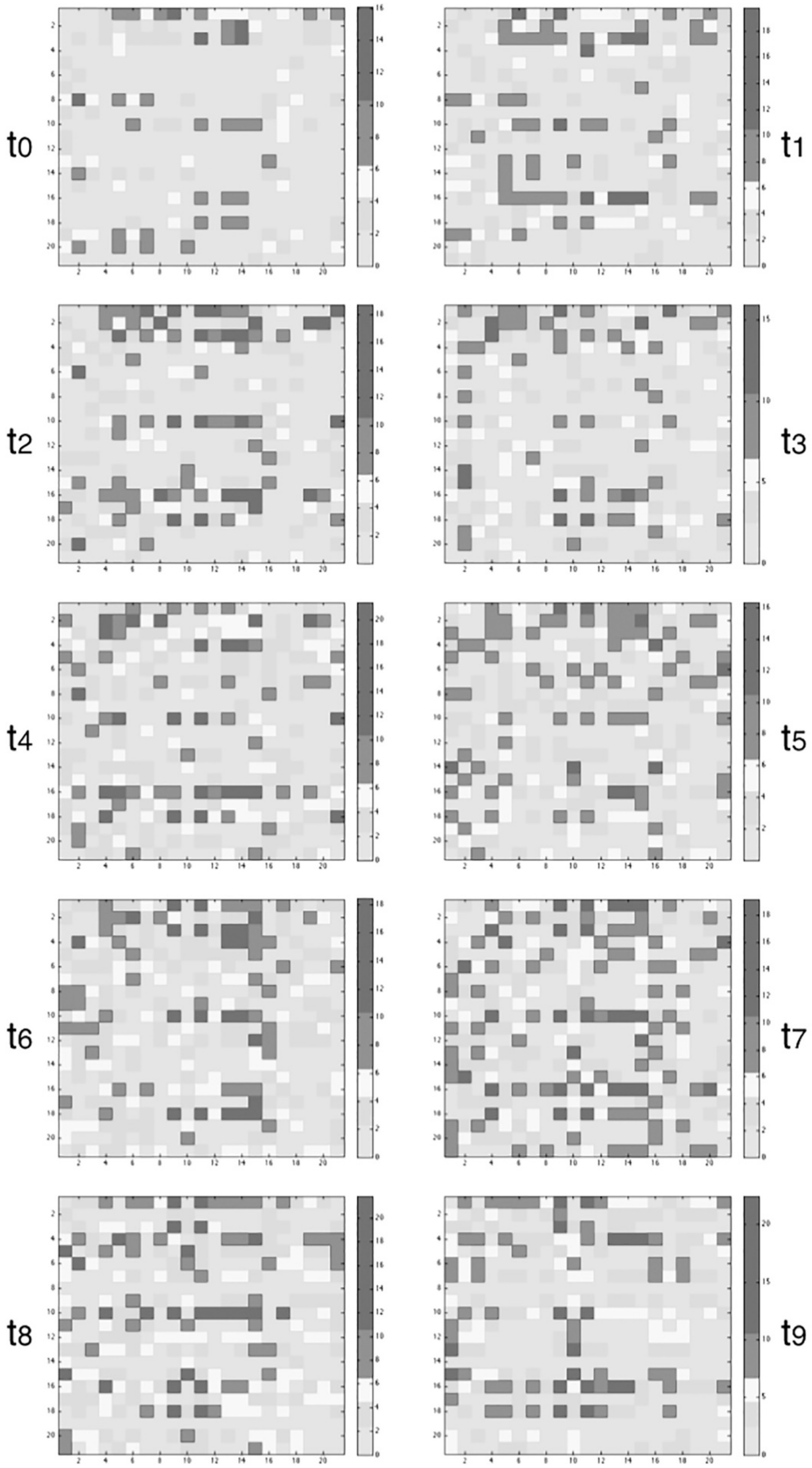


Fig. 9. Coloured differences of travel time between static and dynamic Dijkstra algorithm on different times

6. CONCLUSION

We were able to build a database of historic travel data. The measurement of travel time data is based on the MONICA system. Unfortunately, this data is not public domain. The data was used by a travel agency publishing a real time routing advice on their website. We polled this website to compute the traveling time between the junctions of the highways in the Netherlands. We annotated the sampled data with several parameters such as time of the day, day of the week, season, weather conditions and special incidents. The parameters enabled researchers to filter the data to create homogeneous traffic data. The historic traffic data enabled the design of a dynamic router based on a dynamic version of the well-known Dijkstra shortest path algorithm. In our experiments we found that the saving in travel time, was about 15 % during the rush hours. We have to realize that many traffic jams are caused by incidents or accidents and impossible to predict by historical data. Another problem is that in many cases of traffic jams there are no alternatives to reroute. It also proves that many car drivers are used to regular traffic jams and just join the cue and use the time to communicate or listen to music. But delay of lorries causes enormous economic loss and are rerouted via their central control room. The pick-up and delivery services using solutions of the traveling salesman problem is far from being solved. But the progress in reducing delay of car drivers is promising and additional research is needed.

REFERENCES

[1] Eggenkamp, G., L. Rothkrantz. (**2010**) Intelligent dynamic route planning. KBS&TRAIL Workshop. June 2001.

[2] Tatomir, B., L.J.M. Rothkrantz, A.C. Suson. (**2009**). Travel time prediction for dynamic routing using ant based control. Winter simulation conference, 1069-1078.

[3] Jabbarpour, M.R., A. Jalooli, E. Shaghaghi, R.M. Noor, L. Rothkrantz. (**2014**). Ant-based vehicle congestion avoidance system using vehicular networks. *Engineering Applications of Artificial Intelligence* 36, 303-319.

[4] Radu,A., L Rothkrantz, M Novak. (**2013**) Digital traveller assistant. Informatics in Control, Automation and Robotics, 101-114.

[5] Ehmke, J.F., A. Steinert, D.C. Mattfield. (**2012**) Advanced routing for city logistics service providers based on time-dependent travel times. Journal of Computational Science vol. 3, no. 4, 193-205.

[6] Ritzinger, U., J. Puchinger, R.F.Hartl. (**2016**). A survey on dynamic and stochastic vehicle routi.ng problems. International Journal of Production Research, Francis Taylor, vol. 54, no. 1

[7] Solomon, M., J. Desrosiers. (**1988**). Survey paper—time window constrained routing and scheduling problems. *Transportation Science*. vol. 22, No. 1.

[8] Ling Qiu, Wen-Jing Hsu, Shell-Ying Huang, Han Wang. (**2002**). Scheduling and routing algorithms for AGVs: a survey. *International Journal Prod. Res*. 2002, Vol 40, no. 3, 745-760.

[9] Psaraftis, H., W.Min, C. Kontovas. (**2018**) Dynamic vehicle routing problems: Three decades and counting. Networks, DTU Orbit.

[10] Dijkstra, E.W., (**1959**). “A note on two problems in connexion with graphs,” Numer. Math., vol 1, pages 269-271.