GPS based Public Transport Arrival Time Prediction

Muhammad Umar Farooq, Aamna Shakoor, Abu Bakar Siddique
Department of Electrical Engineering, University of Engineering & Technology, Lahore
mufarooq40@gmail.com

***Abstract*—Users of public transport, in developing countries, face a lot of problems such as long waits on bus stops and often no bus arrival after a long wait. One of the novel solutions to the problem is to display the expected arrival times of the buses on respective bus stops. Many prediction systems have been developed based on different techniques.This paper presents a public transport arrival time prediction system relying on real time Automatic Vehicle Location (AVL) data rather than the old-fashioned techniques. Global Positioning System (GPS) and Global System for Mobile Communications (GSM) are used for developing a Real-Time Locating System (RTLS) and communication between different nodes respectively. On server side, an efficient and optimized algorithm is developed to predict the arrival times accurately. Many test drives were done on a selected route under various test conditions. And resulting predicted times and their deviation from Final Travelling Time (FTT) at each bus stop are shown in graphs at the last of the paper.**

***Index Terms*—Automatic Vehicle Location, Intelligent Traveling System, RTLS, GPS, GSM, IoV**

I. Introduction

Rapidly increasing number of vehicles, poor roads infrastructure, mismanagement of public transport companies and worst conditioned vehicles are causing huge problems for citizens in developing countries. These problems include arrival delays, congestion, extra fuel consumptions and wear and tear of vehicles due to frequent acceleration and breaking sequences.

A case study was done on Lahore, the second most densely populated city of Pakistan, which is the capital of Punjab (most populated province of Pakistan) and has an airport, railway station and a dry port as well. Though the city is enjoying state of art transportation services in some of its areas, but present public transport system is causing extreme hardship and stress for the people of Lahore in their day to day lives. Rapid increase in population, average distance traveled by the commuters and no check and balance for maintenance of the existing public transportation has pushed the transport system to the deterioration.

There are around 700 large size buses operating in Lahore [1]. Besides the insufficiency of this fleet, the buses are old and poorly maintained as well. Most of the time, these buses are off road for repairs. On the other hand, public transport is also being provided by overloaded, poorly managed, and rashly driven mini-buses and vans. Such things cause ill-timed transport availability on bus stops. It calls for replacement of this informal and dangerous transport service medium with properly timed, organized, safe, efficient, and affordable bus transport service. Due to several reasons, principal among them is the non-existence of a proper mechanism for financing the buses and bus service monitoring, the number of buses has been decreasing over the years and the quality of service provided is also getting poorer day by day. Due to these reasons, transport companies are losing the trust of commuters. Long queues on bus stops cause wastage of excessive time which may drive away the travelers and make them reluctant to take buses.

We did a public survey regarding waiting time for buses and arriving times of public transport on bus stops. The survey covered people from every field of life such as students, office workers, businessmen and old citizens with an age group of 18 to 70 years. Main goal of the survey was to judge the intensity of picked problem. About a diversity of 100 people took part in the survey. Out of which, 91% people had to experience long irritating wait on bus stops in Lahore. While 82% had to travel by some other sources due to no bus arrival after a long wait on a bus stop. This shows the severe-ness of the problem regarding arrival times of the public transport.

Such pathetic circumstances demand for a reliable and state of the art system for advancement of public transport to overcome the discussed problem. So, in this research work, an accurate real time system is presented for predicting public transport arrival times on bus stops, technically named as GPS based Public Transport Arrival Time Prediction (PTATP). According to our public survey, all the respondents nodded in the favor of our proposed solution. About 69% people declared it as an excellent solution of the problem while 31% considered it an appropriate solution. Nobody rejected or gave vague statement about the solution.

II. LITERATURE REVIEW

Most local transport companies provide their fleet information through timetable (based on their designed schedule). However, this provides very limited information such as service hours, departure, and arrival times of buses. It does not provide any real-time information to solve the discussed problem.

The developed systems cover various techniques such as historical data, univariate and multivariate traffic modeling, participatory sensing, link-based and path-based prediction, regression techniques, Kalman Filtering (KF) and Automatic Vehicle Location (AVL). Traffic time prediction has been a field of interest for a long time. Initially, the Intelligent Traveling System (ITS) systems were developed based on historical data [2], [3]. Using this technique, the time is predicted based on the data collected over a specified period.

Fig. 1. Some already implemented solutions abroad

The time, taken by different busses on a specified route, is collected over a specified time under different conditions. Then the system uses this collected data and predicts the time for running buses. We can use different speeds for different time intervals based on our historical data [4], [5]. The new data is added into the historical data set to make it updated. Though the system holds good if the conditions of the route remains same as in the collected data. However, in varying conditions, such techniques yield very poor accuracy.

In univariate and multivariate systems, a traffic flow model must be developed for a route. And prediction is made based on that traffic model [6]. Research has shown that multivariate data produces better results than univariate [7]. However, in both cases, route traffic model is must for applying prediction system. Moreover, it still has lower efficiency than the realtime techniques.

In participatory sensing networks, the system relies neither on real-time techniques nor the traffic modeling and historical data. Rather the vital component of the participatory sensing is a man. Some volunteer users, travelling in each bus, are the backbone of the system as the central system detects the location of a bus based on the cellular network signals of those persons [8]. However, this technique is too risky to implement as this involves the human participation that can cause significant human error. Moreover, the cellular network location is not accurate enough to be applied for efficient systems.

So, the better method is to do this task using Kalman Filtering (KF) [9], [10]. An KF based model has been developed to predict time using link travel time (collected from AVL data) and automatic passenger counter (APC) system [10]. Kalman filter was applied to refine the coordinates taken by GPS. However, now a days GPS module come with implemented Kalman Filter and gives very accurate and precise location coordinates.

Most modern and state of the art technology is to use Global Positioning System (GPS) for real-time monitoring of the vehicles. Now a day, smart GPS modules are readily available in the market and can be used to get the exact location of the vehicles along with their speeds, direction of motion and many other parameters. Real-time data can be collected with a very high accuracy and precision. Thats why a number of systems have been developed for public transport arrival time prediction based on GPS data [11], [12], [4].

Some worldwide implemented solutions have been shown in Fig. 1. But no such steps have been initiated in developing countries like Pakistan. So in this paper,a real-time GPS based intelligent public transit arrival time is developed for developing countries.

However, in Pakistan, there are mainly two existing systems. First one is the initiative taken by Lahore Transport Company (LTC). LTC initiated an android application (with the name of Bus Da Pata) [13] in order to provide arrival time of buses on the location of smart phone. The app used GPS of smart phone and provided the times. However, the app was not appreciated by the citizens and went flopped badly due to highly illogical and inaccurate predictions. Most of the users has condemned this application (Users' reviews can be seen here [13]).

Suppose even if the app works accurately, the facility can be availed only by the people with smart phones and connectivity that is not much common in Pakistan. Poor and aged citizens who haven't capability to have a smartphone or use it, can never avail this facility. And most of the public transport users come from these classes. So, there is a need to develop a system that can be enjoyed by both poor and rich equally. That's why we developed a system that shows bus arrival time on bus stop display that can be seen by anyone on bus stop. Poor and old citizens have equal opportunity to avail the facility of PTATP system.

The second alternative system of PTATP is Google Maps. Google Maps provides Direction feature giving information about possible routes, distance and expected times from one location to the another. However, Google shows this time based on historical data. Secondly, the issue with this facility is same as with previous one. It needs a smart phone with GPS, Google Apps, and connectivity. And all these facilities are not common in Pakistan especially availablity of connectivity on public places is a big issue for smart phone users. Data Connection packages offered by ISPs are extremely expensive and Wi-Fi is not available on public places. Moreover, more than half of the population of Pakistan lives below the line of poverty and can't have precious smartphones to avail these facilities.

So, in such cases, PTATP is the best solution of arrival time prediction on bus stops. It does not require any connectivity and smart phones. Public and poor citizens can enjoy this free of cost.

III. Proposed Architecture

The framework of our proposed solution consists of three major components such as;

- In-vehicle device; that is mounted on each vehicle in system and gets vehicles location and communicates with server.
- A server; that receives data from in-vehicle device, predicts time based on time prediction model and send it to respective bus stops.
- Bus Stop Device; that receives time data from server and shows it on bus stop LCD display.

The flow of our proposed system has been shown in Fig. 2.

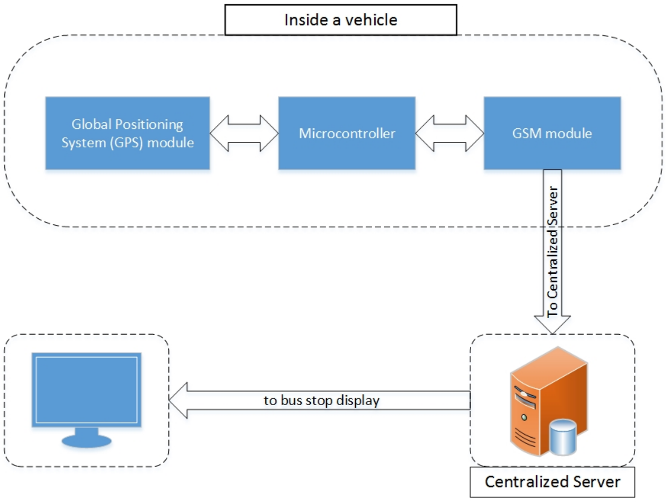

Fig. 2. Top view of PTATP

A. In-vehicle Device

An in-vehicle device consists of a GPS receiver, GSM modem and a microcontroller (we have used ATmega 328 based Arduino UNO). Microcontroller is the heart of the device as it interfaces GPS receiver with GSM modem. The device operates on 12V DC and is powered by readily available vehicles' power. GPS sends its data in National Marine Electronics Association (NMEA) sentences [14], of which we select Vehicle Data over Ground (VTG) that is encoded in $GPVTG$ NMEA sentence. Arduino receives this decoded serial data from GPS receiver and translates it to coordinates and speed of the vehicle. This data is then formatted in to a proper text message containing this decoded data along with the vehicle ID. GSM modems use ATtention (AT) Commands to communicate with each other. AT commands for TCP/IP and their explanations can be found in [15]. These messages are sent to server successively with a time delay of 10 seconds keeping very high resolution of system. All the information in the message is kept comma separated to seperate out easily on server side.

B. Server

It consists of a GSM modem and a microcontroller along with a server computer. A dedicated machine with uninterrupted connectivity must be installed as server to make system efficient (A core i3 system was enough to serve for our institution fleet of about 20 buses). The microcontroller sets GSM into recieveing mode and interpretes the data recieved on it. This message is then trailed to server computer serially. Server separates out the data containing time stamp, coordinates, speed, and vehicle ID. Backend of server interface is developed in Personal Home Page (PHP) language and Java script. It establishes a connection with a Structured Query Language (SQL) database and saves all the data in a database relation. This saved data is used to develop our prediction model as described in next section. Moreover, the server calls coordinate of all bus stops of the respective route from another relation of the database. Google maps Application Program Interface (API) then calculates the distance of bus from all bus stops. Based on this distance and time predictor model, an estimated time for bus to reach all bus stops on its route is calculated. This calculated time is then transmitted to respective bus stops using GSM.

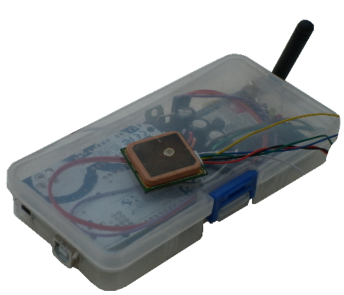

Fig. 3. In-Vehicle Device

C. Bus-stop Device

On each bus stop, a Bus Stop Device (BSD) is installed. It contains a GSM modem, microcontroller and Dot Matrix Display (DMD) display. GSM receives the message from server containing bus ID and predicted time information. The message is interpreted by microcontroller, and arrival time information is shown on a DMD on a bus stop.

IV. Prediction Algorithm

A. Algorithm Flow

An in-vehicle device sends N_p number of packets per minutes out of which a few ill packets are discarded. These are the packets with some information lost during fusion in in-vehicle device and during receival on server side. On observations, it was inferred that about 3% of the total messages in a drive are discarded due to their illness. But this ratio is too small to consider. We may attain still excellent results.

As soon as a packet is received on server side, the validity of the packet is checked. If the packet is valid i.e. not ill, the extracted data from the packet is recorded in the server database for further processing. A SQL database is maintained on server side to store the received data. Database stores the vehicle ID, route ID, packet arrival time stamp and the received coordinates and speed. Predicted times of all bus stops at this point are also recorded for future use.

A unique drive number is assigned to each drive made on the route. This drive number is assigned as soon as a vehicle leaves the terminal for the drive. So, whenever server receives an invehicle's packet, it checks for whether its a new drive or not. If a new drive is detected, a new unique number is assigned to the drive. Unique drive number belongs to integers' set and incremented by 1 on each new drive.

Before starting drive on a route, the only requirement of the system is the pre-feed bus stops of the route i.e. the coordinates of all the bus stops from starting terminal to finishing terminal must be recorded in the server database before starting drive

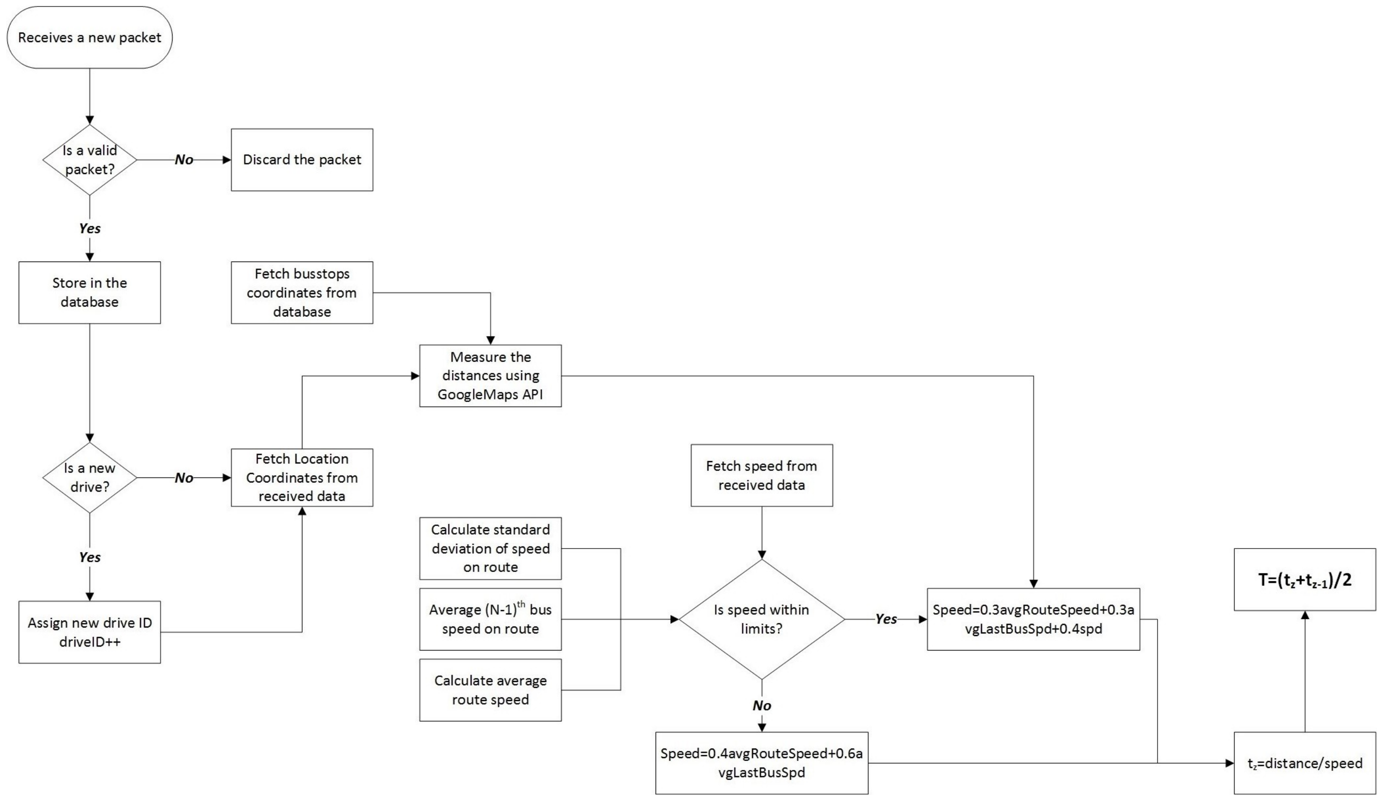

Fig. 4. Flow sequence of PTATP Algorithm

on the route. A separate relation is used for routes bus stops. Each entity is recorded with entity type routeID (a unique ID assigned to each route), status (whether the bus has passed through this bus stop or not), stopID (common name of bus stop) and coordinates of that bus stop.

For further processing, coordinates(latitude; longitude) of the current location are extracted from the received in-vehicle fused data. These coordinates along with coordinates of all the bus stops (of concerned route fetched from database) are fed to Google Maps Distance Matrix API to measure the road distance between current location of the vehicle and the bus stops. This returns a json string containing distance from all bus stops. These retrieved distances, current coordinates and instantons speed of vehicle is then used to calculate the predicted time based on our algorithm.

Complete flow diagram of implemenetd algorithm is shown in Fig. 4.

B. Time Prediction Algorithm

The algorithm takes the distance from bus stops, current coordinates and instantaneous speed as input and returns the predicted time in minutes. If we observe the instantaneous speed of the vehicle on the route, we come to know that there are so many fluctuations in the speed and we can never get a good estimate of time with these fluctuations. The observed instantaneous speed is shown in Fig. 5. With these values of speed, we'll get highly fluctuating time that can never be acceptable. We need to avoid these fluctuations in our algorithm. So, we have applied a filtering technique to smoother it so that we get some acceptable results with minimum fluctuations.

Suppose that $N-1$ buses have been traveled on the route and we are calculating the time for N^{th} bus travelling on the route. The instantaneous speed of the K^{th} vehicle is v_K that is fetched from in-vehicle data. Whereas the average speed of the selected route in V and average speed of K^{th} drive is V_K.

By **average route speed**, we mean the average speed of all the busses on the route till the Kth drive. Basically, it is the average of all the average drive speeds.

Average drive speed is the average of all the instantaneous speeds received during one end-to-end (terminal to terminal)

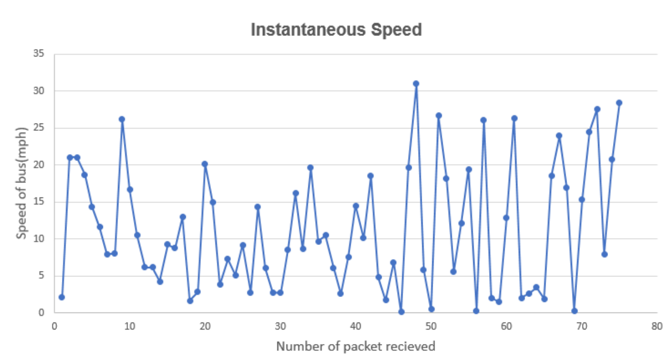

Fig. 5. Fluctuations in instantaneous speed

drive. If there are N perfect packets with v instantaneous speed are received during a drive, the average drive speed would be;

$$V_{kd} = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{1}$$

where k represents the number of drive on the route while V_d means for average drive speed.

By using Eq. 1, we can calculate the average route speed. For N^{th} drive, we exclude the instantaneous speed of N^{th} drive and take average of all the average drive speeds up to $N-1$ drives.

$$V = \frac{1}{N-1}\sum_{i=1}^{N-1} V_{iD} \tag{2}$$

where V_{iD} means the average drive speed of i^{th} drive

Maximum fluctuations occur on application of the brakes or an extra-ordinary acceleration for small intervals. Our goal is to avoid fluctuation due to instantaneous brakes and acceleration. So we don't rely only on instantaneous speed, rather we use all the aforementioned speeds to get a new scaled version of speed that is used to predict the time. During the scaling and accumulation of all the speeds (such as average route speed, average last vehicle speed and instantaneous speed), the largest weight is assigned to instantaneous speed if it is within the 1^{st} standard deviation of the route speed. However, if the instantaneous speeds goes beyond the limits of 1^{st} standard deviation, the instantaneous speed is avoided during scaling. This gives a smoother speed for time calculation. Standard Deviation (SD) defines the variation of all the members of a group from their mean. Standard deviation of all the instantaneous speeds v received during N^{th} drive is defined as;

$$\sigma = \sqrt{\frac{1}{Z-1}\sum_{i=1}^{Z-1}(v_i - V)} \tag{3}$$

where Z are the total number of valid packets received from the start of drive to the Z^{th} packet.

So, if the instantaneous speed received at a time interval t is within standard deviation i.e. $\sigma_1 > v_i > \sigma_{-1}$, then the filtered speed for calculating time prediction will be;

$$V_{filtered} = 0.4v_t + 0.3V + 0.3V_{(N-1)D} \tag{4}$$

However, if the instantaneous speed exceeds the limits, the scaling is changed to;

$$V_{filtered} = 0.4V + 0.6V_{(N-1)D} \tag{5}$$

So, these two equations 4 and 5 are used alternatively for time calculation depending upon instantaneous speed. This gives the much smoother speed that cause very small fluctuations compared with instantaneous values (as shown in Fig. 6). The predicted time is then calculated using simple physic equation;

$$t^{'}_{zs} = \frac{S_s}{V_{filtered}} \tag{6}$$

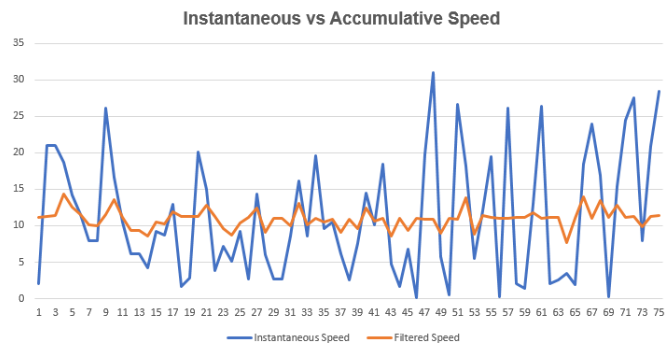

Fig. 6. Comparison of instantaneous and filtered speed

where S_s is the distance of s^{th} bus stop from current location and t_{zs} is the predicted time from s bus stop as soon as z^{th} packet is received on server which is highly accurate prediction.

However, in order to make algorithm more precise as well, the running average on predicted times is also applied. Running average mean to take average of current and previous values rather than relying on current value. So, the resulted time from Eq. 6 is not considered as final predicted time rather the running average is taken as the predicted time. If the predicted time on arrival of z^{th} packet is $t^{'}_{zs}$ and the predicted time on arrival of $z-1^{th}$ packet is $t^{'}_{(z-1)s}$, then the predicted time on arrival of zth packet will be;

$$t_{zs} = \frac{t'_{(z-1)s} + t'_{zs}}{2} \tag{7}$$

This time is the final predicted time and is sent to the respective bus stops.

V. Testing and Results

For real-time testing of the system, we selected a bus from pick and drop fleet of *University of Engineering and Technology, Lahore*. The selected bus provides pick and drop service to students on different points of the city in morning and evening daily. A number of drives were done on this route and the results were analyzed. Testing were done through our web interface.

The selected route starts from UET Lahore and finishes at *Ring Road Harbanuspura interchange, Harbanuspura, Lahore.*

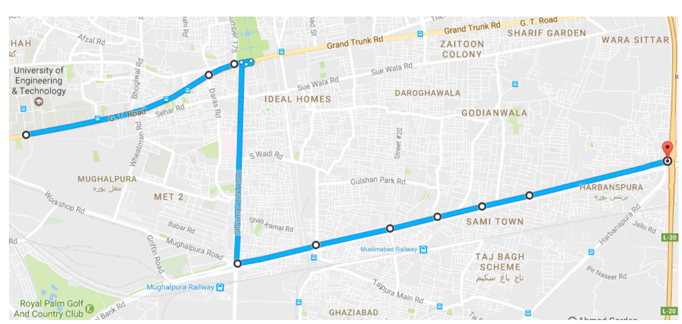

Fig. 7. Selected route for test drive

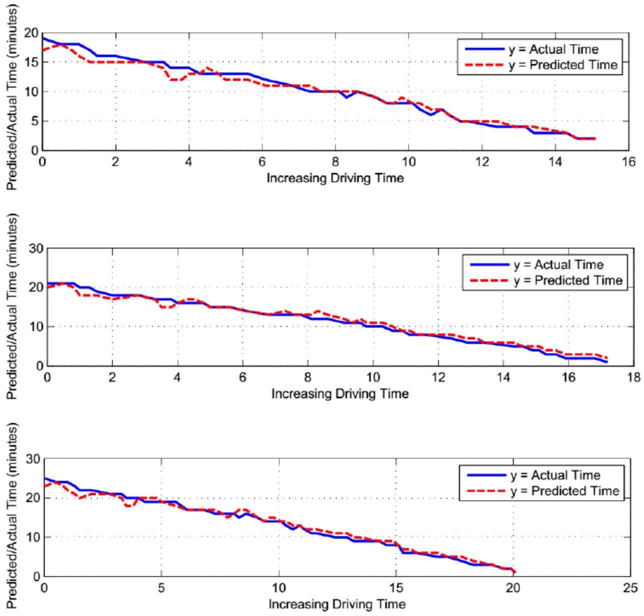

Fig. 8. Actual vs. Predicted Time on three different bus stops

The total length of the route is about 8.21 KM and includes 8 pick and drop points i.e. bus stops. The selected route is shown in Fig. 7.

A number of test drives were done on the route with modifications in algorithm to attain accurate and precise time with optimized resources. After each modification, results were attained to check the accuracy of the algorithm. To handle more and more challenges, we selected a twisted route rather than selecting a straight one.

All the drives were recorded completely. And after each drive, actual and predicted times were compared to view efficiency of our algorithm. The comparison of actual and predicted times on several bus stops has been shown in Fig. 8

The accuracy of the system is measured in terms of average error (deviation from actual time in minutes). Say that on arriving of i^{th} packet, the predicted time from s^{th} bus stop is t_{is} and actual time is T_{is} and a total number of z packets are received during a drive. Then the error calculated for this bus stop of the drive will be;

$$E_s = \frac{1}{z} \sum_{i=1}^{z} |t_{is}| - |T_{is}| \tag{8}$$

It is important to note that the error estimated using Eq. 8 depends directly on distance of bus from bus stop. As the distance between bus and bus stop goes on decreasing, the error goes on reducing. Namely, the error is much higher for last bus stop when bus starts its journey from starting terminal but this error goes on decreasing as bus comes closer and closer to the stop. That's why on a particular bus stop, PTATP shows only the time of latest bus of each route approaching to that stop.

VI. Conclusion

In this paper, we presented a public transport system with bus arrival time prediction and display mechanism. The goal was to facilitate poor people of developing countries by minimizing waiting queues and waiting times for public transport. An efficient, accurate, precise and state of the art algorithm has been used to predict exact arrival time of buses on bus stops. A number of test drives were done to analyze our algorithm and compare it with competitive techniques.

Whole system has been developed making it optimized for power and data usage. Moreover, it has been made much economical so that it can be easily implemented.

Error in predicted times by our algorithm during test drives has been shown in last section. A sufficently high accuracy with less than a minute error was achieved and implemented.

References

[1] Lahore Transport Company."Investment Opportunities". http://www.ltc.gop.pk/investment.php, Last accessed on Apr 30, 2017.

[2] Hoffman C. and Janko J. "Travel time as a basis of the LISB guidance strategy," Proc. IEEE Road Traffic Control Conf., pp. 6-10., 1990.

[3] Thakuriah P., Sen A., Li J., Liu N., Koppelman F.S., and Bhat C. "Data needs for short term link travel time prediction," Advance working paper series number 19, Urban Transportation Center, University of Illinois, Chicago., 1992.

[4] F. Siddiqui, A. Springwala, D. Shaikh, and S. Siddiqui. "GPS Based Bus Arrival Time Prediction System," International Journal of Engineering Research & Technology (IJERT), vol. 3, 2014.

[5] Lin, W. H., and J. Zeng. "Experimental study on real-time bus arrival time prediction with GPS data," Transportation Research Record 1666, Transportation Research Board, Washington, DC, 101-109, 1999.

[6] A. Stathopoulos and M. G. Karlaftis. "A multivariate state space approach for urban traffic flow modeling and prediction," Transportation Research Part C, 11, 121-135, 2003.

[7] Y. Kamarianakis and P. Prastacos. "Forecasting traffic flow conditions in an urban network -Comparison of multivariate and univariate approaches," Transp. Res. Rec., no. 1857, pp. 74 -84, 2003, Transporation Network Modeling 2003: Planning and Administration., 2003.

[8] Y. Zheng P. Zhou and M. Li."How long to wait?: predicting bus arrival time with mobile phone based participatory sensing," 10th International Conference on Mobile Systems, Applications, and Services (MobiSys12), pages 379-392, 2012.

[9] R. P. S. Padmanabanl, K. Divakarl, L. Vanajakshi, and S. C. Subramanian."Development of a real-time bus arrival prediction system for Indian traffic conditions," IET Intell. Transport. Syst., 4(3), 189-200., 2010.

[10] L. Vanajakshi, S. C. Subramanian, , and R. Sivanandan. "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses". Intell. Transport Syst., 3(1), 1-9, 2009.

[11] S.R. Seema and Sheela Alex. "Dynamic Bus Arrival Time Prediction using GPS Data". 10th National Conference on Technological Trends (NCTT09), Trivandrum, Kerala, India., 2009.

[12] Y. Lin, X. Yang, N. Zou, and L. Jia." Real-time bus arrival time prediction: A case study for Jinan," China. J. Transp. Eng., 10.1061/(ASCE)TE.1943-5436.0000589, 1133-1140., 2013.

[13] Lahore Transport Company. Bus Da Pata (2015, June 16) . [Online]. Available: https://play.google.com/store/apps/details?id=com.tpl.ltc&hl=en, Last accessed on June 30, 2017.

[14] "NMEA Reference Protocol", [Online]. Available: https://www.sparkfun.com/datasheets/GPS/NMEA%20Reference%20 Manual1.pdf, Last accessed on June 30, 2017.

[15] "Sim800 Series AT Command Manual V1.01",[Online]. Available: http://cdnshop.adafruit.com/datasheets/sim800 series at command manual v1.01.pdf, Last accessed on June 30, 2017.