International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

A Survey on Data Driven Intelligent Public Transportation using LoRa

Kathyayani Laxman Nayak[1], A. Latha[2], Majid Equbal Akhtar[3]
[1,3]*Student, Dept. of Computer Science and Engineering, Sapthagiri College of Engineering, Bangalore, India*
[2]*Professor, Dept. of Computer Science and Engineering, Sapthagiri College of Engineering, Bangalore, India*

Abstract*: This paper proposes an alternative to the conventional public transport tracking systems which uses Internet of Things(IoT) with GPS. The proposed model uses LoRa wireless transmission to communicate between the bus and cloud. This will help in tracking the buses in real-time. LoRaWAN offers a wide range connectivity while still providing a long battery life. Range of LoRa can extend up to 5 to 10 km depending on the frequency used. The buses will be equipped with LoRa transceiver, which sends out data regarding the bus identity, bus route and bus location to LoRa Gateways which in turn sends the data to cloud. The total crowd in the bus can be monitored with the Electronic Ticketing Machine (ETM) which saves the data regarding the people's journey details. By making use of ETM machine we can carefully analyse the crowd which is present in various bus stops at a given time and day. We can make use real-time Data Analytics to enhance transport planning and operational decision-making. Using data from the bus transit system – buses and passenger stations, the project will map origin-destination trends and identify bottleneck locations, information which can be used to identify whether new buses are needed. The project also explored the possibility of using real-time data to determine passenger-waiting times in order to enhance the efficiency of the bus dispatching system. The efficiency of Public Transportation Networks is a major goal of any urban area authority. Advances on both location and communication devices drastically increased the availability of the data generated by their operations. Adequate Machine Learning methods can thus be applied to identify patterns useful to improve the Schedule Plan. In this paper, we propose K-Mean Algorithms for Machine Learning and dynamic clustering procedures to determine the best Schedule Coverage to be assigned to a given bus network based on passenger count data.

***Keywords*: LoRa, GPS, Public Transportation System, Data Analytics, Machine Learning**

1. Introduction

As cities continue to grow, challenges arise in the design of urban mobility infrastructure. One of the key challenges of rapidly growing cities is to provide effective public transport services, to satisfy the increasing demands for mobility. To this end, IOT (Internet of Things) is being used to overcome existing deficiencies of public transport systems, given its ability to embed smart technology into real-life urban contexts. Public transport tracking systems are being implemented worldwide, providing updates on their availability and timing, to the society. The most commonly used technology in a tracking system, is the GPS (Global Positioning System), a global navigation satellite system used to provide geolocation and time data.

An ideal tracking system can be expected to have the following characteristics:

- *Minimum number of intermediate points between signal source and destination:*

The best case would be one where the geolocation of the source, is sent directly to the destination. In the case of GPS tracking, the intermediaries are:

i. GPS Satellites
ii. Internet gateway

- *Low power consumption*

Once the system is set up, it should run continuously without breaks, and minimal replacements of power supply. A typical GPS module draws around 70 mA on average when in use.

- *Accuracy and repeatability*

Accuracy signifies how close the measured value is close to the actual location. Repeatability shows how many times the same output is obtained when no other parameter is changed. This is of utmost importance in a tracking system, since the object being tracked will most probably be made to pass through the same location more than once, where it should produce identical output every time.

- *Easy scaling of the system*

Once an initial model has been setup, it is inevitable that the system might have to be expanded. This should be easy to implement, with minimal extra costs.

In recent times, tracking systems which do not depend on GPS have been developed. Since it is necessary to provide real time position of the bus to the users, tracking must be done with great accuracy, less power consumption and using low cost methods. This paper explains the development of prototype for efficient Public Transport tracking, using a relatively new wireless platform called LoRa. LoRa offers long range, low power, and secure communication between points up to 5-10 km in dense urban areas and 15km in rural areas. Along with providing the location of the buses plying on the route, the system proposed reduces the installation cost greatly.

In the traditional scheduling mode, the bus departure arrangements are made according to dispatcher's experience, which has the shortcomings of instability. Dispatcher fails to get the connection according to the driver's arrival situation in the process of moving and emergency on monitoring stations and roads cannot be monitored. The real-time performance of operational scheduling management is very poor, unable to update the schedule in time. With the development of intelligent city, bus operation data is increasing. In the era of big data, when there is the global positioning system (GPS) positioning data, dispatchers need only the location information of the current bus for the station arrival & situation of the bus.

The real time scheduling of buses based on big data not only saves human resources, but also meets the needs of passengers, improves the attraction of bus travel and ensures the full load rate of vehicles. City public traffic is an indispensable part of city traffic, to ensure the normal operation of city life and production and one of the important infrastructures to improve the city's comprehensive functions. It plays an important role in its industrial development, economy, cultural business prosperity and the link between urban and rural. At present, almost all cities in world have a bus operation. It must have bus scheduling management, and for scheduling management, there are a lot of cities having bus dispatching management system. However, the intelligent scheduling system is only possessed by a small part of cities, such as large and medium sized cities. But with the continuous updating and progress of data, there are various problems of updating and maintenance in these systems. Under the condition of large amounts of data, there is no accurate calculation method to rearrange the system, so it is difficult to improve the lagging platform. At the same time, in the era of big data, the amount of information is too much and too complex. The current scheduling system does not effectively integrate dynamic traffic information and vehicle location information. In most cases, the first station departure plan is based on the staff experience to determine the departure schedule. Under the condition of constantly updating the advanced vehicle equipment and communication equipment, there is no advanced algorithm for bus dispatching. Obviously, these advanced devices are not being fully utilized, while the bus service has not been improved.

The bus dispatching is the core foundation work of bus companies, which is mutually influenced according to the arrival time and the number of passenger on and off bus. Bus scheduling on the line must be rationally arranged based on the passengers' situation of each site in different time periods, and traffic situation in the sites are affected by bus arrival time, so we need the accurate prediction of bus arrival time and passenger flow. To make average waiting time of passengers the minimum, we should make the frequency of bus high, reducing the waiting time of passengers.

2. Related work

A. A Study of bus location system using LoRa

The information terminal bus stop has been studied widely. It is reported that the information terminal bus stop for citizen support is effective for the purpose of improving citizens' convenience. In particular, the bus location system is considered important for the purpose of eliminating users' concern about the arrival time of the buses. The conventional systems need the communication charges, and also introduction and maintenance costs are high. Therefore, a method using Wi-SUN has been proposed. However, Wi-SUN has a short transmission distance characteristic and requires many repeaters. The problem of introduction cost has not been solved. In order to solve this problem, we propose a bus location system using LoRa, using which the number of repeaters were reduced to 1/4 or less compared with the system using Wi-SUN. The proposed system is superior to the system using Wi-SUN in terms of installation cost.

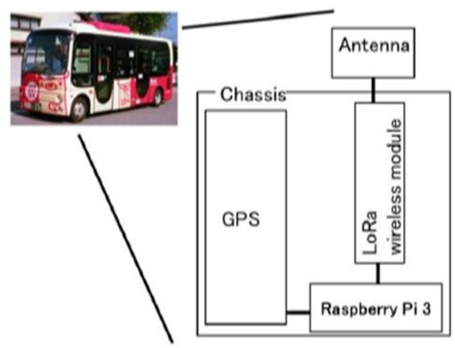

Fig. 1. Hardware configuration of a bus system

The Fig. 1 shows the hardware configuration diagram of the bus system that transmits the bus location mounted on the bus. The system is constructed GPS, Raspberry Pi and LoRa modules. The bus location information of the bus acquired by the GPS is transmitted to each bus stop using the LoRa network.

The bus location system operates on the LoRa network. The significance of leveraging the LoRa network is the reduction of the operational costs.

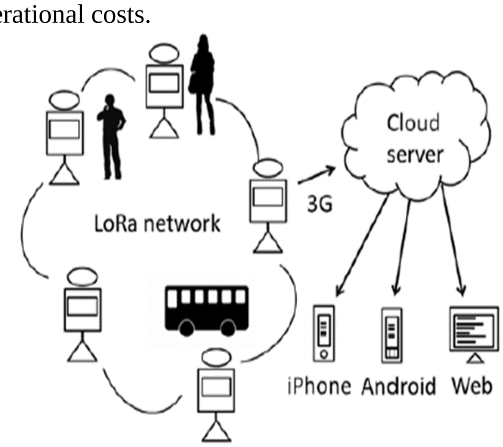

Fig. 2. Network design diagram

The bus system is mounted on the bus, and once every 30 s, the bus location information acquired by the GPS is broadcast to each bus stop by LoRa network. The bus location information can be received at almost all the information terminal bus stops (Figure 2). However, some information terminal bus stops

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

cannot receive the bus location information, because of their proximity to the railroads and the high buildings. For this reason, after broadcasting from the bus, the information terminal bus stops broadcast the bus location information to the other information terminal bus stops in sequence.

LoRa has a long transmission distance when compared with Wi-SUN. Thus, additional repeaters are not needed. That is, installation cost is low, and the time to notify all the bus stops of the buses' current location was short. Hence this system is highly practical.

B. Powering the IoT through embedded machine learning and LoRa

The Internet of Things (IoT) technology is rapidly changing the way, we live and the number of connected devices is increasing at an exponential pace. However, the two key challenges are the battery life for off-grid IoT applications and the ability of edge devices to communicate over long range. There is a need for low-power edge computing devices that reduce the transmission payload and integrate Low-Power Wide-Area Network (LPWAN) technologies, which offer a wide range connectivity while still providing a long battery life. One of the most promising LPWAN technologies today is LoRa.

High bandwidth protocols like 2G/3G/4G are power intensive and protocols like Bluetooth/ZigBee are unable to communicate over long distances (~kilometers). As shown in Fig. 3 LoRa is located at the junction of low power and long range.

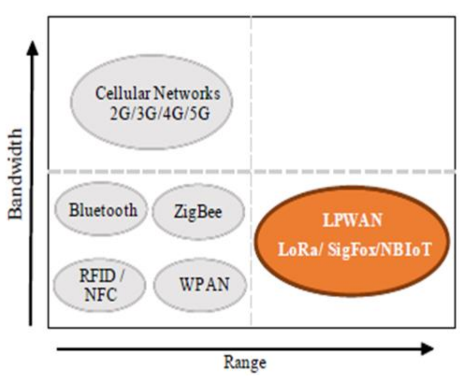

Fig. 3. Bandwidth v/s range of various wireless protocols

The Fig. 4 shows the network architecture of a typical LoRa WAN network. The end-device captures the raw sensor data, and performs data classification directly on the sensor node, reducing the data transmission bandwidth and latency. End-devices communicate with the nearest gateways connected to the network server through an IP connection. The gateways forward messages from the LoRa end-devices to the network server. The communication link is bidirectional. The communication link is bidirectional. The network server acts as the master network controller and manages the data rate of each end-device through automatic data rate (ADR) control mechanism. An application server helps in data visualization and alerts can be sent to the end-user through cellular networks.

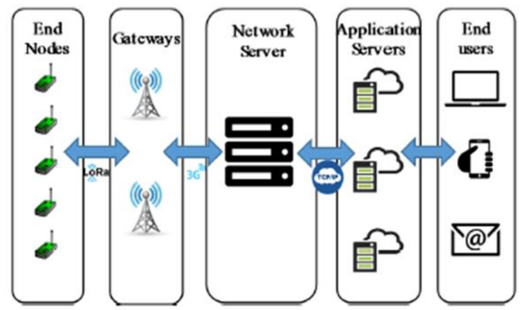

Fig. 4. LoRa WAN system architecture

Integrating LoRa will result in 3 times the saving in energy expenditure resulting in a battery life of 39 days.sd

C. An RFID based system for bus location tracking and display

RFID is a technology similar to that of bar code scanning. An RFID system consists of tags, which use radio frequency signals to transmit its location information to a reader, which usually sends this information to a server that processes it according to the needs of the application. The local server for the city receives the location information, and alerts the forthcoming bus stops in the route of the bus, of the bus's number, route and expected time of arrival, which are then displayed at the stop.

RFID works as a combination of a reader, which can read information from tags. Readers can be passive RFID systems, where the reader and reader antenna send a radio signal to the tag, and tag then uses the transmitted signal to power on, and reflects energy back to the reader. These can operate either in low, high or ultra-high frequencies, with low covering frequency covering 30 KHz to 300 KHz, high frequency covering 3 to 30 MHz and the ultra-high frequency covering 30 MHz to 3 GHz respectively.

The architecture of the system is shown using the figure5. It requires passive RFID tags holding the bus code to be placed in all the buses and low frequency RFID readers to be placed in bus stops which are approximately 20 km apart from one another. Any stops within two bus stops having RFID readers are left out, as the time of arrival of buses these stops are predicted by approximation.

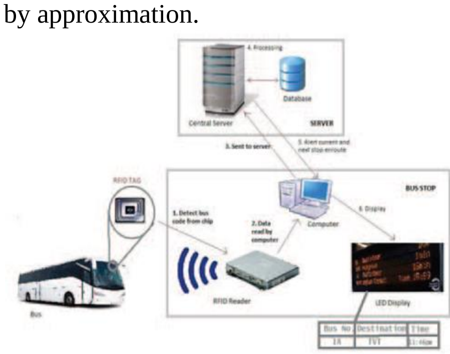

Fig. 5. Basic components of the system

The bus code is read from the tag and is passed to the system placed in the stop. This information is then sent to the central server that is responsible for identifying the bus's route and the earliest time at which it would arrive at the immediately next stop. Once all this is known, it alerts those subsequent stops about the bus number, time and destination of the bus considered. At the receiving end, the details id displayed on the LED screen.

Methodology used are,

- *Bus route mapping* - In order to efficiently identify individual bus from the huge volume of buses, the bus routes are first segmented into 4 regions: North, west, south and central. Individual bus routes are then identified by bumbers 1 through to the maximum number of routes that encompasses that region.
- *Database creation* - A database to map the bus number received from the tag to the route is created. The fields will include all the bus stops in that bus class and the corresponding distances between subsequent stops.
- *Server time estimation* - The bus starts at the first stop and the reader at that terminal collects the bus number from the tag. It sends the bus number and location identifier to the server. The server receives the information and stores it in a temporary data table that is flushed every three to four hours.

Whenever a bus number and location are received, the server checks whether it is already being tracked. If it is entered into the table, then the number of stops covered column is updated. If it is a new one, then it is entered as a new row.

Implementing this technology over a wide area would lead to generation of huge volumes of stream data which has to be handled properly. Data processing techniques for the same, which includes data querying and retrieval, must be applied.

D. A complete observation model for tracking vehicles from mobile phone signal strengths and its potential in travel-time estimation

A mobile phone periodically measures the Received Signal Strength (RSS) levels from the associated cell tower and several (six for GSM) strongest neighbor cell towers. Each such measurement is known as an RSS fingerprint.

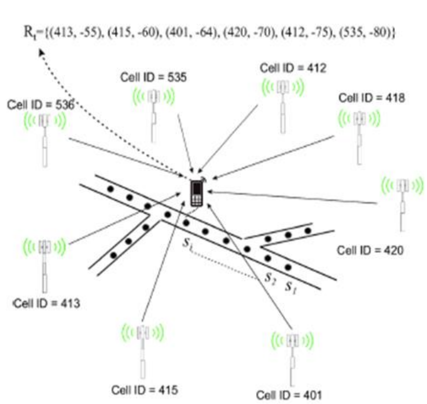

Fig. 6. Received signal strength fingerprint

However, due to various effects, the contents of fingerprints may vary over time even when measured at the same location. These variations have two components. First is the fluctuation of the RSS levels. Second is the variation of the set of cell towers reported in fingerprints. The latter is not properly modeled by traditional methods.

To address both components of variation, a probabilistic model for RSS fingerprints is proposed that specifies for each gird-location in the area of interest, the distribution of the probability of observing any fingerprint at that location. It is then used as the observation model of a Dynamic Bayesian Network to track vehicles.

Suppose the area of interest has been divided into grid-locations (black dots) and a phone is at location si at time t. Then, for example, fingerprint Rt in the figure reports signal strength levels -55dbm, -60dbm, -64dbm, -70dbm, -75dbm and -50dbm as received from cells with IDs 413, 415, 401, 420A complete observation model is presented to represent the distribution of GSM received signal strength fingerprints and demonstrated how it is used with a dynamic Bayesian network to track vehicles from mobile phone signal traces. Our observation model accounts for the variation of cell towers present in fingerprints in addition to the variation of received signal strength levels. It is used to estimate travel-times for road segments, with reasonable accuracy, using mobile phone signal strength data, 412, and 535 respectively, at location si, at time t.

E. RFID for real time passenger monitoring

In this paper they have discussed about the RFID (radio frequency identification) technology embedded on smart card to obtain relevant information about the movement of people who use public transport and thus extend the possibilities of seeking greater efficiency in operation of buses and permit to improve the services to meet the real necessities of passengers.

The RFID technology uses electromagnetic fields to automatically identify and track tags attached to objects. He tags contain electronically-stored information. Passive tags collect energy from a nearby RFID reader's interrogating radio waves. Active tags have a local power source (such as battery) and may operate hundreds of meter from the RFID reader. Unlike a barcode the tag need not be embedded in the tracked object. RFID is termed as one method for automatic identification and data transfer (AIDC).

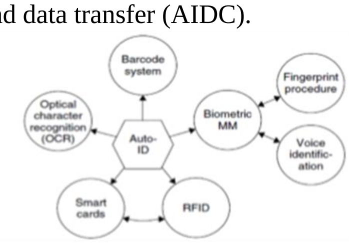

Fig. 7. Overview of most important procedures self-ID

In these studies, they test communication systems, hardware and software that can be used in solving passenger

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

identification and control problems, especially on using concepts and products of Auto-ID (Fig. 7) that are being applied in different systems around the world.

The RFID tag can be affixed to an object and used to track and manage people easily. RFID offers advantages over manual systems or use of bar codes. The tag can be read if passed near a reader, even if it is covered by the object or not visible. The card can be read inside a case, carton, box or other container and unlike barcodes, RFID tags can be read hundreds at a time. Bar codes can only be read one at a time using current device. Facing the possible use of automatic RFID in smart cards, SPTrans invited the NXP Semiconductors, the company who owns of the technology used in the MIFARE "Bilhete Único" for a partnership in order the develop and test an effective solution. The proposal is based on having a RFID tag in the "Bilhete Único" smart card (Fig. 2), thus Recent Researches in Telecommunications, Informatics, Electronics and Signal Processing making it possible to read it and identify it at distances larger than one meter. In the experiment – using an N-bit transponder, read only type – the microchip of each label (TAG) stores in its memory and unique sequential number associated with its "Bilhete Único" card number, thus allowing that, at the time of TAG reading, it conveys the transportation-identifying information to the reader. Thus, records obtained are associated only to the "Bilhete Único" which contains the TAG.

Once processed, recordings collected by equipment installed in the buses will allow the identification of each passenger boarding and alighting locations. Additionally, buses stops will also receive specific tags with a unique identification, thus allowing also for the identification of the infrastructure (route) used by each bus carrying a that activates the bus stop tag. Information on passenger card tags ("Bilhete Único") collected at bus stops and not identified in the bus on route, will also allow estimating the number number of users in bus stops and the average times for bus waiting.

Fig. 8. Smart Card "Bilhete Único" with RF technology

Internally, the bus has installed a reader device capable of managing several data collection antennas. The number of antennas depends on the size of the vehicle and the number of doors.

The demerits of using the RFID technology is that it results in data flooding. Not every successful reading a tag (an observation) is useful for business purposes. A large amount of data may be generated that is not useful for managing or other applications. Event filtering is required to reduce this data flow to a meaningful depiction of moving goods passing a threshold. Various concepts have been designed, mainly offered as middleware performing the filtering from noisy and redundant raw data to significant processed data. Global standardization is another disadvantage as the frequency used in USA is incompatible with those of Europe and Japan.There are also certain security concerns for which we have to use cryptographic methods.

F. *Self-study on ETM machines*

In this paper we are selecting a best suited method which is effective for counting the number of passengers in public transport vehicles. Counting the number of passengers is a requirement as it helps in reducing overloading of the vehicle (bus), which is the main factor affecting the passenger comfort and safety. It also helps to improve the vehicle efficiency and fuel economy. The passenger count mechanism can be effectively implemented using smart card technology over the other methods.

- *Electronic ticketing machines in public transport*

The Electronic Ticketing Machine (ETM) is a Hand Held Computer, in which the program is stored along with all the relevant data, for issue of tickets in the bus during the journey. The storage of program and data is done through a Personal Computer (Host PC), and is called pre-journey configuration for the specific route, bus service, fare, concessions, conductor, driver and day of operation, for handling the passenger and luggage ticketing.

The master data pertaining to the unit fare charges for adult/child and luggage, concessions, route, bus service are stored on the Host PC, and made available for uploading on to the ETM at the time of configuration.

Once the pre-configuration is over the ETM is ready for use during the journey. A collection report can be prepared at the end of trip. After a journey is over the transaction data files are transferred from ETM to the Host PC, over a communication cable connected to serial communication port. The ETM requires about three hours for full charging after which it can be used continuously for 7-8 hours.

- *Features of existing ETM*
    1. Fully customized fare tables for passenger Luggage
    2. Instant calculation of fare
    3. Instant display and report about passengers onboar
    4. Data transfer facility between ETM and host pc using communication cable or contactless infrared port
    5. Day end total collection report
    6. Duplicate receipt
    7. Station specific statements

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

- *Proposed methods of bus transport in bus transport system*

The proposed method consists of input devices smart card reader for acquiring passenger data such as Name, Passenger ID, address, contact details and balance available in account. For this the conductor need to swipe the smart card using the ETM. Then using keypad the travelling information such as source, destination and number of passengers to be passed in. With the help of printer, ticket will be generated with the above details and also includes available balance, transaction ID. Likewise each transaction is transferred to the central database, so that bus can be monitored regarding number of passengers using the service, total fare collected and location of the bus travelling. Additionally, proposed system includes, each card holder will be provided with the online account in Transport Corporation to know their statistics about their travelling statement. It will also reduce the burden of submitting the cash to accounts at end of the day, because the collected cash will be directly transferred to transport corporation electronically.

Fig. 9. ETM machines

G. A data-driven and optimal bus scheduling model with time-dependent traffic and demand [9]

In this paper, it has tried to leverage and make use of the abundant travel data in public transport buses. From the millions of transaction records generated when passengers board and alight at bus stations to infer time-dependent traffic and customer demand and optimize bus scheduling. The main objective of this paper is of minimizing the average waiting time. Compared with the prior bus scheduling system, this model has reduced the waiting time to some extent.

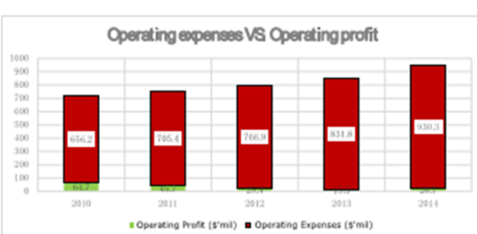

Fig. 10. Operating expenses and profit of SBS Transit from 2010 to 2014

The Fig. 10 illustrates the operating expenses and profit of SBS Transit from years 2010-2014.

As we can infer from the above graph, the operating expenses increase dramatically year by year, resulting in a decreasing profit since 2010. The slight rising of profit in 2014 is due to the increasing of bus fare by 3.2% in that year.

The optimal bus scheduling model proposed by this paper, tried to achieve its objective of minimizing total waiting time with the available time-dependent traffic and demand information in a city in China.

This work only highlights certain initial efforts on the bus scheduling optimization. From the perspective of achieving a proper optimized bus schedule, the smart-card is an important data source. This paper failed to integrate with other techniques, such as video analysis from bus CCTV or other camera techniques to count the number of passengers who have boarded and alighted for more accurate estimation for bus scheduling. Each enquiry by a traveller or a user could specify a concrete demand at a timestamp from a certain bus station. By collecting and analyzing a sufficient amount of such enquired data, we can get a more accurate estimation about user demand at different periods. It failed to create a real-time scheduling system that can respond promptly to on-the-fly passenger demands. We will create a dedicated PWA (Progressive Web App) for our Intelligent Transport System by which users with just from few clicks on their mobile phone can track their buses and estimated arrival time.

H. Big data analytics for transportation: problems and prospects for its application in China [10]

In this paper, it discusses how big data technology infrastructure fits into the current development of China, and provide suggestions for improvement. It outlines relevant big data technologies that are being used in the transportation domain in China. Through standardizing and integrating big data analytics in a national framework in the field of transport system in China, it points out the opportunities to improve the current outdated, unreliable transport system. Ever since the beginning of the industrial revolution, transportation has facilitated economic development by moving materials, resources, products and people however, the development has suffered because of road traffic congestion. Since the bus transport forms the basic mode of transport for people, it should be improved by optimizing bus scheduling. Different governments attempt to formulate effective road strategies that improve the traffic situation and overcome existing traffic problems. Unfortunately, difficulty in formulating a good transportation strategy is the lack of empirical data of actual road conditions.

This paper discusses the problems of traffic in China and point out potential solutions. Specifically, the contributions are the analysis of the current transportation situation in China, and a national multi-level transportation big data analytics framework. Normalizing the data on the length of highways and the number of privately owned vehicles with the starting year 2005, we then calculated a congestion indicator as a ratio of these two normalized values.

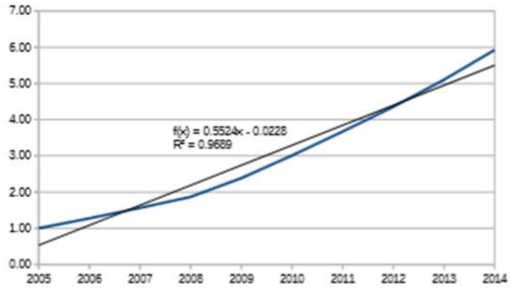

Fig. 11. Congestion indicator data (with trendline)

This indicator is shown in Fig. 11. The rising line indicates the rising ratio of vehicles to available road, and indicates an almost 6-fold increase in the 10-year interval 2005–2014. A trend line (linear regression) is overlaid, with R2 = 0.9689, indicating a very good fit.

Although intelligent transportation technology is developing in China, there is no sufficient effort to use collected data and leverage them for big data analytics. The largest problem in developing big data technology in China is related to the huge data volume captured from a very large network of traffic sensors, which requires high speed technology to capture and aggregate this data. Thus the initial investment for surveillance tools and traffic sensors is high.

Efforts for making better use of data captured by existing intelligent traffic systems would benefit from a greater degree of standardisation. There is not enough standardisation in the field of big data. The large amounts of data that are increasingly detailed, fine-grained and of ever greater coverage, allow traffic and transportation to be tracked to an extent previously not possible. However, there is still lot of scope available for introducing new applications of big data that at long last will make the transportation network better managed, more efficient, and will identify and predict future traffic needs which we will work in this project.

I. A detailed study of clustering algorithms

This paper compared and analysed some highly popular clustering algorithms where some are capable of scaling and some of the methods work best against noise in data. Every algorithm and its underlying technique have some disadvantages and advantages and this paper has comprehensively listed them for the reader. Every paradigm is capable of handling unique requirements of user application. An extensive research and study has been done in the field of data mining and there exist popular real life examples such as Netflix, market basket analysis studies for business giants, biological breakthroughs which use complex combinations of various algorithms resulting in hybrids also and subsequently cluster analysis in the future will unveil more complex data base relationships and categorical data. There exist many measures and initial conditions which are responsible for numerous categories of clustering algorithms. A widely accepted classification frames clustering techniques as:

- Partitional clustering (sum of squared error based)
- Hierarchical clustering
- *Density based*

These classifications are based on a number of factors and few algorithms have been developed bridging the multiple approaches also. In very recent times an extensive amount of algorithms have been developed to provide solution in different fields, however there is no single universal solution provided by an algorithm that solves all prevalent clustering problems.

- *Cluster Analysis*

Cluster analysis or clustering is the task of separating a set of data objects into groups, known as “clusters”, so as the objects which belong to the same cluster are more similar to each other, compared to those in the remaining clusters. It is a widely used technique in many fields of computer science, including machine learning, pattern recognition, image analysis, information retrieval, and data compression.

- *Centroid-based Clustering*

As it is indicated by its name, in centroid-based clustering clusters are represented by their centres, which are known as “centroids”. The most well-known algorithm was developed by Stuart P. Lloyd and is often actually referred to as the “k-means algorithm”. This algorithm provides an approximate solution to the optimization problem, which includes the identification of k cluster centres and the assignment of each data element to the nearest cluster, so as to minimize the Euclidean distance between each element and the corresponding cluster centroid. More precisely, the k-means clustering algorithm includes the following steps of execution:

1. The number k of clusters is provided as input.
2. The starting k “means”, which represent the initiatory centroids of the clusters, are initialized. The most commonly used initialization methods are the Forgy and Random Partition. Using the Forgy method, the centroids are randomly selected from the elements of the data set. On the contrary, in the Random Partition method each data object is randomly assigned to a cluster before computing the initial centroids.
3. The distance values between the objects of the data set and the initial centroids are calculated, so as to be used while associating each object with the cluster with the nearest centroid.
4. After assigning every object to a cluster, new mean values are calculated corresponding to the new cluster centroids.
5. Steps 3 and 4 are repeated until no alteration is observed, neither on the centroids nor on the assignments (convergence).

Regarding computational complexity, even though in the worst-case scenario the k-means algorithm requires exponential time to converge, the smoothed analysis of its complexity is polynomial. As far as its drawbacks are concerned, despite the fact that it is one of the most widely used algorithms for clustering, the definition of the number of clusters in advance is considered as a limitation, since there might exist computational problems demanding dynamic estimation of the number of clusters.

Table 1
Table title comes here

Algorithm	shape	convergence	capability
k- means	Cannot handle arbitrary clusters.	K is required in advance performance degrades with increased dimensionality.	Simple and efficient
Birch	Cannot handle arbitrary clusters.	Deals fairly with robust data. Performs strongly against noisy data.	Most famous HC algorithm
Rock	Random sampling has an impact on selection of cluster shapes.	Can handle large data sets	Usage of links gives better results with scattered points
Cure	Finds richer cluster shapes.	Cannot scale well compared to birch. Merging Phenomenon used in cure makes various mistakes when handling large data sets with comparison to chameleon.	Less sensitive to outliers. A bridge between centroids based and all points approach.

Regarding computational complexity, even though in the worst-case scenario the k-means algorithm requires exponential time to converge, the smoothed analysis of its complexity is polynomial. As far as its drawbacks are concerned, despite the fact that it is one of the most widely used algorithms for clustering, the definition of the number of clusters in advance is considered as a limitation, since there might exist computational problems demanding dynamic estimation of the number of clusters.

J. Comparative analysis of A and basic theta* algorithm in android-based pathfinding game*

This paper compared searching techniques that can be used are A* (A star) and basic theta* (Basic theta star) algorithm. A* algorithm was originally introduced by Peter Hart et al., A* algorithm calculates the cost using a heuristic function to prioritize the nodes (vertex) to be traversed, the cost is the distance it takes to walk from a node to another node. Basic theta* algorithm is variant of A* made by Alex Nash et al. in 2007 because A* has the disadvantage that the result route is often not true shortest path because the routing path is limited by the grid, while the basic Theta* change the limit of routing path becomes to all sides/angles. In this study will measure the performance of search methods which is better for square grids map on android-based pathfinding gams and used variable such as completeness, time complexity, and optimality.

- *Heuristics*

The term “heuristics” is used in computer science in order to describe techniques which offer approximate solutions to time-consuming problems, while limiting the execution time in reasonable frames. As a consequence, the selection of a satisfying heuristic method includes the evaluation of the following trade-off criteria:

- *Optimality:* A heuristic function might not be able to identify the optimal, in case several solutions exist for a given problem.
- *Completeness:* In addition, a heuristic might not offer all the possible solutions.
- *Accuracy and Precision:* Moreover, the proposed solution might lack in accuracy or precision.
- *Execution Time:* Finally, as it has already been mentioned heuristics are designed for finding quick and approximate solutions, when classic methods are either time-consuming or unable to identify any exact solution. As a result, execution time is a notable factor inthe selection of the applicable heuristic function.
- *A-star Search Algorithm*

In 1968, Peter Hart, Nils Nilsson, and Bertram Raphael introduced the A-star or A* search algorithm, an extension of the Dijkstra’s algorithm taking advantage of heuristics in order to achieve better performance, while searching for the path with the lowest cost value, connecting two nodes of a graph. The main difference between the two algorithms is observed in the function which is used for calculating the cost value of each followed path. More precisely, considering the last node of the followed path as n, the cost of the followed path is estimated by the following function:

$$f(n) = g(n)+h(n)$$

where g(n) is the actual cost for travelling from the starting node to n, and h(n) is a heuristic cost estimation for reaching the destination node starting from n. Apart from the distinctness regarding cost estimation the two algorithms include the same steps of execution, utilizing a data structure for keeping priority among the unvisited nodes according to their cost values, and carrying on execution until either the destination node is marked as visited or there no more unvisited nodes. In fact, the Dijkstra’s algorithm could be considered as a special case of A*, where the heuristic function is equal to zero (h(n)=0) for all nodes.

Regarding the running time complexity of the A* search algorithm, it is depended upon the heuristic function. In the worst-case scenario, which includes a search space without limitation, the relation between the number of nodes to be visited and the depth of the selected path is exponential. At this point, it needs to be mentioned that if the search space is unlimited and there is no path connecting the initial node with the destination, then the algorithm will never be terminated. On the contrary, in a more realistic scenario where the search space could be represented by a tree or graph, then the running time complexity of the A* search algorithm could be polynomial.

3. Conclusion

In this paper we have presented a LoRa LPWAN based public transport tracking system for bus or public transport tracking, that can provide real time position of the vehicle, with low cost instalment, and low power consumption. This method could cut down the cost of instalment to one-seventh of the requirements of the existing GPS method. If introduced throughout the state, it could create a massive change in economy by providing an efficient system for transport and logistics. All in all, this system demonstrates the potential of an LPWAN system to provide innovative and efficient transport experience. The project also explored the possibility of using real-time data to determine passenger-waiting times in order to enhance the efficiency of the bus dispatching system. The efficiency of Public Transportation Networks is a major goal of any urban area authority. Advances on both location and communication devices drastically increased the availability of the data generated by their operations. Adequate Machine Learning methods can thus be applied to identify patterns useful to improve the Schedule Plan.

References

[1] M. S. Tanaka, Y. Miyanishi, M. Toyota, T. Murakami, R. Hirazakura and T. Itou, "A study of bus location system using LoRa: Bus location system for community bus "Notty"," 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), Nagoya, 2017.

[2] V. M. Suresh, R. Sidhu, P. Karkare, A. Patil, Z. Lei and A. Basu, "Powering the IoT through embedded machine learning and LoRa," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018.

[3] V. M. Anu, D. Sarikha, G. S. Keerthy and J. Jabez, "An RFID based system for bus location tracking and display," International Confernce on Innovation Information in Computing Technologies, Chennai, 2015.

[4] C. D. Chitraranjan, A. M. Denton and A. S. Perera, "A Complete Observation Model for Tracking Vehicles from Mobile Phone Signal Strengths and Its Potential in Travel-Time Estimation," 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, 2016.

[5] SL, Abdul Haleem & Samsudeen, Sabraz Nawaz. (2016). Real Time Bus Tracking and Scheduling System Using Wireless Sensor and Mobile Technology. Journal of Information Systems and Information Technology. 1. 2016.

[6] D. K. Das Gupta, "Modification and Evaluation of Elastomer Test Machine (ETM)," 1991.

[7] Y. Wang, D. Zhang, L. Hu, Y. Yang and L. H. Lee, "A Data-Driven and Optimal Bus Scheduling Model with Time-Dependent Traffic and Demand," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 9, pp. 2443-2452, Sept. 2017.

[8] R. P. Biuk-Aghai, W. T. Kou and S. Fong, "Big data analytics for transportation: Problems and prospects for its application in China," 2016 IEEE Region 10 Symposium (TENSYMP), Bali, 2016, pp. 173-178.

[9] K. Bindra and A. Mishra, "A detailed study of clustering algorithms," 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, 2017.

[10] E. R. Firmansyah, S. U. Masruroh and F. Fahrianto, "Comparative Analysis of A* and Basic Theta* Algorithm in Android-Based Pathfinding Games," *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, Jakarta, 2016, pp. 275-280.