IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper | Volume-5, Issue-11 | E-ISSN: 2347-2693

# Public Transport Tracking and its Issues

**Jitendra Oza[1], Zunnun Narmawala[2], Sudeep Tanwar[3], Pradeep Kr Singh[4*]**

[1] Institute of Technology, Nirma University, Ahmedabad, India
[2] Institute of Technology, Nirma University, Ahmedabad, India
[3] Institute of Technology, Nirma University, Ahmedabad, India
[4*]Department of CSE, Jaypee University of Information Technology, Solan, India

**Corresponding Author: pradeep_84cs@yahoo.com*

**Available online at: www.ijcseonline.org**

Received: 17/Oct/2017, Revised: 25/Oct/2017, Accepted: 11/Nov/2017, Published: 30/Nov/2017

***Abstract-*** Public transport is a fast and convenient way of travel, but there are many issues related to it. Challenges in current public transport system are: how to estimate the exact arrival time of vehicle and real tracking of vehicle. Solution of these two problems directly save the user time and provide better management for scheduling of vehicles. Many proposal exist in the literature to address above mentioned issues. Keeping the need of intelligent transportation system, this paper provides comparative analysis of all the state-of-art existing proposals. Tracking the vehicles generally takes two types of data: historical, and real time data. For real time tracking of vehicles, Global Positioning System (GPS), sensors, Internet of Things (IoT) devices, etc are used. Due to generation of huge amount of data from IoT enabled devices present in transport system, kalman filtering, artificial neural network, data analytics and machine learning are also used for better scheduling of vehicles. In last section we provide the open issues and challenges that needs to be taken care while designing the Intelligent Transport System (ITS).

***Keywords-*** Vehicle Tracking, ITS, GPS, Smart City, Historical data, Real time data, Sensor, IoT

## I. Introduction

The public transport system is cheap and frequent compared to the private transport system, but it is not very popular due to its availability issues. People don't know what is the current location of the public transport and when a public transport will come at a particular bus stop. We can make the public transport system more convenient and popular if we can solve this problem. We have some systems that tries to solve this problem but prediction and accuracy of these systems are not very reliable.

The problem with the current transport system is that they don't provide us accurate data of vehicles current location and arrival time. This makes the system inefficient because the user doesn't know how much time he/she has to wait for the bus to come to the particular stop. Also, people seating in the public transport don't know how much time it will take to reach their destination or what is the next stop. In addition, it also makes the scheduling task difficult as we don't know that the bus is on time or it is running late.

We have to consider different parameters for accurate prediction of arrival time. Please see Figure 1 for time distribution of public transport. Around 5% time from total time depicts in congestion, 20% at traffic signals, 23% on dwell, the time required for a passenger to get up and down from the bus. So for accurate prediction of the arrival time of public transport, we have to handle all parameters affecting the prediction.

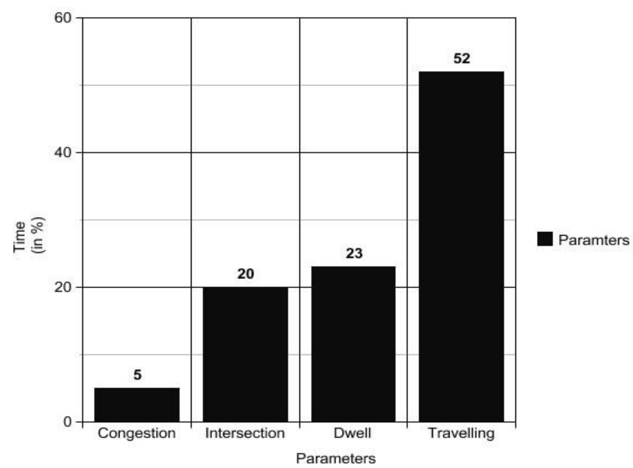


**Figure 1.** Time Distribution of public transport from one stop to next stop [1]

Many researchers have used different methods like historical data, real time data, time series analysis, different machine learning techniques, statistical method and model based methods. Historical data approach is used when we use the previous bus data to forecast arrival time. This type of methods is used when there is steady traffic condition and no external parameter is used for prediction. These methods are not useful when there is a variation from previously collected data. In real time data, we suppose there is a condition in the next time interval is same as in the current time interval. Real time method is not useful when there is a delay in data receiving. The time series is similar to real time. In this, we assume that previous analysis will be same for next time period. Other approaches are Artificial Neural Network (ANN), Support Vector Machines (SVM) and Kalman filter.

## II. LITERATURE SURVEY

Many researchers have given their own solution for the current problem in the transportation system. Many papers have been published on this topic and people are taking interest to implement these concepts. Some of the papers have been implemented successfully. But this is a wide topic and for the different scenario we have a different solution. Also, new technologies play an important part here. With latest technologies, we can improve old concepts or we can make our own. In this survey, we have studied different concepts that have been used by different researchers.

**Historical Data**

Zhou *et al.* [2] have implemented bus arrival time based on smart card data (SCD). They have collected one week data, including SCD and actual arrival time data. From total 1011 dataset, they have used 446 as training dataset and 565 as a testing dataset. For reducing the error in smart card prediction time, they have included the difference between a smart card swiping time and actual arrival time, swiping time density, seating capacity. Distribution of card swiping time is used for bus arrival time calculation. They have implemented a model that estimate time using time lag between two consecutive SCD and the last card swiping time at different stations. Proposed algorithm has around 10% of error rate. For more accurate prediction vehicle mode, load factor, scheduling can also be considered as parameters.

Bai *et al.* [3] have implemented a hybrid model. A hybrid model for predicting the base travel time from the historical data they got from the field. Firstly, they have used the well trained SVM model to predict the base travel time from the historical data they got from the field. Then, they have used the Kalman filtering algorithm to improve the accuracy of travel that we predict using SVM model. Prediction model first takes the previously collected data and give it to SVM. SVM then calculates the base arrival time from the data given. This data is sent to Kalman filtering algorithm where data received from svm and real time data are used to predict arrival times of the bus. The main focus of the paper was on bus route sharing the same road and dynamic travel time of buses with different routes.

Yu *et al.* [4] have implemented a dynamic model that uses the historical data for the prediction of arrival time. GPS data are collected at regular interval. GPS fitted in the base will send data every 15 or 20 seconds. They made a prediction model based on cluster analysis and polynomial fitting. They have used the last 1 year data for the testing of system. To prevent effect of different patterns on prediction accuracy, they have used the euclidean distance with heretical cluster analysis for nearest point. They have also used the hierarchical clustering method and Euclidean distance method because of its low error rate. Prediction model takes the historical data and processes that to save data in a historical database. After that prediction model is called to check for the pattern matching and estimate arrival time.

Li *et al.* [5] have implemented a linear model for the prediction. They divided the total time into two parts: linear part and residual. For the prediction they have included traffic condition, dwell time, intersection and departure time. Parameter calibration process is used for prediction. First, they classify and determine traffic patterns from the historical data. In the second step, estimation of dwell time, travel time and service time are determined. In the last step, least square method is used for parameter optimization. They used the historical data and classify eight different cluster, a statistical approach was developed for identification of different parameters to be included in the cluster. They also built web application to verify practicability and efficiency.

Yu *et al.* [6] have implemented model that predicts the bus arrival time of same stop sharing the different route. For the implementation, they have used the dynamic model which consists of two parts: SVM model for estimating base line travel time and Kalman filter for adjusting the base line with latest information. Predictions are measured using mean absolute error, mean absolute percentage error and root mean square error. For the calculation of the actual arrival time and running time, manual method is used. They have compared the performance of four different methods, in which Linear Regression (LR) gives the worst performance. KNN is better than LR. KNN and ANN are almost same but ANN is slightly better than KNN. SVM gives the best performance among all. The correlation coefficient (r) of four methods, namely LR, KNN, ANN and SVM is 0.84, 0.85, 0.87, and 0.90.

Chien *et al.* [7] have implemented the ANN prediction model. They have used the two different ANN, link-based ANN and stop-based ANN. For the model development, they have used the fully connected multilayer feed-forward network with the Back Propagation (BP) learning algorithm. They have used the Enhanced ANN to improve the performance. While the algorithm is predicting the arrival time using the training data, it also considers real time data. To improve the accuracy, they have integrated both models with an adaptive algorithm. For the implementation of the algorithm, they have used the CORSIM simulation model. For the data collection, they have used the RFID detector to detect the particular vehicle and also for the unique identification. The main advantage of this method is that, we can predict the single stop as well as multi stop prediction because of the two different ANN methods are used in the implementation.

**Real Time Data**

Weng *et al.* [8] have implemented bus speed estimation model. The prediction model is divided into three parts: Data processing & preprocessing, Data matching & travel speed estimation and correction of travel speed. In the first phase, they match the GPS and Geographical Information System (GIS) map data to estimate the location of the vehicle and calculation of vehicle from one point to another to generate the proposition arc. In second phase, this movement arc is diagnosed and missing section is added. In the third phase, estimated calculation of bus arrival time performed. In the fourth phase, direction of the bus is determined. Fifth phase is calculation of travel speed between two points. Then field data and model data is compared and it provides 88.4% accuracy.

Liu *et al.* [9] have implemented a framework called WiLocator. They proposed tool named Signal Voronoi Diagram (SVD) that uses the WiFi access points (APs). WiLocator has three components: WiFi enabled smartphones, backend server and user interface. WiFi enabled smartphones are used for the sending WiFi data to backend server and WiFi sensing. Backend server is used for data storing, processing and prediction. User interface (UI) provides communication between WiLocator and users. WiLocator consists of three parts: SVD based bus position, arrival time prediction and traffic map generation. In the second part, they match GPS data and calculate proposition of position. After that estimation of bus speed is calculated. In the last part, if threshold of estimated speed are not reached, then travel speed is corrected.

Zaki [10] have implemented hybrid neural network for the prediction of public transport arrival time. In their proposed system, they implemented two models: ANN based and Kalman filter based. Proposed algorithm first receives data from hardware module then processes data and store in database. After that it checks for rule matching and if any rules are matches ANN and predict the arrival time as it uses a Kalman filter to calculate arrival time. In proposed neural network model, there are four layers: input layer, two hidden layers and output layer. There are seven nodes in proposed input layer. First hidden layer has ten nodes and the second hidden layer has 3 nodes. The input layer has one node. A modified Kalman filter is used for prediction. Last three similar days in last three similar weeks historical data is used for prediction.

Zhou *et al.* [11] have implemented mobile phone based participatory system for the prediction of the arrival time. System architecture divided into 3 components. The user will convey information about bus stop and interested route. In other side, a user who is seated in bus share data like tower IDs. But first, they have to decide whether the user is on the bus or not. After that, this data is sent to backend server. Backend server is responsible for all computation. Here data received from the user mobile is processed and query of user is answered. They have created an application that the user is required to install for the participation. After that, user will share its bus information with the server using this application. This system is totally dependent upon the participants and independent from other parameters.

Zhu *et al.* [12] have implemented a model for real time bus arrival time prediction. They have created two different models for prediction: point based and path based. For the calculation of point based prediction, instantaneous velocity (V), GPS information and distance (L) is used for calculation of arrival time. For the prediction of path based arrival time, they have used running time of the bus, delay at bus stops and intersection. They have calculated the dwell time separately. They have also used the Mean Absolute Percentage Error. Also, they have compared both models. Point based prediction model provides minimum error rate and is more reliable and accurate compare to path based.

Their proposed model predicts better because of the real time data adoption, which is not available in the traditional model.

Ferris *et al.* [13] have implemented real time public transport tracking system. This paper provides the real time arrival info about Seattle-area bus system. In this, they have created an app and the website for the user to get the bus information as well as for the queries related to the bus trip and bus facilities. They have integrated maps for showing accurate location. They have implemented a system that provides a number of service modules and integrated it using Spring inversion of control framework, a relation database that is handled by hibernate framework. System is able to provide the real time data about bus location and also estimated arrival time of bus at particular station.

**Hybrid Model**

Maiti *et al.* [14] have implemented historical data based bus arrival system. They claim that their system is much faster than ANN and SVM. In their proposed model. they first take previously collected data then analysis and pre-process data. At this stage, they try to remove noise, missing values, and bus stops correction. After that, they take this refined data and predict bus arrival time. Arrival time prediction procedure consists location wise speed calculation and time slot wise speed calculation. Also, for data collection, they have used the Connect-Port X5 R devices. The sensor sends speed, location, time stamp, card swiping time stamp, and passengers boarding and getting down. The sensor is cheap and lightweight. GPS enabled mobiles phone can also be used as replacement of sensors. Proposed model shows better result than standard ANN and SVM used in the paper for comparison.

Gong *et al.* [15] have implemented a hybrid model for the bus arrival time. For the prediction, they have selected both historical and real time data. Real time data can't be directly used. So they preprocessed data so that it can be used in the prediction model. They have used the Moving Average Model (MAM), Moving Average Dynamic Adjustment Model (MADA) and Hybrid method based on Hybrid Moving Average and Dynamic Adjustment (HMADA) model for prediction of arrival time. They suggest that HMADA can make up the shortening of MAM and MADA. Also, they suggest that GPS data must be preprocessed before it can be used for the prediction model directly. Also, they stressed the importance of upper bound of station in the prediction.

Padmanaban *et al.* [16] have implemented real arrival time prediction system. This system used the historical data methodology for the prediction. For data collection, they have taken route no. 21-L in Chennai. They have collected one week data of two public transport. They have collected latitude, longitude and time stamp from GPS fitted in a bus. After that Haversine formula is used to calculate distance. Then, they have developed an automated procedure for calculating delay at bus stops. Then, they have used their algorithm to predict the arrival time of the third bus with the use of the previously collected data. For prediction, they have divided travel into two different parts: running time and delay time. They have also included the different types of delay like traffic signal delay, bus stop delay, etc. The have also used the Kalman filtering algorithm for better accuracy. They have merged the historical data with the real time data to improve the performance of the prediction algorithm.

Chien *et al.* [17] have implemented two prediction model: historical data model and real time data model. They have collected data from the New York City Thruway. For the data collection, they have used the road side terminals. To improve data they have eliminated result of a particular vehicle if data of vehicle highly deviates from the average vehicle data in a particular time period. They process and filter few data as per the prediction algorithm requirement. Two prediction methods are tested in this, path based and link based. They have also used the Kalman filtering algorithm for better accuracy. For the implementation, they have used the improved version of the CORSIM simulation model. They have not used the real time data directly as it needs to be refined to make it appropriate for use in implementation. From the experiments they have performed, they conclude that the link based prediction is accurate and reliable compare to path based. Path based is accurate when there is uniform traffic condition and path is short.

Table 1 represents the summary of whole survey. In the table there are four column. First column is a name of author, second is what algorithm/method they have used for implementation, third is what is the strength of paper and fourth one is what can be added/improved in paper.

Table 1: Literature Survey Summary

| **Authors** | **Methodology** | **Pros** | **Cons** |
|---|---|---|---|
| Zhou et al. [2] | Frequency distribution, regression | Smart card data is used efficiently for calculation of arrival time. | Require appropriate infrastructure for better result, which is not easy. |
| Bai et al. [3] | SVM | Used the SVM and Kalman filter to improve the accuracy of algorithm. | Requires large data and more training time for SVM. |
| Yu et al. [4] | Historical Data | Well-designed model that does not required extra computation time to access real time data. | Prediction model does not consider delay at traffic signals and intersection. This can reduce the performance of algorithm. |
| Li et al. [5] | Statistical approach, Historical data | They divided data into two parts for better analysis. | Algorithm can be improved if more parameters are used and historical data can be used more efficiently. |
| Yu et al. [6] | SVM, kNN, ANN, LR | They have implemented four models and conclude that the SVM gives better results than other three. | Performance can be improved if we use more parameters for prediction. |
| Chien et al. [7] | Artificial Neural Network (ANN) | Link based and Path based enhanced ANN is used for prediction. | Prediction algorithm requires large training time as well as data for the accurate prediction. ANN makes system slow and complicated. |
| Weng et al. [8] | Real time data | They use the GPS and bus Geographical Information System (GIS) map to get accurate location of bus. | More parameters can be included to improve the performance of overall system. |
| Liu et al. [9] | Real time, passenger participation | Use of WiFi hotspot for prediction and communication. | Not applicable when there is no WiFi nearby and requires passenger participation. |

| | | | |
|---|---|---|---|
| Zaki [10] et al. | ANN, Kalman filter | Use of hybrid model improves the accuracy of algorithm. | Algorithm requires lot of data and time for the prediction. |
| Zhou [11] et al. | Passenger participation | Use of mobile phones for collecting and sending data makes system efficient and easy. | Totally dependent on participants data which makes system less reliable. |
| Zhu [12] et al. | Real time data | Separates calculation for the delay at bus stops and intersection. | Real time method is not applicable where there is delay or loss of data. |
| Ferris [13] et al | Real time data | User can access real time location of vehicle using application, website, SMS, IVR. | Proposed system requires huge resources and infrastructure as real time data cannot be used directly. |
| Maiti [14] et al. | Historical data | They used the sensors fitted in buses and used limited parameters to make prediction model less complex and more accurate. | Lots of historical data is required for accurate prediction and not applicable where situation changes quickly. |
| Gong [15] et al. | Historical Data,Real time data | Three models are used for prediction. | Model uses average speed which is not efficient. |
| Padmanaban et al. [16] | Historical Data based Approach | Consider delay separately for better result | Based on Historical data which is not accurate when there is delay in system and required more data for accurate prediction |
| Chien et al. [17] | Kalman Filtering Algorithm | Road side terminals (RST) and E-ZPass system is used for data collection. | Real time data cannot be used directly for prediction and not applicable on heterogeneous traffic condition. |

## III. OPEN ISSUES

Much work has been done in public transport arrival time prediction. But, still there are issues that have been not solved properly. To get accurate location, we have to use proper methodologies and appropriate tools/hardware to get accurate location of public transport. For the prediction of arrival time, we have to consider parameters like: prediction algorithm and data. Selection of prediction algorithm is one of the major factor. The whole system relies on this algorithm. Type of data also plays an important role. Different data provides different parameters and we have to choose data according to our algorithm. Data filtration/clustering techniques are useful when we have too much data to process. Another issue with prediction algorithm is efficiency and accuracy of the algorithm and feasibility of the algorithm in real world. Also, we can merge two or more methods and check whether the performance improves or not.

## IV. CONCLUSION

Although many researchers have given their solution to solve the problem of public transport arrival time prediction, there are still many issues left. Public transport arrival time prediction is still open issue. Different researchers have used different techniques to solve these issues. We have mentioned these issues and what approaches has been used by researchers to solve these issues and what is the problem with their solution. We have also discussed what we can do to improve arrival time prediction algorithm.

## REFERENCES


[1]. “Brt (bus rapid transit). metro rapid.,” Accessed October 1, 2000.

[2]. Y. Zhou, L. Yao, Y. Chen, Y. Gong, and J. Lai, “Bus arrival time calculation model based on smart card data,” *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 81–96, 2017.

[3]. C. Bai, Z.-R. Peng, Q.-C. Lu, and J. Sun, “Dynamic bus travel time prediction models on road with multiple bus routes,” *Computational intelligence and neuroscience*, vol. 2015, p. 63, 2015.

[4]. H. Yu, R. Xiao, Y. Du, and Z. He, “A bus-arrival time prediction model based on historical traffic patterns,” in *Computer Sciences and Applications (CSA), 2013 International Conference on*, pp. 345–349, IEEE, 2013.

[5]. F. Li, Y. Yu, H. Lin, and W. Min, “Public bus arrival time prediction based on traffic information management system,” in *Service Operations, Logistics, and Informatics (SOLI), 2011 IEEE International Conference on*, pp. 336–341, IEEE, 2011.

[6]. B. Yu, W. H. Lam, and M. L. Tam, “Bus arrival time prediction at bus stop with multiple routes,” *Transportation Research Part C: Emerging Technologies*, vol. 19 , no. 6, pp. 1157–1170, 2011.

[7]. S. I.-J. Chien, Y. Ding, and C. Wei, “Dynamic bus arrival time prediction with artificial neural networks,” *Journal of Transportation Engineering*, vol. 128, no. 5 , pp. 429–438, 2002.

[8]. J. Weng, C. Wang, H. Huang, Y. Wang, and L. Zhang, “Real-time bus travel speed estimation model based on bus gps data,” *Advances in Mechanical Engineering*, vol. 8, no. 11, p. 1687814016678162, 2016.

[9]. W. Liu, J. Liu, H. Jiang, B. Xu, H. Lin, G. Jiang, and J. Xing, “Wilocator: Wifisensing based real-time bus tracking and arrival time prediction in urban environments,” in *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on*, pp. 529–538, IEEE, 2016.

[10]. M. Zaki, I. Ashour, M. Zorkany, and B. Hesham, “Online bus arrival time prediction using hybrid neural network and kalman filter techniques,” *International Journal of Modern Engineering Research*, vol. 3, no. 4, pp. 2035–2041, 2013.

[11]. P. Zhou, Y. Zheng, and M. Li, “How long to wait?: predicting bus arrival time with mobile phone based participatory sensing,” in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 379–392, ACM, 2012.

[12]. T. Zhu, F. Ma, T. Ma, and C. Li, “The prediction of bus arrival time using global positioning system data and dynamic traffic

information," in *Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP*, pp. 1–5, IEEE, 2011.

[13]. B. Ferris, K. Watkins, and A. Borning, "Onebusaway: results from providing realtime arrival information for public transit," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1807–1816, ACM, 2010.

[14]. S. Maiti, A. Pal, A. Pal, T. Chattopadhyay, and A. Mukherjee, "Historical data based real time prediction of vehicle arrival time," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 1837–1842 , IEEE, 2014.

[15]. J. Gong, M. Liu, and S. Zhang, "Hybrid dynamic prediction model of bus arrival time based on weighted of historical and real-time gps data," in *Control and Decision Conference (CCDC), 2013 25th Chinese*, pp. 972–976, IEEE, 2013.

[16]. R. Padmanaban, K. Divakar, L. Vanajakshi, and S. Subramanian, "Development of a real-time bus arrival prediction system for indian traffic conditions," *IET intelligent transport systems*, vol. 4, no. 3, pp. 189–200, 2010.

[17]. S. I.-J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with realtime and historic data," *Journal of transportation engineering*, vol. 129, no. 6 , pp. 608–616, 2003.

 