

## Research Article

# Reducing the Amount of Data for Creating Routes in a Dynamic DTN via Wi-Fi on the Basis of Static Data

**Yulia Shichkina<sup>1,2</sup> and Alexander Koblov<sup>1</sup>**

<sup>1</sup>*Saint Petersburg Electrotechnical University "LETI", Ul. Professora Popova 5, St. Petersburg 197376, Russia*

<sup>2</sup>*Saint Petersburg State University, Universitetskaya Nab. 7-9, St. Petersburg 199034, Russia*

Correspondence should be addressed to Yulia Shichkina; [strange.y@mail.ru](mailto:strange.y@mail.ru)

Received 22 March 2017; Accepted 18 June 2017; Published 26 July 2017

Academic Editor: Rui Zhang

Copyright © 2017 Yulia Shichkina and Alexander Koblov. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article presents the results of research on the acceleration of data processing in the construction routing in dynamic DTN, realized through the devices established in moving objects and connected by means of Wi-Fi. The routes are constructed based on the prehistory of the movement of objects, which is stored in the database in the csv format. This database has a large volume and contains unnecessary and incorrect information. The article shows the possibility of speeding up the process of constructing routes using data precleaning. Reducing data volume is proposed by constructing an enlarged grid on the object movement map. This article also demonstrates the use of fuzzy logic and graph theory for constructing routes.

## 1. Introduction

Analysts predict that, by 2020, wireless network traffic will much exceed the traffic on the wired Ethernet. Wi-Fi technology is one of the most important perspectives today in the field of computer communications. Wi-Fi is the ability to transfer data between devices for short distances without wires. Devices connected via wireless technology form a network. One of the characteristics of the efficiency of such a network is the technology of organizing data transmission in the network, routing.

This article describes the case of constructing a dynamic DTN between devices that are in motion and communicating with each other via Wi-Fi. The authors create a routing system based on information about previous movements of devices using graph theory.

Devices that communicate via Wi-Fi are installed on vehicles moving in a given area. Data on the movement of vehicles are taken from the database Glonass. The original format is csv; an example of a record is shown in Box 1.

The main task is to generate rules for transferring data in a DTN from an arbitrary device A to an arbitrary device B via Wi-Fi with a range of 100 meters.

Further, it will be assumed that data transmission should be carried out only through channels with an 80% probability of delivery. Since the network is constructed between moving devices and from this 100% probability of delivery along one chosen route is lost and the flood-fill algorithm is excluded due to network overload, then several alternative routes should be created.

The main task is thus divided into several subtasks:

(1) The preliminary data preparation, which includes selection of a database model for accelerated data processing, transfer of data from the csv format to the format of the selected database, data cleansing, and other operations on data, contributing to a faster and better solution of the original problem

(2) The construction of routes using methods of fuzzy logic and graph theory.

The main emphasis in this article is made on presenting the results of solving the first problem: reducing the initial volume of data. The second task is also solved by the authors, but in this article, only a little is described about how routes on the reduced data set are constructed.

The remainder of the paper is organized as follows. Section 2 provides an overview of the results of other scientists

*Names of Fields*

rxDevice, fileId, txDevice, gentime, txRandom, msgCount, dsecond, latitude, longitude, elevation, speed, heading, ax, ay, az, yawrate, pathCount, radiusOfCurve, confidence

*Example of a Record*

1566, 1158354, 1566, 291864241884364, -9018, 19, 35800, 42.41119, -83.874352, 269.10001, 18.280001, 176.1125, 9.9999998E - 3, 0.56999999, 0.0, 1.86, 9.0, 1829.0, 95

Box 1: Field of names and example of record in the original database.

in related fields of research. At the end of this section, the results of the closest research and difference between our approach and it are presented. Section 3 contains the formulation of the problem, the purpose of the research, and the main stages of the solution. Sections 4–7 describe the stages in the solution of the problem. Section 9 presents the results of the work.

## 2. Related Works

A lot of research has been devoted to the development of dynamic networks and routing methods in them, in particular, in the last decade. All the conducted research can be divided into the following classes:

- (1) Methods of data compression and encryption: as the wireless network has limited bandwidth and insecure shared media, the data compression and encryption are very useful for the broadcasting transportation of Big Data in IoT (Internet of Things). For example, a new combined parallel algorithm named “CZ algorithm” which can compress and encrypt the Big Data efficiently can be found in paper [1]. Paper [2] describes several broadcast encryption schemes from the point of view of wireless sensor networks.
- (2) Methods for finding the shortest paths for creating an effective routing system in a network.

In existing networks, as a rule, a number of traditional highly specialized routing algorithms are used. The main feature of these algorithms is the estimation of the cost of the route based on the set of metrics used in the protocol.

The task of providing the required quality of service in data transmission networks is solved by introducing rules for processing the streams of transmitted information in accordance with a certain type of traffic. However, setting static rules for certain data streams in modern converged networks is not effective enough [3, 4].

To solve this problem, researchers at Stanford University developed a concept of software-configurable networks. It is based on the use of OpenFlow technology, which allows more flexible network control by analyzing the data streams passing through the switches. Dynamic rules, that redistribute data streams to less loaded routes according to specified conditions, are set on specialized OpenFlow controllers. This architecture also allows collecting information about the network status in the critical nodes [5, 6].

The solutions offered by OpenFlow developers use the approaches inherent in traditional routing protocols, adapting and dynamically changing the rules tables for composing new routes. This approach takes into account the convergence of the network but still does not allow classifying the streams of transmitted data according to the given criteria (type of application or service, amount of data transferred, bandwidth, required bandwidth, delay, etc.). Besides, this approach does not provide the ability of load balancing in communication channels and between network devices. In addition to these problems, modern routing algorithms should solve the problem of increasing the reliability and performance of critical network segments when transmitting real-time information, such as video data streams or VOIP services, and others.

But, one thing is the devices being static; another thing is the devices being, 80% of their time, in motion. In this case, the number of problems becomes even greater.

Static routing algorithms, in contrast to dynamic routing, do not take into account the constantly changing topology of the network. This makes them unsuitable for use in most networks.

Static routing algorithms, in contrast to dynamic routing algorithms, do not take into account the constantly changing topology of the network. This will make it unsuitable for use in most networks.

All routing algorithms use one of three mathematical models: Dijkstra’s model, Bellman-Ford’s model, and Floyd-Warshall’s model. However, if static algorithms apply them for the entire described subnet, then the dynamic algorithms are only local, using the developed metrics of optimality.

Flood-fill algorithm is the most reliable and fastest [7] of all existing algorithms. The principle of functioning is to send the incoming packet to all the lines except the one on which it came. But its only and main disadvantage is unacceptably high value of traffic. This algorithm is evaluative when testing new developments and it is still used in specialized networks (e.g., military).

The stream-based routing algorithm uses the assumption that traffic within the network can be described by some statistical law, on the basis of which the optimal route schemes are selected [7].

One of the most advanced algorithms now for routing is routing with the account of the state of the lines. The metric for this algorithm is the average delay value for the test packet, which reflects not only the length of the route, but also the load of the channel. The practical implementation of this algorithm is the OSPF protocol.

Paper [8] solved the problem of dynamically updating all-pairs shortest paths in a distributed network while edge update operations occur in the network.

Multihop data delivery between vehicles is an important technique to support the implementation of vehicular ad hoc networks (VANETs). However, many inherent characteristics of VANETs (e.g., dynamic network topology) bring great challenges to the data delivery. In particular, dynamic topology and intermittent connectivity make it difficult to design an efficient and stable geographic routing protocol for different applications of VANETs. To solve this problem, paper [9] proposes an adaptive routing protocol based on QoS and vehicular density (ARP-QD) in urban VANETs environments. Generally, path efficiency and path stability are two important criteria in designing routing protocol for VANETs. To achieve high efficiency, authors of article [9] selected the shortest path with minimum hop count as the best path. In contrast to the authors of paper [9] we, in our research, put the reliability of delivery in the first place. Another difference between our routing method and the method proposed in [9] is its basing on static information about the movement of vehicles. And also data is transferred via Wi-Fi.

Paper [10] described Greedy Perimeter Stateless Routing (GPSR) algorithm, which uses the positions of routers and a packet's destination to make packet forwarding decisions.

Another method similar to GPSR is Greedy Perimeter Coordinator Routing (GPCR) [11], which assigns the routing decision to the nodes located at the street intersections and uses the greedy forwarding strategy to route the packet path between the street intersections.

Another system of wireless mobile nodes is mobile ad hoc network, which can freely and dynamically self-organize network topologies without any preexisting communication infrastructure. Due to characteristics like temporary topology and absence of centralized authority, routing is one of the major issues in ad hoc networks. In paper [12], a new multipath routing scheme is proposed by employing simulated annealing approach.

Recently, a few more routing schemes for ad hoc networks have been presented in research literature. In particular, these articles describe

- (i) multicast power greedy clustering (MPGC) scheme [13], which is an adaptive power-aware and on-demand multicasting algorithm;
- (ii) double-layered effective routing (DLER) scheme for peer-to-peer network systems [14];
- (iii) entropy-based multirate routing (EMRR) scheme [15];
- (iv) ant-colony based routing algorithm (ARA) scheme [16];
- (v) genetic algorithm based routing (GAR) scheme [17].

Interesting results of the research are presented in article [18]. The authors propose an anonymous and nonintrusive geographic routing protocol, GeoVanet, which ensures that the sender of a query can get a consistent answer.

Since in our studies we used Yen's algorithm, it should be noted that in the last decade new algorithms for finding shortest paths in the graph appeared [19, 20]. However, due to the fact that the ultimate objective of this research was to show how it is possible to reduce the amount of data, used to create a route, then in this article we do not consider the problem of optimizing methods for finding shortest paths.

- (3) Protocols for data transmission in mobile networks: the proliferation of mobile devices has changed the trends by which people access and share contents in the communication world, leading to migration from the wired to wireless networks with high expectations of ubiquitous connectivity. These trends have motivated researchers to have interests in the future advanced mobile communications, such as MANET [21]. In paper [22], the authors propose a self-aware network architecture in which protocol stacks can be built dynamically. Those protocol stacks can be optimized continuously during communication according to the current requirements.
- (4) Methods of storing data in dynamic networks: in paper [23] authors propose a replica-based data-storage mechanism and undelivered-message queue schemes to provide reliable data storage and dissemination. They also propose replica update strategies to maintain data consistency while improving data accessibility. These solutions are based on a clustered MANET where nodes in the network are divided into small groups that are suitable for localized data management.

The problem associated with a large volume of data was discussed in the literature, in two aspects.

(1) *Reducing the Data Volume Transmitted in Wireless Sensor Networks (WSNs)*. For example, in [24] a solution to the problem of transferring large volume of data in the WSN used to monitor the structural state of underground subway tunnels is discussed. Being considered as a feasible solution to these issues, data compression can reduce the volume of data travelling between sensor nodes. In paper [24], an optimization algorithm based on the spatial and temporal data compression is proposed to cope with these issues appearing in WSNs in the underground tunnel environment.

A variety of data compression approaches appeared in the literature: a distributed data compression approach [25] and a local data compression approach [26].

Network performance can be improved in different network conditions using the caching procedure [27].

In paper [27], the authors propose a service-customized 5G network architecture by introducing the ideas of separation between control plane and data plane, in-network caching, and Big Data processing and analysis to resolve the problems that traditional cellular radio networks face.

But, in our case, data compression or caching procedures will not help speed up the transfer of information. In our case, processing of data of large volume occurs before transmission of information.

Research, very similar to our research, is discussed in [28]. In this article, the authors suggest routing protocol for cognitive radio vehicular ad hoc networks (CR-VANETs) using a Kalman filter algorithm. The protocol first selects an idle channel from among all the channels available to a vehicle while moving on a straight road and then finds the best relay node to deliver the packet to the destination. The selection of a relay node is done by dividing the vehicular transmission range into five regions, and then the source/relay node selects the one that is in the region having a higher preference than other regions. Communication between two vehicles occurs only when both the vehicles are on the same channel. Delay may increase while sensing the channel and selecting the relay node.

To reduce the delay, authors use a Kalman filter to predict the future positions of all moving vehicles in the network.

But, the authors of [28] apply their methods to the original data set. This data set can have a very large volume. In this article we show how to reduce this volume. And already it is possible to apply both the method described in [29] and various graph methods and methods based on fuzzy logic to reduce the volume. All the methods listed above will work much faster if the source data is much smaller.

(2) *The Problem of Processing a Large Volume of Data in Real Time.* The literature [29–31] reveals some important problems that need to be solved when processing large data. It is, for example, designing a generic communication model, real-time Big Data analytics, and acquisition of data from sensors. In paper [30], authors present the efficient and generic communication model for future smart cities based on Big Data analytics and integration of WSN. The MapReduce paradigm is used for the data analysis, while manipulation and storing are performed by Hadoop distributed file system (HDFS), HBASE, and HIVE.

The proposed scheme stores and processes data in Hadoop framework. MapReduce has been selected as the mechanism for analyzing filtered data. MapReduce works in two steps. First is the mapping process where the set of filtered data is converted into another set of data. Next is the reduction process which combines the data created in mapping process and results in a set of values that are reduced in amount.

The difference between our technology and the technology of the authors [30] is that our approach allows us to reduce once again a set of data which has already been reduced with the help of fuzzy logic in our case or MapReduce in the case [30].

### 3. The Formulation of the Problem and the Stages of Its Solution

A database with vehicle location information contains a very large amount of data. So only for 2 months it contains 267,163,140 records on 19 attributes. A preliminary analysis of these data showed that more than 50,000 vehicles will be involved in routing in a limited area. When creating routing based on these data using the theory of graphs, the graph is

very large, and the time spent searching for several alternative paths ranked in descending probability of delivery is very long.

In connection with this, one of the research objectives was to reduce the amount of initial data and to develop an approach to accelerate the processing of the graph model.

In accordance with this objective, the following tasks were solved, the results of which are described in this article:

- (1) Selection of a database, queries to which can be successfully paralleled on high-performance computing systems (Section 4.1)
- (2) Importing data from the source format to the format of the selected database (Section 4.2)
- (3) Reduction of the initial volume of data by removing unnecessary attributes and incorrect information (Section 5)
- (4) Creation of a grid on the coordinate map for the construction of a graph of communications between vehicles with a minimum set of vertices (Section 6)
- (5) The second iteration of reducing the amount of initial data using fuzzy logic. As a result, only those vehicles remain, the probability of delivery through which is above a certain value, in our case 0.8 (Section 7)
- (6) Construction of an optimal route between grid cells (Section 8)
- (7) Construction of a given number of alternative data transmission routes between vehicles along the route found between the grid cells (Section 8).

## 4. Preliminary Preparation of Data

*4.1. Choosing a Database.* The most widely spread SQL databases are undesirable for this task, since their performance noticeably decreases with the increase in the amount of information.

The main advantage of NoSQL databases is their mass scalability. NoSQL databases have a large storage space, easily cope with large volumes of data, and have predictable performance (1.5–2 times greater than that of SQL).

Comparative analysis of free distributed database management systems MySQL and MongoDB showed that the running time of MySQL on large data sets is much lower than MongoDB (Figure 1). The experiments were done using a special library that allows processing data in several parallel streams

It is also noted that the performance of the MySQL relational database with four threads on large data sets is comparable to the performance of MongoDB, which processes the data sequentially.

To solve the task NoSQL database MongoDB 3.2 was selected under the Ubuntu Server 14.04 LTS operating system.

*4.2. Importing Data into a Database.* According to the proposed documentation for the Safety Pilot Model Deployment (SPMD) program, the following data fields were selected for



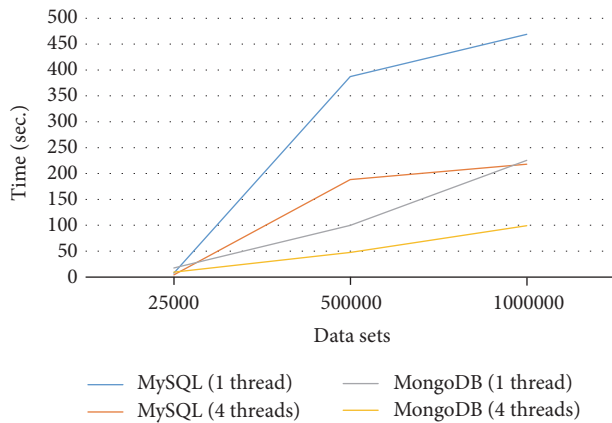


FIGURE 1: The graph of the dependence of the execution time on the number of records in the data set.

the BsmPI data set and the command «mongoimport» was applied to them:

```
mongoimport -db moto -collection traffic --fields rxDevice, fileId, txDevice, gentime, txRandom, msgCount, dsecond, longitude, latitude, elevation, speed, heading, ax, ay, az, yawrate, pathCount, radiusOfCurve, confidence --type csv --numInsertionWorkers 40 --file april.csv
```

As a result of this operation with the files of the test database with data for April and October 2014, about 2,267,163,140 records from the original test database files were imported to the MongoDB database.

## 5. The Method of Extracting Knowledge from Databases: The Concept of Creating a Grid

The Knowledge Discovery in Databases (KDD) method originated in the second half of the 20th century and did not describe specific algorithms but described a certain sequence of actions that must be followed to extract useful knowledge.

First, the initial data must be obtained from the data sources. Often such data should not simply be collected but consolidated from several sources. After the data is received and placed in some storage, it is necessary to clear it of «junk» data (data that does not have value for obtaining useful knowledge). In the third stage the cleared data must be subjected to transformation. Transformation is the preparation of cleaned data for further analysis, because many of the methods require a certain data format for their application. The fourth stage is Data Mining, the process of extracting useful knowledge from the initial data. At the last stage, interpretation of the received knowledge is carried out.

When it comes to large data, the KDD approach can be applied with some modifications; namely, it is necessary to go through the steps of cleaning and transforming data repeatedly. This approach allows reducing the amount of data and therefore more quickly and efficiently using methods of knowledge extraction.

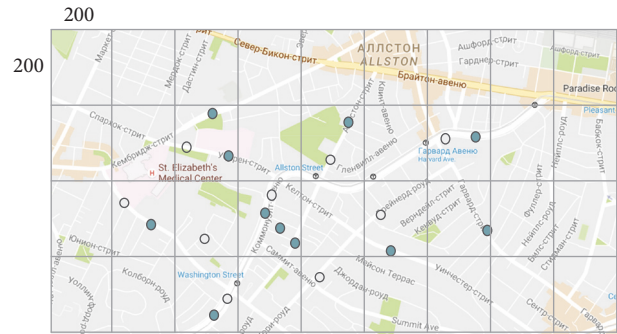


FIGURE 2: Type of coordinate grid.

The data set received in BsmPI is redundant to solve the task. The data set contains 13 types of data reflecting various attributes of vehicles.

The following 19 attributes are described in the BsmPI file: RxDevice, TxDevice, Gentime, TxRandom, MsgCount, Dsecond, Latitude, Longitude, Elevation, Speed, Heading, Ax, Ay, Az, Yawrate, PathCount, RadiusOfCurve, and Confidence.

Only the information about device identifiers (TxDevice), geodata (longitude and latitude), and time (Gentime) is needed from the entire set of data for constructing routes. Thus, only 3 fields will remain from 19 initial fields after cleaning. From these fields, the field «Gentime» must be converted into two other fields: «TimeZone» and «Weekday». As a result, instead of 19 fields of the initial database, only 5 fields will be included in the new database, which will contain «useful» data.

Also, it is necessary to exclude incorrect records from the processing, for example, records with coordinates corresponding to geographical poles or located at a considerable distance from densely located data. Such records could be obtained with incorrect operation of the geopositioning devices.

## 6. Construction of a Grid on a Coordinate Map

**6.1. The Concept of a Grid.** The main idea of simplifying the processing of data and speeding up the process of accessing data is grouping data on a common feature. Since for the BsmPI data set the common parameters (which do not depend on other parameters) are the coordinates and time, it will be logical to combine the data into groups by time and location.

The simplest and at the same time the right solution will be the grouping of records for each section of space in which the devices will be at a given time, for example, dividing the map into rectangles or squares due to the two-dimensionality of the coordinates received from the GPS satellite. Such a partition will be called a coordinate grid (Figure 2).

Initially, the route is constructed between cells of coordinate grid. In this way, we get a route that will connect the set of devices localized in the corresponding coordinate grid cells. Such a route will be called an enlarged route. According

to the created enlarged route, in the next step, several final information transmission routes will be constructed, which will be ranked in descending order of probability of information delivery. This approach allows speeding up the data processing by reducing the amount of records in the database for the construction of an enlarged route.

It is necessary to choose a unit measure to create a coordinate grid on a map, for example, meter. And it would all be much simpler if the degrees of longitude and latitude were uniquely translated into a metric system. As it is known, the length of the meridian is constant, and, proceeding from this, it is possible to translate uniquely the required unit into degrees of latitude. For longitude it is more difficult, since the length of 1 degree in meters varies depending on the coordinate.

If accurate calculations are performed when creating a two-dimensional coordinate grid, then, for each record processing it is necessary to make calculations of length of the meter, depending on the coordinates of the object, and all these data must be stored in a separate repository. This approach requires additional storage and access to these repositories. Therefore, for localized creation of a coordinate grid within a single city or region, it is permissible to use an approximate mean value, based on the difference of the extreme coordinates divided by two.

The grid step can be selected depending on the final destination of the obtained data set.

For accelerated navigation on the grid, field «nSquare» should be added to the database, which is the unique identifier of the grid cell.

Turning to the grouping by time, it should be noted that the load of certain areas by different vehicles, on which it is possible to collect telemetry of geolocation, is sufficiently periodic with respect to the days of the week.

Based on the periodicity of the data it makes sense to divide the entire data set into 7 parts by the days of the week and work with each new set separately. In the future, this approach will allow distributing the processing of newly received data on a cluster of several machines, each of which will process a smaller part of the data, for example, only one of the days of the week.

It also makes sense to enter one more time coordinate in addition to the day of the week, the time zone of the day. The choice of such a coordinate is due not only to the simplification of navigation along the time axis, but also to smoothing out the errors in the location of the device in a certain coordinate grid cell.

In this case, after the transformation of the data, a new data structure will be obtained. In this structure the field «loc» contains the coordinates of the device, the field «fileId» contains the ID of device «RxDevice», field «daytime» contains the number of full minutes passed since the beginning of the day, field «weekday» is the day number of the week, field «timeZone» is the number of the time zone of the day, field «nSquare» is the unique identifier of the grid cell, and the field «Square» contains the coordinates of the cell in the grid.

After the first transformation, the number of records did not decrease, but the number of fields changed. As a result,

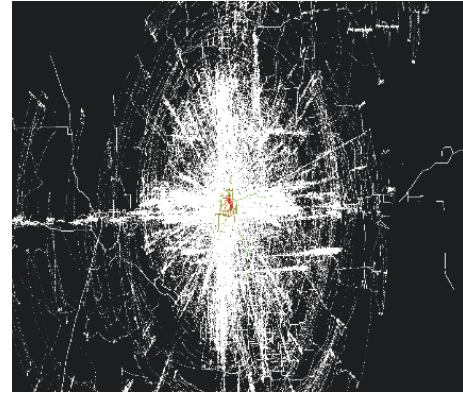


FIGURE 3: Visualization of traffic density of vehicles with devices for data transmission.

only the necessary and useful data are included in the new database. This data set will be called the preliminary grid (preMesh).

In the future, above this data set, actions will be performed for deep data cleaning using fuzzy sets, which significantly will allow reducing the amount of data for accelerated application of knowledge extraction methods.

In our case the similar data set allowed parallelizing the task of constructing routes to 7 computational nodes, one data set for each day of the week.

**6.2. Visualization of Received Data.** To estimate received data and make a decision on the further appropriateness of using knowledge extraction methods, a work was carried out to obtain a graphical representation of the data. This process was implemented using a software module written in PHP 5.6. The name of the collection in MongoDB and the resulting maximum value of the devices or records in the grid cells were input data of the visualization function.

Figure 3 is an example of the resulting pregrid image for Monday. During the analysis it was determined that, in the center, where the most loaded grid cells are located, there is such a large city as Indianapolis.

The resulting files are limited to the grid's extreme coordinates for a particular set. Each point of the image is a separate sector plotted on the grid.

With the help of the images we can understand where and how many records we have and in which directions we can make the delivery of messages.

## 7. Using Fuzzy Logic to Reduce Data in a Grid

Most of the data processed in modern information systems is of a numerical nature. However, in queries to databases that a person tries to formulate, there are often inaccuracies and uncertainties.

Fuzzy slices are a good example of the addition of one technology (fuzzy logic) to another (database). Under fuzzy slice, filters on measurements are understood, which include fuzzy values, for example, "all objects moving to the north of the city." In this example, the term "north of the city" is fuzzy.



Number of records before the screening		Number of records after the screening
387504024	preMesh → New grid	1018
396409188	preMesh → New grid	988
360181158	preMesh → New grid	1124
334144522	preMesh → New grid	1137
342934063	preMesh → New grid	1078
241741346	preMesh → New grid	924
204248857	preMesh → New grid	820

FIGURE 7: Decrease in the number of records after the screening of unreliable vehicles.

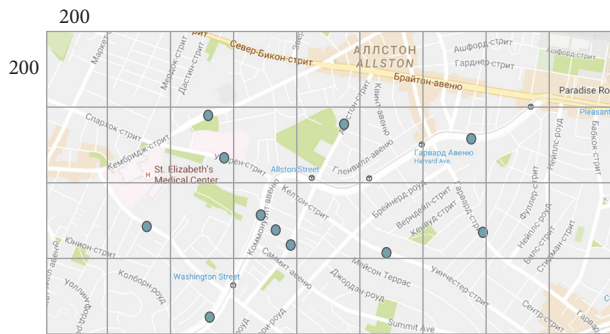


FIGURE 8: Grid after removing information about the devices from it, the probability of data transfer through which is small.

## 8. Further Work on the Creation of Routes for Data Transfer Using Devices Installed on Vehicles

Correspondingly, a connected graph was constructed between the cells of the coordinate grid. This graph is constructed taking into account possible directions of information transmission (Figure 9).

Later, the depth-search algorithm and the modified Yen's algorithm for creating routes were applied to the resulting graph. The first of them receives JSON-lines from the database and creates the main route according to the chosen weights. With the help of the modified Yen's algorithm, no channels of communication are deleted and nodes and alternative routes are created.

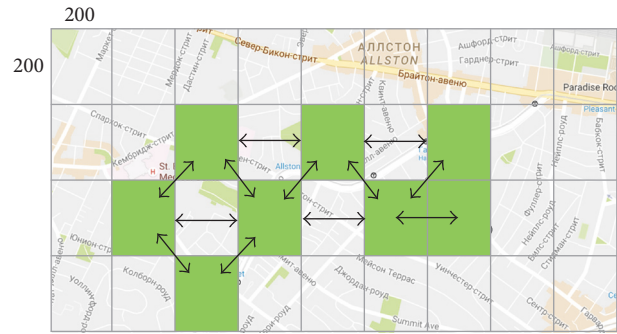


FIGURE 9: A connected graph constructed from an enlarged grid.

The proposed approach to data transformation and data clearing allowed reducing, in our case, the number of records more than 256,000-fold.

Such a significant reduction in the number of records allows not only reducing the cost of storing the prepared data, but also greatly accelerating the application of methods of extracting knowledge. And with the use of the library «pthreads», PHP applications can use the processor time carefully and get a gain in the execution time of the solution of the task of preliminary cleaning and transformation of data more than 2 times on a sufficiently weak processor.

## 9. Conclusion

In today's world, the volume of data produced and processed inevitably grows, and the question of rational use of available resources for solving applied problems, especially in real time, becomes relevant.



This article presents the results of studies on the construction of routes using fuzzy logic and graph theory. How, with the help of fuzzy logic, it is possible to significantly reduce the data volume on the prehistory of vehicle movement is shown. How it is possible to create routes in two stages is demonstrated, first, to create a route between the sets of nodes and, then, to create an updated route already directly between devices.

The approach to routing which is considered in the article is not suitable for vehicles moving constantly along new routes. But, with the help of this approach, it is possible to exclude those vehicles that do not have a relatively permanent route.

It is also shown that the use of classic relational databases, such as MySQL, is only rational for small projects in which there are no large amounts of data or in which there is no need to quickly obtain a result. However, setting the purpose of qualitative data processing, we inevitably come to increase the sample to increase the representativeness. But, when increasing volumes, do not forget about the relevance, which data can lose due to too long processing. Therefore, in order to satisfy the need for these criteria, the use of NoSQL databases as the most suitable for processing large data will be expedient. This was shown in the example of MongoDB.

Considered in the article, the approach to the construction of routes can be adapted to vehicles moving constantly along new routes. After the grid is created based on the history of the movement of vehicles, new data can be added to it. But, these data must be preprocessed using the same methods of fuzzy logic or other methods that guarantee a certain probability of information delivery using these vehicles. In the graph for grid cells, then weights will change. But due to the fact that the graph is small, recalculation of the enlarged path through the grid cells does not take much time.

With the proposed approach of data cleansing using fuzzy logic, it is also possible to filter out those vehicles that do not have a relatively constant route. This is another way to solve the problem of data transfer using vehicles.

Summing up, we can say the following: the described technology allows significantly speeding up the process of constructing a data transmission route using moving objects due to the reduction of the initial data and a very significant reduction in the graph of relationships between objects.

Research is not complete. At the present time, additional research, as well as other studies, is being done to expand the area of transportation of vehicles, to select the set of parameters most influencing the quality of information delivery, to select protocols for transferring information, and to take into account the architectural features of devices on vehicles.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This paper has been prepared within the scope of the state project "Initiative Scientific Project" of the main part of the state plan of the Ministry of Education and Science of Russian Federation (Task no. 2.6553.2017/BCH Basic Part).

## References

- [1] Q. Jiancheng, L. Yiqin, and Z. Yu, "Parallel algorithm for wireless data compression and encryption," *Journal of Sensors*, vol. 2017, Article ID 4209397, 11 pages, 2017.
- [2] P. Szalachowski and Z. Kotulski, "One-time broadcast encryption schemes in distributed sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 536718, 9 pages, 2012.
- [3] I. Bolodurina, D. Parfenov, and A. Shukhman, "Approach to the effective controlling cloud computing resources in data centers for providing multimedia services," in *Proceedings of the International Siberian Conference on Control and Communications, SIBCON 2015*, pp. 1–6, May 2015.
- [4] P. Charuenporn and S. Intakosum, "Qos-security metrics based on ITIL and COBIT standard for measurement web services," *Journal of Universal Computer Science*, vol. 18, no. 6, pp. 775–797, 2012.
- [5] F. Dandria, S. Bocconi, J. G. Cruz, J. Ahtes, and D. Zeginis, "Cloud4SOA: multi-cloud application management across paas offerings," in *Proceedings of the 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2012*, pp. 407–414, September 2012.
- [6] J. Zhu, Z. Zheng, and M. R. Lyu, "DR2: dynamic request routing for tolerating latency variability in online cloud applications," in *Proceedings of the IEEE 6th International Conference on Cloud Computing, CLOUD 2013*, pp. 589–596, July 2013.
- [7] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, Prentice Hall Press Cloth, 5th edition, 2011.
- [8] M. R. Khoudja, E.-G. Talbi, L. Jourdan, B. Sarasola, and E. Alba, "Multi-environmental cooperative parallel metaheuristics for solving dynamic optimization problems," *Journal of Supercomputing*, vol. 63, no. 3, pp. 836–853, 2013.
- [9] Y. Sun, S. Luo, Q. Dai, and Y. Ji, "An adaptive routing protocol based on QoS and vehicular density in urban VANETs," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 631092, 13 pages, 2015.
- [10] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 243–254, August 2000.
- [11] C. Lochert, M. Mauve, H. Füßler, and H. Hartenstein, "Geographic routing in city scenarios," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 1, pp. 69–72, 2005.
- [12] S. Kim, "Adaptive MANET multipath routing algorithm based on the simulated annealing approach," *Scientific World Journal*, vol. 2014, Article ID 872526, 8 pages, 2014.
- [13] J. J. Leu, M. H. Tsai, C. Tzu-Chiang et al., "Adaptive power aware clustering and multicasting protocol for mobile ad-hoc networks," in *Ubiquitous Intelligence and Computing*, pp. 331–340, 2006.
- [14] J. Kim, K. Lee, T. Kim, and S. Yang, "Effective routing schemes for double-layered peer-to-peer systems in MANET," *Journal of Computing Science and Engineering*, vol. 5, no. 1, pp. 19–31, 2011.

- [15] C. T. Hieu and C. Hong, "A connection entropy-based multi-rate routing protocol for mobile Ad Hoc networks," *Journal of Computing Science and Engineering*, vol. 4, no. 3, pp. 225–239, 2010.
- [16] M. Günes, U. Sorges, and I. Bouazizi, "ARA—the ant-colony based routing algorithm for MANETs," in *Proceedings of the International Conference on Parallel Processing Workshops*, pp. 79–85, British Columbia, Canada, August 2002.
- [17] L. Barolli, A. Koyama, T. Suganuma, and N. Shiratori, "GAMAN: a GA based QoS routing method for mobile ad hoc networks," *Journal of Interconnection Networks*, vol. 4, no. 3, pp. 251–270, 2003.
- [18] T. Delot, N. Mitton, S. Ilarri, and T. Hien, "GeoVanet: a routing protocol for query processing in vehicular networks," *Mobile Information Systems*, vol. 7, no. 4, pp. 329–359, 2011.
- [19] M. Holzer, F. Schulz, and D. Wagner, "Engineering multilevel overlay graphs for shortest-path queries," *ACM Journal of Experimental Algorithmics*, vol. 13, article 5, pp. 1–26, 2008.
- [20] L. Roditty, "On the  $k$  shortest simple paths problem in weighted directed graphs," *SIAM Journal on Computing*, vol. 39, no. 6, pp. 2363–2376, 2010.
- [21] M. R. Bosunia, D. P. Jeong, C. Park, and S.-H. Jeong, "A new routing protocol with high energy efficiency and reliability for data delivery in mobile ad hoc networks," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 716436, 8 pages, 2015.
- [22] A. Keller, D. Borkmann, S. Neuhaus, and M. Happe, "Self-awareness in computer networks," *International Journal of Reconfigurable Computing*, vol. 2014, Article ID 692076, 16 pages, 2014.
- [23] M. K. Denko and H. Lu, "Replica dissemination and update strategies in cluster-based mobile ad hoc networks," *Mobile Information Systems*, vol. 2, no. 4, pp. 193–209, 2006.
- [24] B. He and Y. Li, "Big data reduction and optimization in sensor monitoring network," *Journal of Applied Mathematics*, vol. 2014, Article ID 294591, 8 pages, 2014.
- [25] A. Ciancio, S. Patten, A. Ortega, and B. Krishnamachari, "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks (IPSN '06)*, pp. 309–316, April 2006.
- [26] C. M. Sadler and M. Martonosi, "Data compression algorithms for energy-constrained devices in delay tolerant networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 265–278, ACM, 2006.
- [27] H. Yao, C. Fang, Y. Guo, and C. Zhao, "An optimal routing algorithm in service customized 5G networks," *Mobile Information Systems*, vol. 2016, Article ID 6146435, 7 pages, 2016.
- [28] H. Ghafoor and I. Koo, "Spectrum-aware geographic routing in cognitive vehicular ad hoc network using a Kalman filter," *Journal of Sensors*, vol. 2016, Article ID 8572601, 10 pages, 2016.
- [29] X. Li, W. Shu, M. Li, H.-Y. Huang, P.-E. Luo, and M.-Y. Wu, "Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1647–1653, 2009.
- [30] B. Nathali Silva, M. Khan, and K. Han, "Big data analytics embedded smart city architecture for performance enhancement through real-time data processing and decision-making," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9429676, 12 pages, 2017.
- [31] I. J. Lee, "Big data processing framework of road traffic collision using distributed CEP," in *Proceedings of the 16th Asia-Pacific Network Operations and Management Symposium (APNOMS '14)*, Hsinchu, Taiwan, September 2014.
- [32] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

