



# Estimation time of arrival for tram vehicles

M. A. Dagli, O. Ozdemir, E. Dede & L. Akcil  
*Istanbul Transportation Corporation, Turkey*

## Abstract

In the public transportation industry remarkable Big Data is formed. The data which are generated from tracking the vehicles shall be used for different kinds of purposes. The arrival time of the vehicles is one of the most desired data for the passengers. In this study, big data which is produced for tracking the vehicles at the Command Centre are used. The T1 tram line of Istanbul City is segmented into 300 parts using RFID sensors and each part is observed for estimation processes. Also, for computing the location accurately, odometer data of the train is gathered. Elapsed time of each train at each part of the line is inspected for the previous 2 hours. Then the average of the elapsed time of each part is calculated. To determine the remaining time of the closest trains, the location of all trains is observed. Finally, according to the distance between the trains and the station, the arrival time of a train is calculated. As test laboratory generated data for each station are displayed on a web environment. For observation of estimation accuracy, estimated times are compared with the real arrival times. As a result, ninety percent accuracy is obtained. This study will generate the data for the information display on the train platforms.

*Keywords: tram tracking, Big Data, estimation of arrival, railway vehicle, RFID, passenger information system.*

## 1 Introduction

After a revolution of sensor technologies and IOT applications, the importance of the Big Data and Analytics is increasing every day. Public Transportation (PT) industry is one of the lead areas for the need of Big Data Analytics. At many parts of PT notable data production occurs. Several of these data are recorded. These recorded data get bigger according to the complexity and intensity of the related PT environment. Recorded big data may not make sense to the operation staff or to the passengers.



Henry [1] listed several of the applications implemented for the PT as

- Travel Demand Modeling;
- Train Signal and Control Systems;
- Route Planning and Scheduling;
- Automated Fare Collection;
- Automated Passenger Counting;
- Automatic Vehicle Location.

For tram line operations, Automatic Vehicle Location is the most important of the mentioned applications. Unlike underground metro systems Tram Operations take place with the road cars and the passengers. Advanced underground metro signaling systems cannot be implemented at those environments. Special automation systems are needed to manage the operations. For managing purposes, precise locations of the vehicles are the most important data for the operation staff. When the location and several data prepared for the Operations Centre there becomes Big Data accumulation. These Big Data can be used at different kinds of purposes after converting them to meaningful data (fig 1).

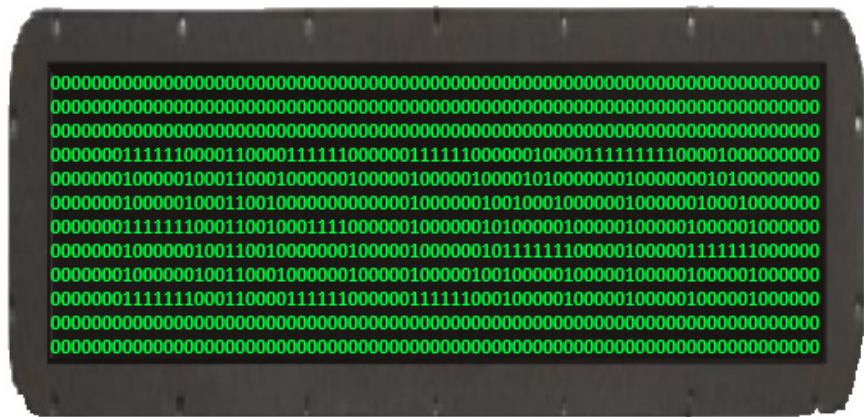


Figure 1: Big Data might be clear after inspection.

Estimation of train arrival (ETA) is one of the products of these data. With this paper, Tram Tracking System for the Istanbul T1 Kabatas-Bagcilar line and the study about ETA for the Passenger Information System is presented.

## 2 Tram tracking system

Istanbul Metropolitan T1 Kabatas–Bagcilar tram line is one of the busiest tram lines around the world. Every day, an approximate ridership figure of 350,000 is achieved on this line [2]. Unlike metro or light rail transit (LRT) lines, tram

vehicles are travelling with the road cars and pedestrians which make automated solutions ineffective. As a result, the Control Center (CC) operators and staff have an important role in ensuring the safety of the whole tram system.

The CC operators need real-time information about the location of the vehicles for an efficient operation. In unusual situations when an accident or a malfunction occurs, the CC requires the whole information of the line in detail to provide reliable and safe operation. A sample screenshot of a CC is given in Fig. 2.



Figure 2: Command Center of T1 Tram Line.

For the Tram Tracking System data acquired from more than 1,500 different points and these data presented to the CC staff as meaningful information. At the interface of the CC staff several information like direction, speed, daily distance covered, coupling, door status of the vehicles are displayed. In addition, the name of the driver, trip number, warnings about vehicle faults are shown.

### 3 RFID-based blocking

Using the 92 low floor tram vehicles; 19 km length line is operated. There are 31 stations, 74 cross section areas for the road vehicles and 32 switching areas. For tracking the important events these areas should be observed.

Therefore, the line segmented into more than 300 parts (fig. 3). At the beginning of those parts there are RFID tags mounted under the concrete surface. The unique data of those tags are read by the antenna mounted on the train. In this

way the section number of the vehicle is obtained. However, the lengths of those sections vary from 50 meters to 400 meters so the location of the vehicle cannot be determined precisely. Only at those points exact location is gathered. Another problem is that, when antenna fails to read a tag there will be a positioning error. At RFID systems this happens occasionally due to the environmental effects like rain, snow and dust.

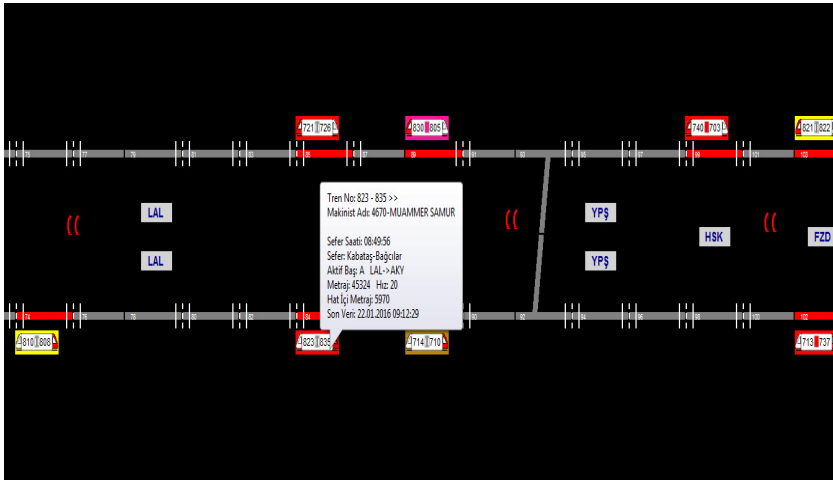


Figure 3: Station, crossroad and switch areas.

Providing more accuracy and reliability to the system additional data is required. For this purpose odometer data can be used. In that system 5 meter accuracy is obtained from the odometer. In this system not only the reader data but also the odometer data is used for positioning. The length of each block is calculated and recorded on the system control unit. Thus a train's 5-meter accurate position data can be calculated by the control unit and if any of the RFID tags fails to be read it is declared by the control unit that train is entered to a new block.

Also, only the odometer data is insufficient too. In some situations, due to sliding of the train location by braking, the data produced by the odometer can be inconsistent. The odometer cannot measure the sliding distance. Besides, the crossing of the vehicle from one rail track to another by using points should also be detected, in order to display the vehicle on the new rail track. This time RFID part of the system rectifies the location data. Also it is not possible to detect the violation of the section speed limits by using the data taken from the control unit, for that purpose speed limit of the sections are recorded on RFID tags.

Additionally, the alternation in the direction of movement plays an important role for the localization of the vehicle. After changing the direction of the movement of the vehicle, the data produced by the odometer becomes insufficient because, the odometer keeps producing data as the vehicle is travelling in its

previous direction. The data acquisition cannot be provided after changing the direction of the movement. To cope with this, the data of which side of the vehicle is being used by the driver is perceived by the control unit. If it is needed to be turn back to opposite direction as completing a trip, driver activates the other side of the vehicle [3].

## 4 Estimated time of arrival

For estimating the arrival times *Moving Average Prediction* algorithm is used. As Pongnumkul *et al.* [4] defined, the inspiration for this algorithm is the assumption that trains normally take longer to run between two stations than duration specified in the schedule due to unforeseen circumstances. Therefore, we can better predict the arrival time at downstream stations using the historical amount of time that the train takes to run between stations. We can predict the arrival times at downstream stations by averaging the historical durations that the same train takes to run from station  $S_p$  to a downstream station  $S_i$  and adding that duration to the last known departure time, which can be calculated using the following formula:

$$PA_i = RD_p + \frac{\sum_{b \in H} (RA_i^b - RD_p^b)}{|H|} \quad (1)$$

where  $RA_i^b$  is the historical arrival time at station  $S_i$  on date  $b$ ,  $RD_p^b$  is the historical departure time at station  $S_p$  on date  $b$ ,  $H$  is the set of historical dates that both  $RA_i^b$  and  $RD_p^b$  are available, and  $|H|$  is the size of the set  $H$  [4].

Different from the study of Pongnumkul *et al.*, at this study within the tram tracking system tram line divided into *parts*. For the estimation of the train arrivals, data which are acquired “when the tram’s location changes between the parts” are used. Arrival time of a train to a destination station is calculated by “adding the estimated elapsed time of each part between the destination and the current location of the train”. Estimation of elapsed time of each line part is calculated by arithmetic average of the “total elapsed time of the trains which are passed through this related part” during the last 2 hours.

For finding the closest 2 trains to each station, metric location value of each train and station are obtained. Then according to the direction and the location of the train, first two of the closest trains to the related station selected. After finding the first two closest trains, the line parts between the train location part and station location part are inspected. Then estimated elapsed times of all parts are summed up.

Estimated arrival times are compared with the actual arrival times. In a sample station named “Yusufpaşa” every day more than 300 train journeys occur. All of these trains’ arrival times are recorded. And all of the estimated arrival times for each train are listed. Then time difference between the actual and estimated value is calculated. The difference is compared with the estimated duration to the station and percental success rate is found.

In the graph it is seen that, according to the day’s different hour periods success rates are changing. At the rush hours of this tram line (07:00–09:00 and 17:00–

19:00) success rate is low than other hours and success rate is seen more at the below of the average value line (fig. 4). Also, extraordinary states like abnormal dwelling time (ex. 06:24:41 or 13:17:21) of a train at the previous stations decrease the success rate.

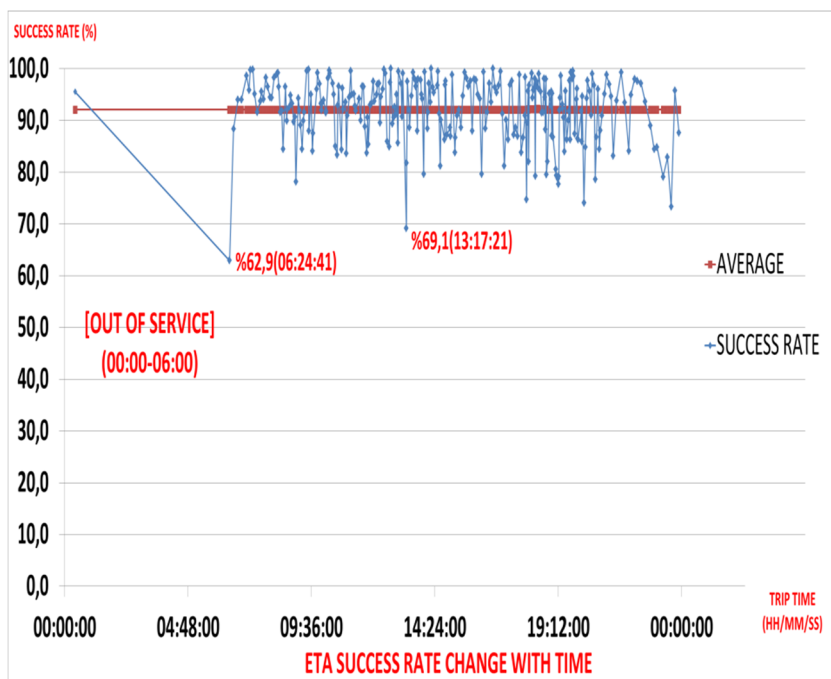


Figure 4: Percent success rate of the estimation according to time.

## 5 Conclusion

For tracking the tram vehicles, location and much other information sent from the vehicles. Everyday nearly one million raw data received and recorded at the database servers. Using the recorded data acquired from the trams, time of arrival has been estimated. According to the tests, 92% success is achieved in average. Estimated data is prepared for the data generation of the Passenger Information Systems. For every station platform there will be an information display. For verifying the generated arrival times, a web page built as test environment. In this page trip name, remaining distance and estimated time of arrival is displaying (fig. 5).

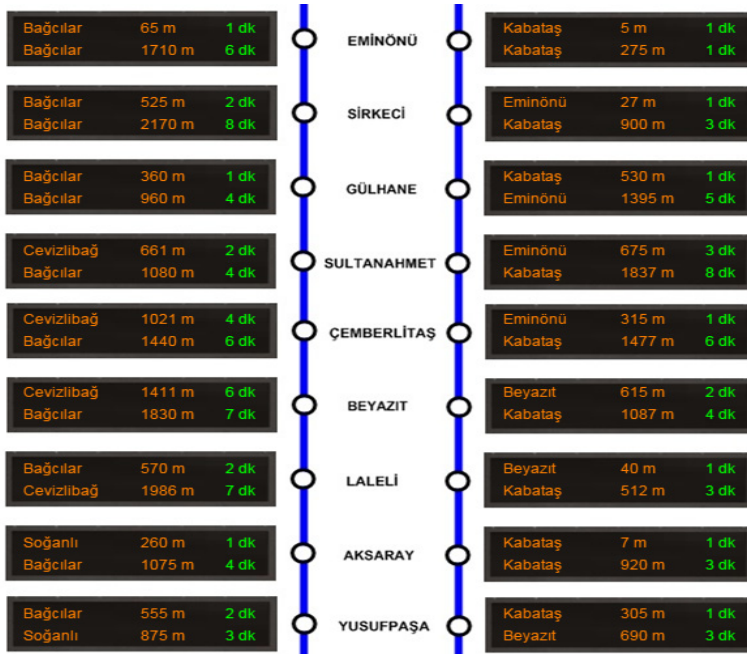


Figure 5: Estimated data for each station.

## References

- [1] Henry, L. (2013). Analytics and Big Data – Rail Public Transportation is a Leader *Rail Conference of the American Public Transportation Association*, Philadelphia (APTA).
- [2] Istanbul Transportation Company Annual Operation Report. (2015). URL: <http://www.istanbul-ulasim.com.tr> (Reached on: 04.01.2016).
- [3] M. A. Dagli, K. Ozmal, S. Saygin, M. Kurtkaya, L. Akcil (2014). Positioning of tram vehicles using RFID system. In: 10th ITS European Congress, Helsinki, Finland.
- [4] S. Pongnumkul, T. Pechprasarn, N. Kunaseth, K. Chaipah (2014). Improving Arrival Time Prediction of Thailand's Passenger Trains Using Historical Travel Times. In: 11th International Joint Conference on Computer Science and Software Engineering (JCSSE).