

## SEMUT: Next Generation Public Transportation Architecture in the Era IoT and Big Data

<sup>1</sup>Agus Sukoco, <sup>2</sup>Ary Setijadi Prihatmanto, <sup>3</sup>Rifki Wijaya, <sup>4</sup>Ilyas Sadad and <sup>2</sup>Reza Darmakusuma

<sup>1</sup>Department of Informatics, Faculty of Computer Science,  
Universitas Bandar Lampung, Lampung, Indonesia

<sup>2</sup>Institut Teknologi Bandung, Department of Electrical Engineering,  
School of Electrical Engineering and Informatic, Bandung, Indonesia

<sup>3</sup>Department of Computer Engineering, Faculty of Electrical Engineering,  
Telkom University, Bandung, Indonesia

<sup>4</sup>Department of Civil Engineering, Faculty of Engineering,  
Universitas Bandar Lampung, Lampung, Indonesia

---

**Abstract:** Transportation problems in various forms are one of the main problems faced by world cities including cities in Indonesia. Vehicle growth increases the intensity of traffic in goods and services and increases in population associated with users that are not comparable to the availability of road facilities and supporting facilities that generate crucial problems. This leads to a decrease in the effectiveness and efficiency of the transportation system to support the activities of civilization. It must be realized that information technology is the need for a road transportation system in the future. This research program is based on the understanding that solving transportation problems in major cities in Indonesia cannot be done only in the usual way that has been done in developed countries but requires a unique approach that requires the application of all the available technology in the ICT field. Public transport has a new IoT breakthrough that provides more efficient and value-added public transport services to get the latest big data to gain new power and services for citizens and thus make public transport more efficient and happier. The SEMUT architecture used is prepared to be used at the instrumentation level with massive data for that architecture is needed to be able to send all data requests and post data of at least 500 thous. devices, especially, for domain transportation. In this study we use 300 existing nodes that do not cause a burden to the CPU or memory.

**Key words:** SEMUT architecture, big data, big data analysis, ICT field, public transport, CPU

---

### INTRODUCTION

Daily transportation, especially, public transportation is the main domain problem in the world. Public transportation is one of the most public service centers related to daily life (Samaras *et al.*, 2015). Through the system, everyone works to live more independently and connect with each other (Hallgrimsdottir *et al.*, 2016). The increasing population is one of the important things for related parties to identify the determinants of reliability of public transport services (Murad *et al.*, 2018).

The most important thing about sustainability is how the operational strategies that determine the sustainability of the company, especially, from the financial aspects (Tanaka *et al.*, 2016), so that, data that can be processed into information is needed to support decision-making systems (Duarte *et al.*, 2016).

Service improvements, especially, in the field of public transportation, rules limit the number of private vehicles and campaigns to use public transportation are not enough to improve a reliable public transportation system. Public transportation is designed to provide value for efficiency but now there are several alternatives that are better for public transportation, sometimes not in line with the quality of service (Nguyen *et al.*, 2016). One other problem is the uncertainty of the arrival of public vehicles (Tanaka *et al.*, 2016). However, this uncertainty causes public transportation to find passengers that results in congestion (Zhu *et al.*, 2016).

Because information and communication technology provide opportunities to transform public transportation services based on IT, especially, on IoT, public transportation has a new breakthrough (Puiu *et al.*, 2017) for sustainability. A new breakthrough in the form of IoT

in the form of a simple GPS tracker, Mobile, provides public transportation services that are more efficient and provide added value to get the latest big data (Mohanty *et al.*, 2016). Big data products make it possible to solve problems and develop urban public transportation innovations. Big data provides the possibility to solve urban public transportation problems (Li *et al.*, 2017). It is therefore, important to develop a new type of intelligent transportation system to coordinate people, vehicles and roads more smoothly and happily and then increase traffic efficiency to reduce traffic congestion and make people more comfortable. SEMUT App. is a public transportation application to provide community services aimed at improving traffic management, real-time traffic information, public transportation and solving local problems for existing problems, especially, public transportation. Application innovation to support large data provides new opportunities for many solutions to public transportation problems.

## MATERIALS AND METHODS

**Application layer of modern digitalize:** Basically, many services today are complicated combinations of complete implementations and popular concepts such as big data and cloud computing, advanced instrumentation and internet of things, social networks and digital gamification, modeling and simulation, content-human interaction and machine and system learning intelligent. The possibility of data availability is not only possible because of technology but is mainly driven by the feasibility and economic opportunities of abundant and inexpensive data communication through the Internet and the availability of instrumentation for various phenomena to be connected to the internet (Advanced instrumentation and internet of things).

In addition, the availability of extraordinary data for certain phenomena allows the process of calibrating mathematical models from phenomena that allow the model not only to have similar qualitative behaviors and relevance to real phenomena but also quantitatively similar. As a result, modeling and simulation activities become increasingly relevant and are used as a major part of the analysis with increasingly quantitative predictive capabilities (Fig. 1). At the user level, the results of data processing will form various meta-information that must be able to interact with customers (human-content interaction as a top-up of human-computer interaction) optimally in accordance with:

- The user's physical and mental abilities
- Purpose of use as well as
- Unique user habits and usage preferences

The combination of technology above provides an opportunity for the emergence of premium services and technologies that can serve a large number of users at affordable costs. The combination of technology above is carried out in a complex level of implementation as in Fig. 2. In the picture above, interactive digital media services are built in 4 application layers.

Level of instrumentation application that is when information from the real world is digitized. For example, in transportation services, the application level can be information from CCTV sensors, traffic density sensors, GPS sensors in each vehicle, water level sensors in several places or actuators in the form of controls for traffic lights, notifications on various media on the road, indications in mobile device applications, etc.

Information system application level that is when information from instrumentation equipment is not only used in a momentary situation but is collected structurally and systematically in a time domain (historically) as well as in the spatial domain, so that, it can be displayed or processed richer and more meaningful for the user. For example in the transportation service above, historical traffic information from intersections can be displayed so that users can draw more precise conclusions about the data.

Business intelligence system application level that is when information obtained from the application level of information systems is managed on a more massive scale to get richer, more diverse and in-depth insights that can be obtained in spatial information system layout, so that, it can help users make decisions which is more on target. This measure usually involves processing big data using mathematical models and more explicit modeling concepts (modeling and simulation). In the example of the transportation service above an accurate estimate of travel time can be obtained by processing information collected over the years from hundreds of CCTV camera sensors, traffic density sensors, information from social media, etc. It is used to calibrate the network traffic of mathematical models, so that, it can be simulated with high fidelity to be able to predict estimated travel times at certain times.

Gamification system application level which is when all quantity of metadata that can be inferred from the level of business intelligence system application is managed and processed by engineering the system user behavior

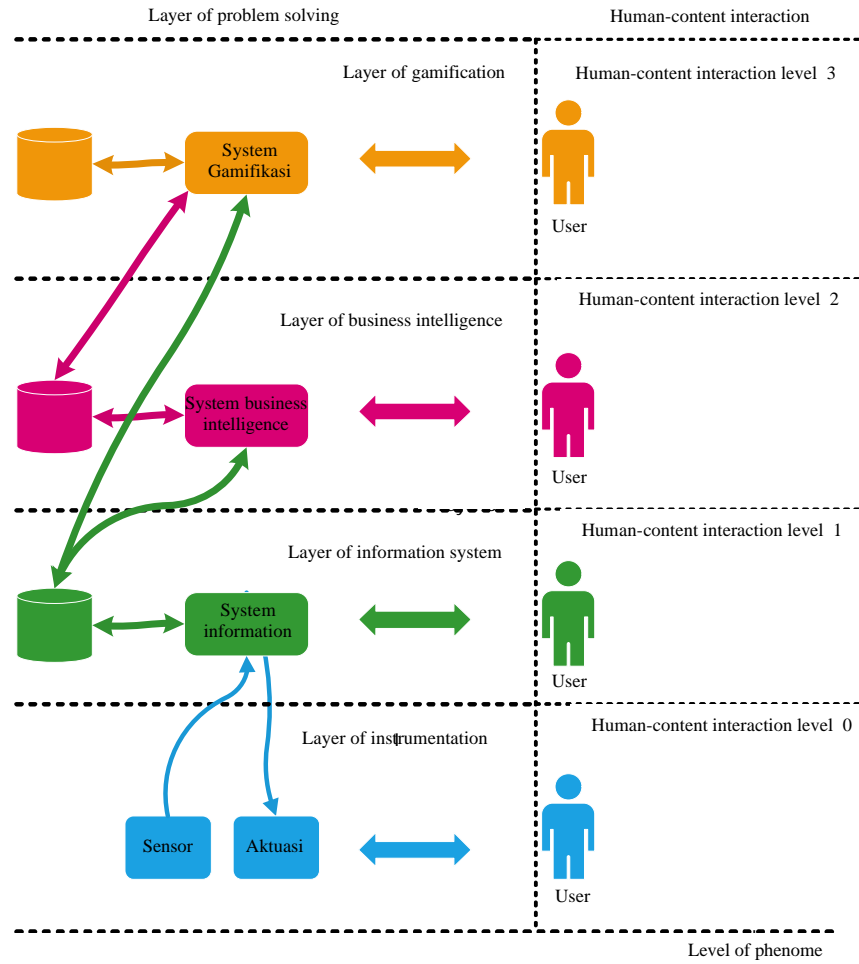


Fig. 1: Application layer in modern interactive digital media service

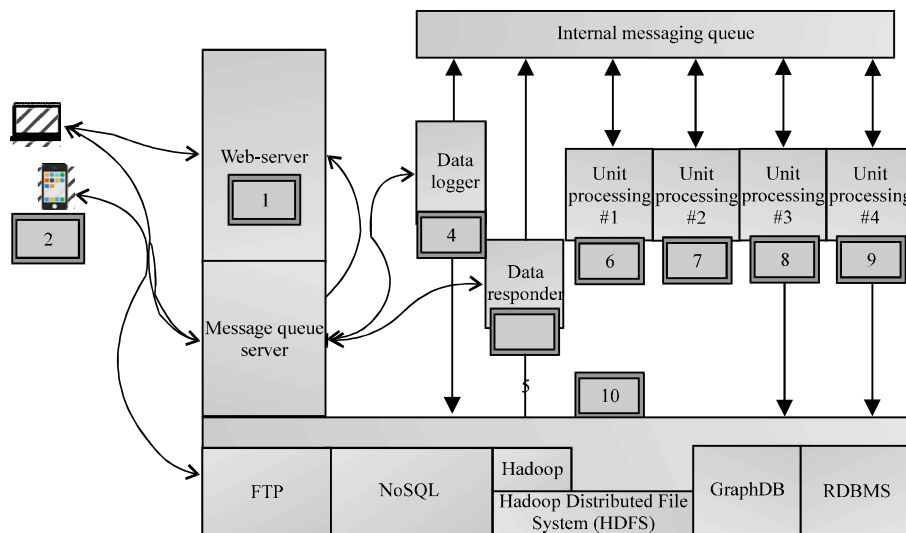


Fig. 2: SEMUT architecture for public transportation



Fig. 3: Exchange implementation

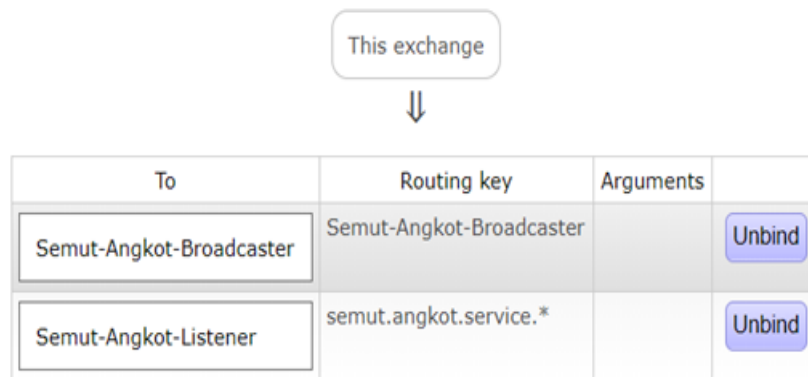


Fig. 4: Queue implementation

to achieve certain KPI systems. In the example of a transportation service, this level can be guaranteed as an example:

Gamification system application level which is when all quantity of metadata that can be inferred from the level of business intelligence system application is managed and processed by engineering the system user behavior to achieve certain KPI systems. In the example of a transportation service, this level can be guaranteed as an example.

By knowing the estimated travel time under certain user conditions, the system can perform user behavior on the trip, so that, users can pass it more productively and save energy. Can manage user behavior with a variety of patterns that are not easy to do before (e.g., even odd number systems), so that, the actual travel time can be obtained optimally. The laminar application process is the implementation norm in modern digital media application services. This is very important in developed countries but in reality our country needs technology and technology-based solutions that are typical of departing because:

The pattern of interactive digital media services is a service pattern that has the potential to reach all users who may be relevant compared to conventional services with limited capacity. Digital media application services can be built with large capacity at a relatively low cost compared to conventional services. Digital media application services in developing countries such as Indonesia are built without basic infrastructure at the same level as developed countries. As a result, the quality and performance pressures that must be gained from digital media application services in developed countries enable the emergence of unprecedented breakthroughs (Fig. 3 and 4).

Further development of applications such as Waze allows all transportation users not only to know the actual traffic conditions more precisely but also to predict future traffic conditions, so that, they can plan their activities better and planned. Furthermore, the system can utilize simulation results and data analysis processes to carry out the traffic engineering process in a timely manner. Distribution suggestions can be made directly to users, unique to their circumstances, conditions and needs.

On the other hand the logistics system, the same approach in the framework of production and distribution of goods and services allows all parties to get a precise picture of the actual market situation and distribution chain but also to predict what will happen in the future. This allows better planning and decision-making processes from all parties that minimize overall costs and price volatility.

SEMUT architecture system for public transport consists of several components. Data acquisition sensor, messaging broker, services that serve data, database for storing data. The implementation of a client-server architecture with the publish-subscribe paradigm using an intermediary message middleware with asynchronous communication is expected to be applied to the server to scale easily. The basic form of middleware communication is by sending a message from sending to the recipient via a queue where the post is called the publisher who will send the message via the queue topic. Recipients are called customers who will receive messages according to the topic being followed. The system by applying the message architecture broker middleware consists of several components.

**Messaging broker:** The broker message is basically a data queue that connects the producer and the consumer. The Messaging Broker implemented is RabbitMQ (Roy, 2018). RabbitMQ provides message queues as intermediate messages sent between sensors and services. Message queues will receive messages sent from sensors. Then, it will send a message to the destination service only if the service is active or online. RabbitMQ uses the Advanced Message Queuing Protocol (AMQP) and the adapter is available to serve the MQTT protocol (Lin *et al.*, 2017). Queues are made to hold the queue of data from sensors (Ayanoglu *et al.*, 2016). Service subscribes to the queue. A queue is tied to an exchange of topic type, so, it's flexible in doing message routing.

## RESULTS AND DISCUSSION

The sensor retrieves the geolocation data for acquisition of data. Sensors become producers for acquisitions and publish possible data with the most important forms of technology. Sensors can send data to the message middleware with protocols that support the message broker. In this study, the sensor consists of two forms, namely the angkot mobile application and the GPS tracker microcontroller module.

**Mobile application:** This application is built with an Android operating system that has a GPS sensor and internet data connection. The application sends messages periodically for 10 sec after the longitude coordinate data,

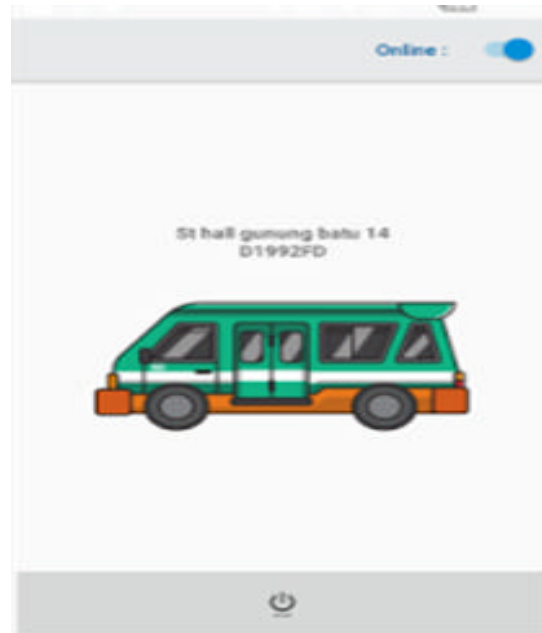


Fig. 5: Mobile application for angkot sensor

the latitude position to the message intermediary uses the Advance Message Queuing Protocol (AMQP) with JavaScript Object Notation (JSON) format. This format will facilitate the process of parsing data (Fig. 5) (Algorithm 1):

### Algorithm 1; Parsing data:

```
{
  "session_id": "587c7115672fe09bc2c14ad"
  "time": "2017-10-12 14: 05: 23"
  "longitude": "-6.8907437"
  "jumlah_penumpang": "5"
}
```

**GPS tracker modules:** GPS tracker sends data with internet connection via WiFi module ESP8266 connected with MiFi modem internet. The protocol used to connect with messaging brokers is MQTT. The GPS tracker module is set to transmit data every 10 sec. A unique and registered value MAC address is used to perform data validation (Algorithm 2):

### Algorithm 2; MAC data validation:

```
{
  "Mac": "5C: CF: 7F: 1D: B9: 2C"
  "Date": "21/10/2017"
  "Time": "20:34:11"
  "Data": [
    -5.378157
    105.251014
  ]
}
```

**Service:** This service is a program that becomes a subscriber who subscribes from the messaging broker to receive data from the sensor which is then saved to the

```

MINGW64/e/S E M U T/semut-angkot-rest-api
POST /api/v2/notif/flash 200 2.258 ms - 226
POST /api/v2/cctv/bandung 200 4.003 ms - 30149
POST /api/v2/users/status 200 2.029 ms - 60
POST /api/v2/notif/flash 200 2.425 ms - 226
POST /api/v2/cctv/bandung 200 3.916 ms - 30149
POST /api/v2/cctv/bandung 200 3.997 ms - 30149
POST /api/v2/users/status 200 2.146 ms - 60
POST /api/v2/notif/flash 200 2.524 ms - 226
POST /api/v2/cctv/bandung 200 4.280 ms - 30149
POST /users/login 200 18.873 ms - 891
POST /users/status 200 6.335 ms - 60

Incoming msg : {"session_id":"08daf60b89272a22388ea3fb8545cad4","time":"2017-12-23 15:32:42","longitude":107.6515223,"latitude":-6.9640975,"speed":0.001566931875751716,"jumlah_penumpang":0,"status_online":1}
-----
UPDATE LOKASI ANGKOT
process success : {"ok":1,"n":1}
duration [ms] = 10

Incoming msg : {"session_id":"08daf60b89272a22388ea3fb8545cad4","time":"2017-12-23 15:33:11","longitude":107.651523,"latitude":-6.9640976,"speed":0.004341061744425032,"jumlah_penumpang":0,"status_online":1}
-----
UPDATE LOKASI ANGKOT
process success : {"ok":1,"n":1}
duration [ms] = 8
    
```

Fig. 6: Service implementation

```

Result: Query Explain
1 [{"_id" : ObjectId("5a28e6b02fa8cfa27cf4e750"),
2  "location" : {
3    "type" : "Point",
4    "coordinates" : [
5      107.609628,
6      -6.8891742
7    ]
8  },
9  "speed" : NumberInt(0),
10 "jumlahPenumpang" : NumberInt(0),
11 "lastupdate" : ISODate("2017-12-07T06:58:56.719+0000"),
12 "typeTrip" : NumberInt(1),
13 "ID" : NumberInt(3746)
14 }
15 ]
16 [{"_id" : ObjectId("5a28e6b02fa8cfa27cf4e751"),
17  "location" : {
18    "type" : "Point",
19    "coordinates" : [
20
    
```

Fig. 7: Implementation of database

database. Services can be made with various programming languages by implementing protocols that match the message of the broker (Duarte *et al.*, 2016). The number of producers used will increase the amount of data that must be possessed by the number of consumers who can add excess load.

The service is tasked with subscribing data from the messaging broker with the AMQP protocol. Service is built using NodeJS with the Javascript programming language. Then, the service to store data into the database that is implemented using Mongo DB. The advantage of using Mongo DB is that it is not necessary to define a schema as it needs to be done on a Relational DataBase (RDBMS) (Kang *et al.*, 2016), velocity processing data especially for big data (Celesti *et al.*, 2018), So, the difference of JSON format from the resulting mobile application and GPS tracker module is not an obstacle (Fig. 6 and 7).

**Database:** The database allows storing data. This overload also, occurs when the process of reading and writing in the database but some databases are capable of scaling.

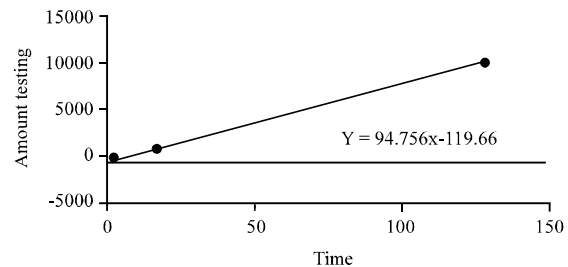


Fig. 8: Architecture performance

**Testing of performance:** To measure performance performed calculations with some data in the form of time required during the process of data available until all the data is delivered to the subscriber. From the above values can be made a graph with a linear analysis approach from the testing performance. The linear equation is close to the true value, so, it can be predicted for values >10000 (Table 1; Fig. 8 and 9):

$$y = 94.756x - 119.66 \tag{1}$$

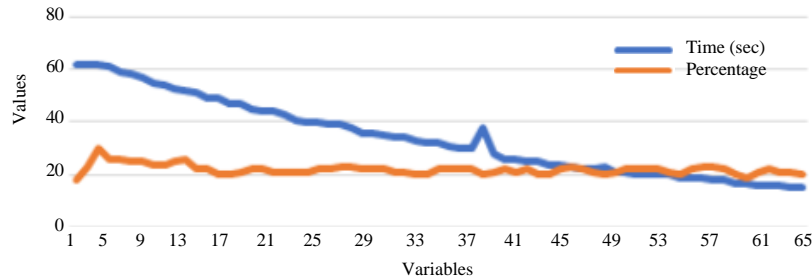


Fig. 9: Computer system performance (CPU performance)

Table 1: Testing performance

Total	Time
10	1 msec
100	1.2 sec
1000	13.5 sec
10000	01:46.6

Suppose for the value 50000 can be inserted into the value of, so that, the value of x becomes:

$$x = \frac{y-119.66}{94756}$$

$$x = \frac{50000-119.66}{94756} = 526.41s$$

With 50000 data can be processed with the time of 8.77 min.

**CONCLUSION**

Based on the tests that have been carried out on the system built it can be concluded that the following architecture can be prepared for big data management.

**REFERENCES**

Ayanoglu, E., Y. Aytas and D. Nahum, 2016. Mastering Rabbit MQ. Packt Publishing Ltd., Birmingham, UK., ISBN:978-1-78398-152-6, Pages: 232.

Celesti, A., A. Galletta, L. Carnevale, M. Fazio and A. Lay-Ekuakille *et al.*, 2018. An IoT cloud system for traffic monitoring and vehicular accidents prevention based on mobile sensor data processing. IEEE. Sens. J., 18: 4795-4802.

Duarte, P.H., L.F. Faina, L.J. Camargos, L.B. de Paula and R. Pasquini, 2016. An architecture for monitoring and improving public transportation systems. Proceedings of the 30th International Conference on Advanced Information Networking and Applications (AINA), March 23-25, 2016, IEEE, Crans-Montana, Switzerland, ISBN:978-1-5090-1857-4, pp: 871-878.

Hallgrimsdottir, B., H. Wennberg, H. Svensson and A. Stahl, 2016. Implementation of accessibility policy in municipal transport planning-Progression and regression in Sweden between 2004 and 2014. Transp. Policy, 49: 196-205.

Kang, Y.S., I.H. Park, J. Rhee and Y.H. Lee, 2016. MongoDB-based repository design for IoT-generated RFID/sensor big data. IEEE. Sens. J., 16: 485-497.

Li, J.Q., F.R. Yu, G. Deng, C. Luo and Z. Ming *et al.*, 2017. Industrial internet: A survey on the enabling technologies, applications and challenges. IEEE. Commun. Surv. Tutorials, 19: 1504-1526.

Lin, J., W. Yu, N. Zhang, X. Yang and H. Zhang *et al.*, 2017. A survey on internet of things: Architecture, enabling technologies, security and privacy and applications. IEEE. Internet Things J., 4: 1125-1142.

Mohanty, S.P., U. Choppali and E. Kougianos, 2016. Everything you wanted to know about smart cities: The internet of things is the backbone. IEEE. Consum. Electron. Mag., 5: 60-70.

Murad, D.F., B.S. Abbas, A. Trisetyarso, W. Suparta and C.H. Kang, 2018. Development of smart public transportation system in Jakarta City based on integrated IoT platform. Proceedings of the International Conference on Information and Communications Technology (ICOIACT), March 6-7, 2018, IEEE, Yogyakarta, Indonesia, ISBN:978-1-5386-0955-2, pp: 872-878.

Nguyen, V.S., Q.D. Pham and M.H. Ha, 2016. Improving the connectivity of a bus system: A case study of Ho Chi Minh city. Proceedings of the 7th International Symposium on Information and Communication Technology (SoICT '16), December 8-9, 2016, ACM, New York, USA., ISBN:978-1-4503-4815-7, pp: 73-78.

Puiu, D., S. Bischof, B. Serbanescu, S. Nechifor and J. Parreira *et al.*, 2017. A public transportation journey planner enabled by IoT data analytics. Proceedings of the International Conference on Innovations in Clouds, Internet and Networks (ICIN), March 7-9, 2017, IEEE, Paris, France, ISBN:978-1-5090-3673-8, pp: 355-359.

- Roy, M.G., 2018. RabbitMQ in Depth. Manning Publications Company, New York, USA., ISBN:9781617291005, Pages: 264.
- Samaras, P., A. Fachantidis, G. Tsoumakas and I. Vlahavas, 2015. A prediction model of passenger demand using AVL and APC data from a bus fleet. Proceedings of the 19th Panhellenic Conference on Informatics (PCI '15), October 1-3, 2015, ACM, New York, USA., ISBN:978-1-4503-3551-5, pp: 129-134.
- Tanaka, M., T. Kimata and T. Arai, 2016. Estimation of passenger origin-destination matrices and efficiency evaluation of public transportation. Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), July 10-14, 2016, IEEE, Kumamoto, Japan, ISBN:978-1-4673-8986-0, pp: 1146-1150.
- Zhu, M., X.Y. Liu, F. Tang, M. Qiu and R. Shen *et al.*, 2016. Public vehicles for future urban transportation. IEEE. Trans. Intell. Transp. Syst., 17: 3344-3353.