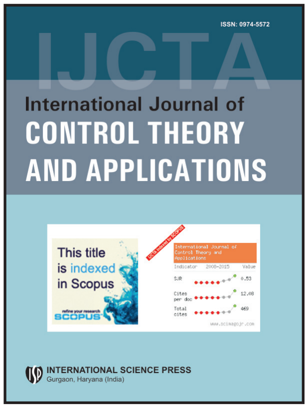

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 43 • 2016

Real Time Traffic Congestion Detection and Optimal Path Selection using Big Data Processing

Aparnasree R[a], Ani Sunny[b] and Surekha Mariam Varghese[c]

[a-c]Department of Computer Science and Engineering, M.A College of Engineering, Kothamangalam, Kerala, India. Email: [a]aparnasree.r@gmail.com; [b]anisunny88@gmail.com; [c]surekha.laju@gmail.com

Abstract: Traffic monitoring plays a vital role in management of routes. Taxis and buses are probe vehicles producing real time traffic trajectory data. Usage of this data can improve the efficiency of real time traffic estimation in a cost effective manner. Most vehicle routing systems use local map data, and can only handle static customer demand and traffic information. A solution to develop dynamic vehicle routing system based on online map service is being proposed. Google Maps API is well suitable for obtaining shortest route information. Congestion can be evaluated on that path. Sometimes, an alternative route with less congestion can take you faster to your destination than a shortest route. The congestion detection is based on the speed of the probe vehicle. At some point of time, there can be a lack of velocity information in the segment under consideration. Multiple linear regression can be used to estimate traffic conditions in such a case. In this method, the average velocity is found from the average velocities of the surrounding regions. The weather information about a region can also help in the selection of a route. The Ant Colony Optimization can be applied in finding the optimal path considering the distance and congestion.

Keywords: Traffic Congestion, Global Positioning System (GPS), Vacancy Estimation, Multiple Linear Regression, Ant Colony Optimization.

1. INTRODUCTION

One among the major crisis lies in managing the road traffic in many cities. The number of vehicles in the road is increasing at exponential rates and due to several reasons, the traffic infrastructure development is happening at a very low speed. The reasons may be difficulty in expanding the roads, fast developing city infrastructure etc. The major cause of increase in travel time is the traffic congestion. The speed of travel decreases considerably and all vehicles including the emergency services are getting delayed. There is no other way to control congestion than rerouting the vehicle, because road infrastructure expansion for avoiding congestion is not feasible. There are lot of studies happened and going on in this area, but most of them are local to an area or static with respect to time. The aim of this work is to build a congestion detection system that is capable of identifying congestion at real time and can be used without any special accessories being installed. One among the available routes having least congestion will be suggested to the users of the system.

The probe vehicles can be well used as the aid for congestion detection. These vehicles can be taxis running through the city. Their position, velocity etc. are tracked with the help of GPS trackers installed in the vehicles. These data are useful to infer the congestion of an area. The velocity of vehicle is reduced considerably from a normal rate then that can be because of congestion. The reduction of speed directly affects the travel time. So by monitoring the velocity of vehicles in a route we can reach on a decision to choose a route or not. The Online Map services like Google Map help us to get locations and routes in an efficient way. The Google Map API can be used for this purpose. The route information is obtained and can be analysed with the help of the traffic trajectory data from the taxi probe vehicles

2. RELATED WORK

The traffic congestion detection has been approached in several ways. Earlier there were various works that concentrated on the identification of congestion at one particular place by using dedicated devices. In all these, some devices were installed in such a way that the vehicle velocity is calculated or the number of vehicles that passes is counted. In [1], Inductive loops were used for finding congestion. It can detect the vehicles passing through it. But it requires the inductive loop to be installed under the road bed.

Another one is using Cameras and image processing as discussed in [2]. The cameras are installed in airplane at a high altitude. Series of images and image processing are used in the detection of congestion. Processing such a huge amount of image data is a cumbersome task. Similar method is used in [5], based on video data collected. Using Digital Image Processing the useful information is extracted. The RFID and GSM technologies were deployed by Siuli [4]. The active RFID tag sends a beacon to the nearest router and thus the vehicles are monitored and several congestion prevention techniques are also discussed. Changes in Earth's magnetic field caused by the movement of vehicle are taken as a measure to count the number of vehicle. This type of detection is discussed in [6].

There were several works done in the monitoring system using GPS technology. There were systems using mobile phones as tracking devices. Such a system was discussed in reference [3], and it uses map matching techniques to know vehicle location. The results were visualized using Google Map APIs. The online and real time dynamic route system was proposed by Riad [8]. The system uses a GPS based tracking system where the GPS tracker is the main input to the system. The probe vehicle carries the tracker and vehicle movement data is sent from vehicle to control centre in wireless mode. The Geo-info is collected at specified time interval and stored at the control centre. The speed information is used for analysis. It is used for calculating cost matrix of current time slice. Another work by Dhar [7], discusses the real time traffic congestion detection and optimal path selection. To monitor the traffic level it uses tracking of mobile phones. This work concentrates mainly on Smartphone for gathering information and GPS system. Google Map APIs are used in this to get the route information between the source and destination. The GPS devices can be installed in probe taxies, providing real time traffic information. In [10] considers this large data and uses a suitable method to process the big data and produce relevant information.

In [9], a method of estimating online vacancies in real time road traffic monitoring is proposed. The method is proposed as a solution for inhomogeneous sparseness introduced by probe vehicle. The method of finding congestion using the probe vehicle may face the problem of sparse availability of probe data and hence being not able to determine the congestion level. Then a vacancy estimation method is used and with the help of the velocity of the surrounding region, the velocity is estimated for the region under consideration.

In [11], [12] the application of the Optimization algorithm, the Ant Colony Optimization is being used as a means for optimal route finding. Based on the distance the best routes are suggested considering congestion as well.

3. PROPOSED WORK

The proposed system performs route finding on the basis of distance and congestion detection is done on this obtained shortest path. Google Map API is used to locate the source and destination, find route between them and plot them appropriately. The congestion detection is done with the help of probe vehicles. The probe vehicle share information required for congestion detection using GPS tracker system. These information helps to reach on a decision about the amount of congestion in the path. To solve the issue of sparseness in probe vehicle data, a region based approach based on multiple linear regression is used. Figure [1] shows the architecture of the proposed system.

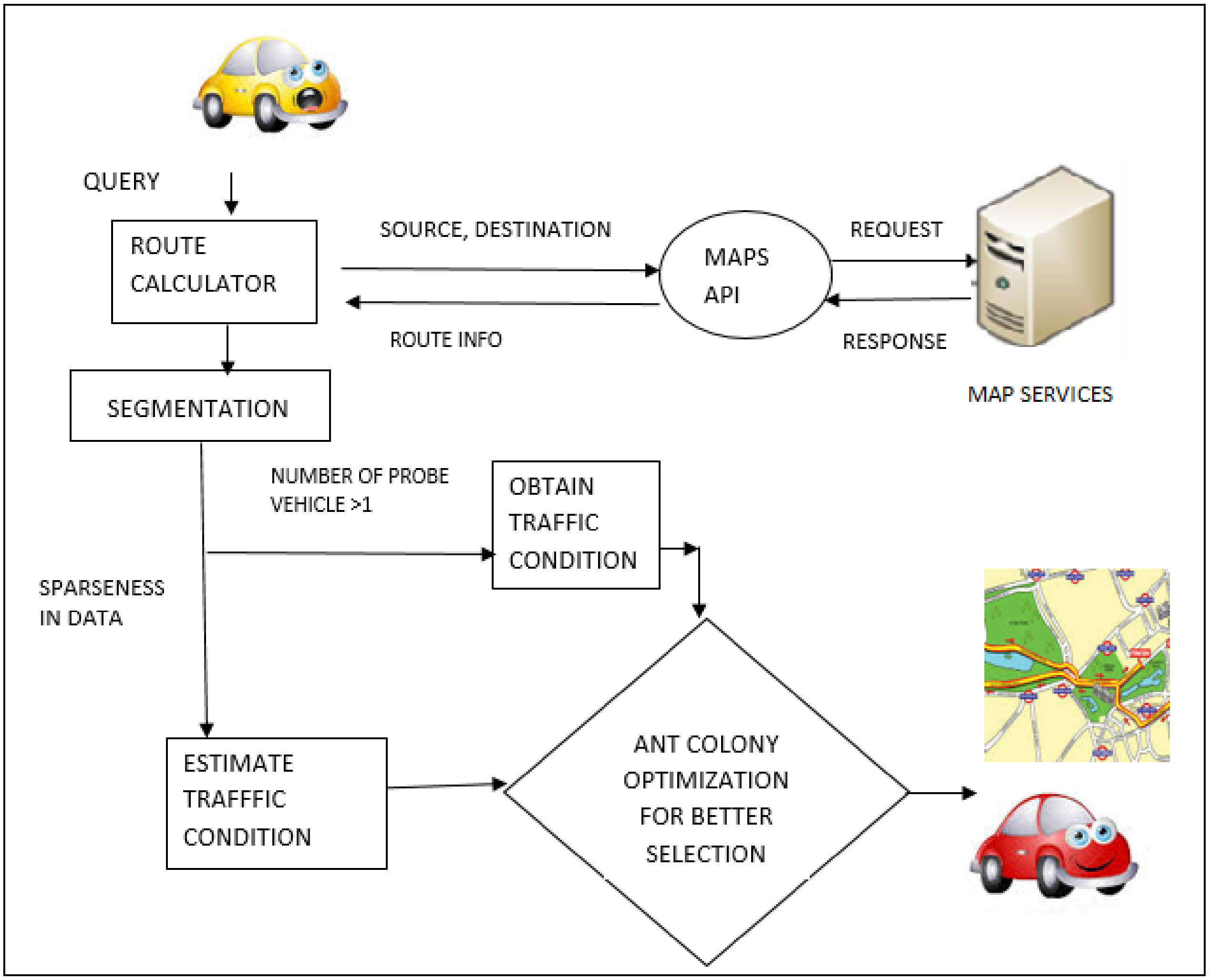

Figure 1: Proposed System Architecture

A. Obtaining Route Information

Google Map API helps developers of Map based applications to easily integrate Map into their work. This avoids the cumbersome task of buying expensive map data or learning complex development platforms for including Map services. Map can be integrated into an HTML page using Java Script. The Map (DivisionId, Map Options) function helps to load the Map in the Web Page. We can specify the level of zoom and initial centre for loading map. The Markers can be used to display locations. Direction Services are used for obtaining the shortest route information. The intermediate points in the route and the length of the route is obtained from the Google map API itself. The way points are intermediate locations between the source and destination, in the selected path. The Distance Matrix API helps to obtain the intermediate points, major junctions, etc. in the selected route. With the help of Map APIs, can effectively visualize the results obtained. By altering the parameters in the Direction Services, we can obtain multiple number of routes.

The communications with the Google Map API is performed in JavaScript. The route information is retrieved and stored as an array in the Java Script. For sending this to the python code we use Flask Framework.

B. Estimating the Congestion

Real time traffic congestion is to be estimated. The real time estimation considers only the data available for the corresponding route for a very short time period prior to the current time.

1. *Probe Vehicle:* The probe taxis have GPS tracker devices installed on it, to track the details of the taxis. The probe taxi data will contain the latitude and longitude of the probe vehicles. Usage of this type of data doesn't require any type of special device setup or privacy violations. Some existing methods for congestion detection need heavy device installation and maintenance overhead. Most of them are able to find the congestion local to an area. This system leverages the data available already in a useful manner for detecting congestion. This information are collected and transmitted via cellular network to the data centre. The data collected are then processed to provide the relevant information. All these data collection steps are being done by the taxi companies as part of tracking their taxi locations and managing their operations effectively. This data can be leveraged to have a look into the traffic rhythm of the city. Basically, such a database contain the vehicle identification number location information and the time stamp. The system aims to reduce cost of installation and maintenance by leveraging this kind of data.

2. *Route Management:* The route under consideration can be divided based on the distance in between. The congestion estimation is done on a segment by segment basis. The way points obtained from route information will be useful in segmenting the entire route. Traffic congestion is estimated in each segment and the decision is taken based on that. The segments are identified considering the four bounding points based on each intermediate points returned by the API.

3. *Mapping Trajectory Data:* The route data and the trajectory data have to be coordinated. The taxi probe data is a large streaming data. The Spark Framework is used for the management of this large taxi data. The data is read and analysed. The taxi vehicles travelling through the specified route are identified and stored in a segment by segment basis. The vehicle which are to be considered for obtaining the speed information, are found after map matching process. The latitude and longitude values of the vehicle are obtained from the tracking data and this is compared for mapping the data with the required route.

4. *Congestion Estimation:* The velocity or speed of the vehicle can be used as a measure of congestion. The vehicle cannot move in a faster speed when the road is affected with congestion. Thus monitoring the number of vehicle in the area and its velocity, the estimates of the congestion is easily estimated. The segment by segment processing is to be done to determine the congestion. The entire route is divided into segments and vehicles in those segments are identified. The average velocity of a region is found. All the segments are processed in the same way and the route average velocity is obtained. The system works smoothly if there are more than one vehicle in a segment. Otherwise the segment suffers from sparseness in the vehicle data.

5. *Vacancy Estimation:* The system is completely based on the availability of probe vehicle data. So lack of taxis in a segment we are monitoring can be an issue. The problem is approached in a different way. The congestion of a region is to be obtained. Based on Multiple Linear Regression, the velocity of vehicle in a region can be estimated based on the average velocities of surrounding regions. Figure [2] explains the gridding concept. There are eight regions being considered for estimating the velocity of vehicle. The vehicle i reports there states at time $T_i = \{t_{i1}, t_{i2}, ..., t_{ik}\}$ and report set, $S_i = \{s_i(t) \mid t \in T_i\}$. In a segment, the traffic at time t_0, average speed, in region r_0 and direction d_0. Reports from neighbours of region (r_0, t_0, d_0), let $R(r_0, t_0, d_0)$ denote the set of reports from the neighbourhood.

$$R\,(r_0, t_0, d_0) = \{S_i(t) \mid r(r, r_0) < \Delta r \wedge |t - t_0| < \Delta t \wedge d = d_0\} \quad (1)$$

Then, the average speed is computed as

$$\hat{v}\;(r_0, t_0, \mathrm{d}_0) = \frac{1}{N} \sum\nolimits_{\mathrm{Si}(t) \in R(r_0, t_0, d_0)} S_i(t) \cdot v,$$

where, $N = |R\,(r_0, t_0, d_0)|$ (2)

Need to overcome the issues related to sparseness of data in some region, we can estimate the traffic condition depending on the states of neighbouring regions. By considering urban traffic network connectivity, we can find an enclosed neighbouring area to compensate this. Given a set of regions R in a set of time series T in the direction d0, where

$$R_n = \{r_0, r_1, ..., r_{n-1}\},$$

$$T_m = \{t_0, t_1, ..., t_{m-1}\}$$

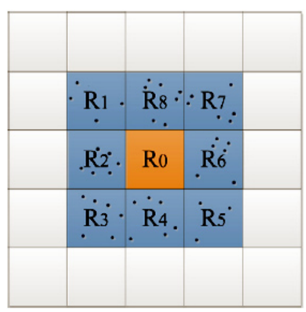

Figure 2: Region view

If $N = |R(r, t, d)|! = 0$, we compute a traffic condition matrix (TCM), denoted by $C_{TCM} = (v_{ij})_{m \times n}$, where, v_{ij} is the average speed of $R(r, t, d)$. If $N = |R(r, t, d)| = 0$, we find estimated TCM, denoted by $E_{TCM} = (v_{ij})_{m \times n}$.

Linear regression models contain more than one predictor variable. In the system, it is assumed that there are m neighbouring regions available around the estimation region r_i, denoted as $r_1, ..., r_m$, and the neighbour node set denoted as $R_{nb}(i) = \{r_1, r_2, ..., r_m\}$. Because regions are geographically close resources, not only do they correlate spatially with every neighbouring region, $r_j, r_j \in R_{nb}(i)$, but there are also spatial correlations among neighbouring regions, denoted as r_j is spatial correlation with r_k, where $\forall r_j, r_k \in R_{nb}(i)$. Therefore, all neighbouring regions are considered a whole in the estimation, since using a single region will introduce random errors. We establish multiple regression models for representing spatial correlations between r_i and its neighbouring regions at time t, denoted as

$$v_{it} = \beta_0 + \beta_1 v_{1t} + \beta_2 v_{2t} + \cdots + \beta_m v_{mt} + \mu_t \quad (3)$$

where, v_{it} is the traffic condition in region r_i at time t, v_{kt}, $k = \{1, 2, ..., m\}$ is the traffic condition in region r_k, $r_k \in R_{nb}(i)$ at time t, β_k are the partial correlation coefficients for v_{kt}, and μ_t is the random error term. V_{it} will be considered an explanatory variable, and partial correlation coefficients β_k represent the impact of v_{kt} to v_{it}, $k = \{1, 2, ..., m\}$. Obviously, we can estimate $\widehat{v_{it}}$ by using above equation when it is a vacancy in the region. We will sample h , $(h - m \geq 2)$ items to establish multiple linear regression models for computing the estimates $\widehat{\beta_k}$ for β_k. B_k is substituted for $\widehat{\beta_k}$ and can be further derived.

$$\widehat{v_{it}} = \widehat{\beta_0} + \widehat{\beta_1}\, v_{1t} + \widehat{\beta_2}\, v_{2t} + ... + \widehat{\beta_m}\, v_{mt} \tag{4}$$

where, $\widehat{v_{it}}$ is an estimation for v_{it} and v_{kt} is the true value in the region k, $k = \{1, 2, ..., m\}$ at time t. Equation is called the vacancy estimate equation, and the error between the estimation and the true value is called the residual, which is denoted as et, $et = v_{it} - \widehat{v_{it}}$. During the computation of correlation coefficients β_k, we employ a vector to represent the traffic condition collected from h samples, denoted as $V = (v_{i1}, v_{i2}, ..., v_{ih})^T$, if there are m neighbouring regions and they construct the matrix X denoted as

$$X = \begin{matrix} 1 & v_{11} & ... & v_{m1} \\ 1 & v_{12} & ... & v_{m2} \\ 1 & v_{1h} & ... & v_{mh} \end{matrix}$$

Similarly, we can rewrite the estimation of correlation coefficients β_k as follows

$$\hat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, ..., \widehat{\beta_m})^T = (X^T X)^{-1}(X^T V) \tag{5}$$

6. *Finding Optimal Route:* The multiple routes are returned by the API, can be considered as the road network where we need to perform Ant Colony Optimization (ACO). The graph G, G = (N, L)

Here N is the set of nodes, here the set of major cities or junctions and L is the links connecting them, here the roads. The best path should be the path having least travel time considering distance congestion and weather. The traffic congestion T_{ij} is the factor that changes with time. The edge L_{ij} is not based on the distance d_{ij} between i and j only. It depends on the traffic T_{ij} also. Each route $R = r_{ij}$ where,

$$r_{ij} = \begin{cases} 1, \text{ if there is a link between } i \text{ and } j \\ \quad 0, \text{ otherwise} \end{cases} \tag{6}$$

The travel time Time (R) of route R is given by,

$$\text{time(R)} = \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij} d_{ij} T_{ij} \tag{7}$$

The objective is to choose a path that minimizes Time (R)

$$\text{Best Score (S)} = \text{Score at min } (x_{ij} T_{ij})$$

While travelling from node i to j cars take decision based on the probability

$$P_{ij} = \frac{T_{ij}^{\alpha} d_{ij}^{\beta}}{\sum_{q \in Q} T_{qj}^{\alpha} d_{qj}^{\beta}} \tag{8}$$

The set Q is the set of all cities, and gives the influence of traffic (pheromone) and distance on the solution. The traffic (pheromone) update depends on the evaporation rate (the rate at which cars leave) ρ and the deposition rate Δ_{ij} (the rate at which cars arrive). The pheromone or traffic update is governed by the equation:

$$T_{ij} = (1 - \rho)T_{ij} + \Delta_{ij} \tag{9}$$

Δ_{ij} depends on whether a car used the edge r_{ij} or not. The total amount of pheromone added or traffic added can be calculated as follows:

$$\Delta_{ij} = \sum_{k=1}^{N} \frac{r_{ij}^{(k)}}{\text{time}(k)} \tag{10}$$

time(k) is the time taken by the car k in covering that section of the road and is a function of the speed of the car $v(r)$. N is the total number of cars in that section of the road

$$\text{time}(k) = \frac{d_{ij}}{v(k)} \tag{11}$$

Thus with the help of ACO the best path can be selected based on the pheromone deposit on each of the links.

4. CONCLUSION

The congestion detection is one among the hot areas of research due to its relevance in the present scenario. The system helps to reach on a decision on selection of one among the available shortest routes based on the congestion. By leveraging the existing taxi trajectory data the system help to avoid lot of additional expenses and overheads. Installing and monitoring of devices for congestion detection is the major overhead. Data from Smart Phones are rarely shared by users for such purposes. The usage of Online Map Service helps to get realistic route finding results. The problem of lack of trajectory data at some location is solved with the help of the vacancy estimation. It uses the multiple linear regressions to estimate velocities of a region using the information from adjacent regions. Thus the system help to identify a least travel time path from a source to destination in a high efficiency and least cost.

REFERENCES

[1] Cherrett, T., Waterson, B. and McDonald, M. "Remote automatic incident detection using inductive loops". Proceedings of the Institution of Civil Engineers: Transport, 158, (3), 149-155, 2005.

[2] Palubinskas, G., Kurz, F., and Reinartz, P., 2009. "Traffic congestion parameter estimation in time series of airborne optical remote sensing images". In: Proc. of ISPRS Hannover Workshop 2009 – High Resolution Earth Imaging for Geospatial Information, 2-5 June, 2009.

[3] Hnin Thant Lwin, Thinn Thu Naing , "Estimation of Road Traffic Congestion using GPS Data", IJARCCE, 2015

[4] Siuli Roy, Somprakash Bandyopadhyay, Munmun Das, Suvadip Batabyal, Sankhadeep Pal, "Real time traffic congestion detection and management using Active RFID and GSM technology", LAP Lambert Academic Publishing (2012-10-09).

[5] Pradip Singh Maharjan, Ajay Kumar Shrestha, "Automatic Vehicle Detection and Road Traffic Congestion Mapping with Image Processing Technique", International Journal of Computer Applications (0975 – 8887) Volume 114 – No. 16, March 2015

[6] Nitish R Chavan, Ravinarayana .B, Ravikumar .D, "Wireless Sensor Networks for Traffic Congestion Monitoring", International Journal of Science and Research (IJSR) ISSN (Online), 2013

[7] Joydip Dhar, Gaurav Garg, "Real Time Traffic Congestion Detection and Optimal Path Selection Using Smartphone", IEEE, 2014.

[8] A. M. Riad, M. E. A. El-Mikkawy, B. T. Shabana, “Real Time Route for Dynamic Road Congestions” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012

[9] Feng Wang, Liang Hu, Dongdai Zhou, Rui Sun, Jiejun Hu, Kuo Zhao, “Estimating online vacancies in real-time road traffic monitoring with traffic sensor data stream”, ELSEVIER, 2015

[10] Sumit Kumar, Gautam Srivastava, Md. Shameem, Amber Khan “Vehicle Positioning System Using GPS, GSM & GIS ”

[11] Saurav RANJIT, Masahiko NAGAI, Itti RITTAPORN, Thanomsak AJJANAPANYA,Fredrik HILDING, Apichon WITAYANGKURN, Ryosuke SHIBASAKI, “GPS Enabled Taxi Probe’s Big Data Processing for Traffic Evaluation of Bangkok Using Apache Hadoop Distributed System”,2014

[12] Joydip Dhar, Gaurav Garg, “Real Time Traffic Congestion Detection And Optimal Path Selection Using Smartphone”,IEEE 2014

[13] Mehmet Yildirimoglu and Nikolas Geroliminis. “Experienced travel time prediction for congested freeways”, Transportation Research Part B: Methodological, 53(0):45 – 63, 2013.

[14] Pavel Kr¨omer, Jan Martinoviˇc, Michal Radeck´y, Radek Tomis, V´aclav Sn´aˇsel, “Ant Colony Inspired Algorithm for Adaptive Traffic Routing”, IEEE, 2011.

[15] Jerry Kponyo , Yujun Kuang , Enzhan Zhang, Jerry Kponyo. “Dynamic Travel Path Optimization System Using Ant Colony Optimization”, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014