Delay Tolerant Network Potential in a Railway Network

Eugene Tikhonov, Donat Schneps-Schneppe
AbavaNet Ltd.
Moscow, Russia
eugene.tikhonov, donat.shneps-shneppe @abava.net

Dmitry Namiot
Lomonosov Moscow State University; RUT (MIIT)
Moscow, Russia
dnamiot@gmail.com

***Abstract*—The article is devoted to the analysis of the effectiveness of the DTN (Delay Tolerant Network) system in a communication network on a railway line. Trains act as moving objects that are sending and receiving messages to/from an external network. As well trains may transmit telemetry, collected during movement. The article presents data on effectiveness of the DTN network in terms of reducing the delay in message delivery, reducing the delay in transmitting telemetry, and increasing the total number of messages in the railway scenario. The data on the theoretical efficiency of the DTN system is presented, including the results of the generation of many maps of railway networks and the corresponding train schedules. An analysis of railway lines with various loads and different coverage of mobile network was made. These models are based on real railway maps and train schedules. Effectiveness of DTN during migration to networks of a higher data rate (5G) and networks with satellite connection of trains is discussed.**

I. INTRODUCTION

DTN (Delay Tolerant Network) as a part and the enhancement of the railway communication system is being analyzed in the article. Messages from a train to the server in the external fixed network and return messages to the train are transmitted over the radio-linked network. Depending on whether the train is in the coverage area of the radio network (the coverage area of the base stations, BS), the message can be delivered either almost instantly, or it can get a long delay: several minutes, tens of minutes or even more than an hour. If a train is out of the coverage area during an attempt to send a message, then this message is saved and then is sent to another train (oncoming or passing) when these trains are in each other's radio access zone. Also, when a train reaches the coverage area, a message is sent if it was not delivered yet through the retransmissions chain by another train, which arrived in the coverage area of the network earlier (see Fig. 1).

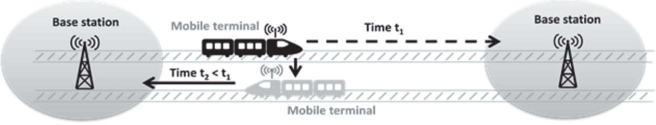

Fig. 1. Two possible data delivery ways example. Data is generated on the mobile terminal (black train) and could be delivered directly (dotted arrow) or partially via another mobile terminal (gray train) that is faster

Since message delivery can take quite a while, the article focuses on two types of messages (delay tolerant types):

- Subscriber/passenger messages (from a train to subscribers in an external network and, conversely, to subscribers/passengers on a train from subscriber in an external network).
- Telemetry data. The train collects information during the movement (about the state of the tracks, about the condition of the goods on board) and tries to send it to the network. The value of information is higher when the telemetry data is delivered faster.

The rest of the paper is organized as follows. In Section II we consider a number of related works and common DTN protocols, in Section III – mathematical model for evaluating the effectiveness of DTN and analysis of simulations of the DTN functioning on the set of synthetical maps of railway networks with the generated train schedules, in Section IV – we consider the increase in the efficiency of the existing network due to DTN on the example of a real section of the railway line with the real train schedule, in Section V – we consider the increase in the efficiency of existing sections of the railway when migrating to a communication network of a higher data rate and/or satellite Internet on some trains.

II. RELATED WORKS

A system of movable mobile objects that could interact with each other directly could be characterized as a mobile ad hoc network [1]. This wireless mesh network is dynamically self-organized and self-configured by peer-to-peer connected clients and stationary external infrastructure routers (base stations) and so could be classified as hybrid type network [2]. There is no full graph for communications (communication possibility is determined by proximity [3] and not all objects can directly communicate with each other), and intermediate relay nodes should store and forward packets [4]. Also, data should be delay and disruption tolerant; therefore we are dealing with DTN.

There are a number of works devoted to maximizing the effectiveness of DTN in various scenarios. The system tries to transmit the largest amount of data as quickly as possible. The general theoretical foundations for DTN networks are considered, for example, in [5]. The basic definitions and concepts (infrastructure objects, interfaces, load balancing methods, priorities and message classes) are proposed. The main difficulties arising in networks without constant stable communication between nodes are considered. The most general principle of DTN operation is hop-by-hop forwarding through intermediate nodes/objects/peers, which, in contrast to classical communication networks, provides relatively long-term intermediate data buffering.

A general categorization of possible transmission routing protocols was proposed in [6]. Various possible knowledge or reliable predictions of the future are divided into classes of "oracles". Examples of general algorithms are proposed for them. "First Contact" makes a random choice of transmission path and does not require any a priori data. "Minimum Expected Delay" selects the best route from alternatives according to statistical information on average delivery time using the Dijkstra shortest path in graph search algorithm. "Earliest Delivery" uses current knowledge of upcoming delivery times for each alternative. "Earliest Delivery with Local Queuing" and "Earliest Delivery with All Queuing" additionally balance the load of intermediate nodes at a meeting or in the entire system, respectively.

It is shown that it is theoretically possible to absolutely minimize the total delays using Linear Programming, but in practice, such a solution requires enormous computing resources and complete accurate knowledge of the future. Therefore, it is necessary to rely on not optimal, but more realistic algorithms.

A simple flooding routing protocol "Epidemic" was proposed in [7]. According to it, each node transmits to each other, if possible, all the data, which transmitter has and receiver doesn't have yet (taking into account the existing buffer limitations and transmission speed). With unlimited possibilities of storage and transmission resources, this algorithm provides the fastest delivery of all messages and therefore can be used as a reference to assess the potential for the implementation of DTN technologies, which was used in Sections IV and V.

Since the Epidemic protocol requires huge resources, numerous options for limited copying of information have been proposed with attempts to maintain comparable delivery speeds.

This is done mostly for models of wandering objects that do not have reliable information about upcoming events, but for which there are statistical patterns of movement. Thus, on the basis of the accumulated history, the possibilities and potential of objects for future retransmissions and propagation are estimated. The protocols "PRoPHET" [8], "MEED" [9], "EBR" [10] use a probabilistic assessment depending on statistics of past meetings.

"MaxProp" [11], developed on the basis of practical experiments on regional regular buses, additionally uses the history of specific messages (the number of hops already completed and current age of message) to prioritize the transmission and deletion of messages.

In [12], an approach was proposed for sending a message only if the expected delivery time of met node is better than the one of those to whom copies were sent earlier.

"DTC" [13] in addition to encounters statistics analysis uses physical communication parameters. The position of the destination node in the lists of the last encounters, often met objects, signal power level, re-detection interval (with user-specified parameter weights) are taken into account.

The direction of the object's movement can also serve as a criterion for making a decision about sending a copy: the message is not sent if the receiver's direction is within the specified limits with respect to the sender's direction - "Vector Routing" [14].

Some protocols rely on the restriction of flooding without the use of additionally collected information. In "Spray and Wait" [15], only a certain distributed number of copies is released. When they are distributed, the nodes simply expect to meet their destination. "Spray and Focus" [16] after the distribution of copies starts the forwarding procedure with deleting of copy from sender's buffer (if the receiver is more promising).

Some protocols rely on sending a message without saving a copy at the sender. "Contact Graph Routing" [17] involves considering all possible paths of upcoming transmissions with full recursion and, thus, is suitable only for a few contacts.

SABR protocol [18] relies on the choice of the best route at the last moment of decision making (the choice is made by Dijkstra's algorithm). Conflict resolution of a limited buffer is thus achieved by the principle of "who was the first to take." In addition, sending a higher priority message crowds out the existing ones in the queue - and for them a different initial choice might be preferable.

An important way to reduce the used buffers is to distribute between the nodes information about the message status and delete already delivered packets from the buffers [19].

In all these cases, it is assumed that the message is sent from one arbitrary node in the network to another arbitrary node, without the presence of dedicated permanent communication nodes. In the railway network, the typical scenario is different. Usually objects do not communicate with each other, but with an object/server in a fixed outer network.

Moreover, in addition to the offline zones (or the presence of weak coverage with insufficient transmission speed in relation to the generated data volumes), there are coverage areas with guaranteed communication. These zones can be considered as a single selected object with which all the others communicate.

In addition, routing protocols generally do not use the specifics of having a relatively predictable schedule of movements along fixed trajectories — rail lines.

Articles discuss mostly various models of random motion. The most common are Random Walk (Brownian Motion) [20], Random Waypoint (RWMM) [21], Random Direction, Boundless Simulation Area, Gauss-Markov, Probabilistic Random Walk, City Section, Group coordinated node movements, Cluster Mobility Model (the nodes are divided into clusters with constant connection inside; clusters are connected through several moving nodes), Shortest Path Map Based mobility model [22]. An interesting graph-based model (PANDORA) for data transfer tasks has been discussed in paper [23].

Separately, the reverse case of data collection by a single mobile data collector (Data Mule) from many offline nodes is considered, similar in the sense of having a dedicated communication node [24], [25].

The cases of sending data by transport couriers to remote from the Internet areas [26] and collecting and retransmitting data from sensors mounted on zebras (ZebraNet) [27] were considered.

The packet format for use in DTN ("bundles"), as well as a general specification of the buffering and transmission formats proposed in [28]. For the purposes of this article, these issues are not essential; it is enough to rely on the possibility of fragmentation on such bundles and the subsequent collection of data from them.

The utility concept used in Section III is based on the ideas [29] of the utility of individual message packets. This utility reflects the expected contribution of replication to the metric (expected delay, maximum delay, number of packets delivered). It can be used as a basis for a routing-decision algorithm.

III. THE MATHEMATICAL MODEL FOR PERFORMANCE EVALUATION

Due to the retransmissions to "faster" trains, the DTN system makes possible faster message delivery to the external network. We will try to evaluate the effect of accelerating the delivery of messages from retransmissions.

Consider a primitive scenario when two trains travel towards each other and a meeting occurs. One train goes from city A to city B, the other from city B to city A. There is an external network connection only in cities A and B for both trains. The time spent by trains in each other's radio access zone allows sending a message completely from one train to another. One or the other train generates a message at an equally probable random moment. Consider the gain in message delivery time due to DTN.

If trains travel the same time T on the way (their speeds are equal), then without DTN (with Direct delivery), the average message delivery time is $T/2$. With DTN, the average delivery time depends on the trains' meeting point and the minimum average delivery time is $(7/16)T$. This means, with DTN messages are delivered 0-12.5% faster.

If the difference between the speeds of two such trains increases and tends to infinity (the travel time of the "slow" train is still T), then the average Direct delivery time tends to $T/2$, and the minimum (depending on the meeting point of the trains) average message delivery time with DTN tends to $T/4$. This means, the maximum gain from DTN when delivering one message at different train speeds is 50%.

The diagram in Fig. 2 shows the distribution of the number of passenger train meetings per railroad section between mobile network coverage areas (in one offline path) for real train schedule in Tyumen region. The population density in Tyumen region (2.5-3.0 people/km^2) is within the boundaries of the average population density in Siberian and many other regions of Russia. Assuming a positive dependence of the number of passenger trains in the region and the coverage area of mobile network on population density, we can conclude that the prevailing scenario is a single-meeting, less two-time meeting of trains on a section between two coverage areas of base stations of a fixed radio network in a significant part of Russia.

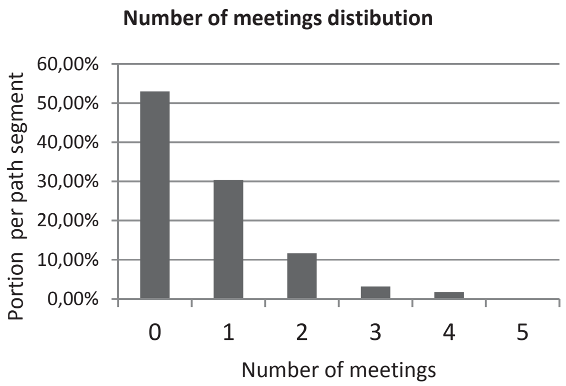

Fig. 2. The number of train meetings in one out of coverage area (offline path) in Tyumen-Surgut section.

Thus, the range of 12–50% can serve as a guideline for the boundaries of decreasing average message delivery time. From the analysis of train schedules, we can conclude that the prevailing scenario is almost the same speeds of meeting trains, so the upper limit of the expected efficiency range will be much lower in the low population areas.

Next, we consider the scenario, when trains generate/collect telemetry data.

A. Approach to the quality estimation of data transfer

Quantitative information utility function (utility density function) is considered to compare different ways of data delivery. Let define this function as the dependence of the value of a small portion of the information on the delay between the moments of its generation and final delivery.

For instance, the following functions act as utility density:

- threshold: data have the same value if the delay does not exceed a certain limit, after which the value is zero (data is no longer relevant).
- linear: the value of the data decreases linearly with time (corresponds to the summary delay).
- and others: exponential, power, maximum, etc.

For evaluation of the general delivery scheme, we assume that each object constantly collects information during its movement (many small portions of data are gradually formed). The overall utility is the sum of the utilities of data portions.

Objects transfer accumulated data (all or part of it) during the meeting, changing these partial delays and thus overall utility. We compare results to determine the best route scheme.

B. Delay information utility density function

Utility density function $i(t)$ is a value for an infinitely small portion of information with delivery delay t.

Any reasonable function of such type satisfies certain requirements. Usually earlier information receiving is preferable, so the function is non-increasing. It has the limit 0 (very old information is useless) or $-\infty$ (all information is critical and must be delivered ever). Also we suppose that utility density function is finite everywhere. Several utility density function examples are shown in Fig. 3.

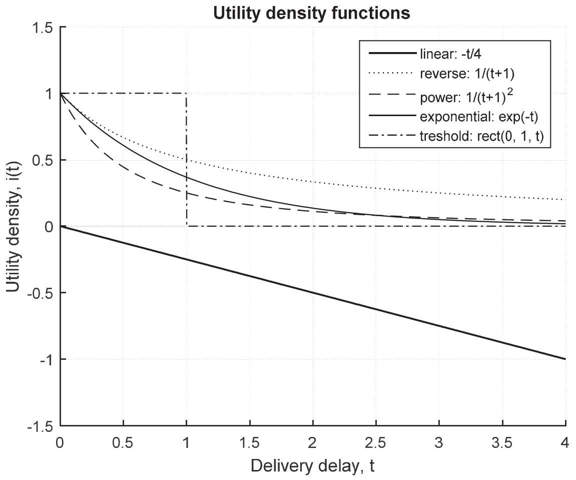

Fig. 3. Examples of different types of utility density functions: scaled linear, reverse, power, exponential, threshold

Utility of a continuous packet (a set of consecutive portions) with duration P, delay t, uploading rate dr (i.e. 1 second of uploading data equals to dr seconds of generation):

$$I_P(t) = \frac{dr}{dr+1} \int_0^{P(1+1/dr)} i(t+\theta)d\theta \qquad (1)$$

Equation (1) relies on the most recent data being transmitted first; when transferring the oldest data first, the utility will be less. For infinite uploading rate order is not important.

Unless otherwise indicated, we further suppose that the uploading rate is very high and the additional delay can be ignored.

Linear utility function is used further, but other functions can be used in specific cases.

C. *Average delay (linear utility density function)*

If during the time t_0 the value of the information decreases by 1, then the linear form of the utility function (2) should be used:

$$i(t) = -\frac{t}{t_0} \qquad (2)$$

If $t_0 = 1$, then the utility of the packet corresponds to the total delay during the continuous generation in units of time (up to the sign "-").

Also, t_0 can be interpreted as the characteristic period $[0; t_0]$ in which usually occurs one random event with uniformly distributed probability, i.e. it's probability density is $p = \frac{1}{t_0}$. In this case, the negative utility of the packet $-I_P(t)$ characterizes the average expected summary delay for these events that occurred in time range [0; P] after time t from its end. In this range there will be an average number of events $\frac{P}{t_0}$ with an average delivery delay $t + \frac{P}{2}$ for each one.

So, this measure could be used as the estimation for the purpose of QoS (Quality of Service).

The linear measure is used further for comparing delivery schemes as the most general characteristic (although similar calculations can be carried out for the rest utility functions, using the appropriate function for the information type used).

Measured usefulness of a packet (1) for function (2) is (3):

$$I_P(t) = -\frac{P}{t_0}\left(t + \frac{P}{2}\right) \qquad (3)$$

Visualization example for linear packet utility (with $t_0 = 4$) is shown in Fig. 4.

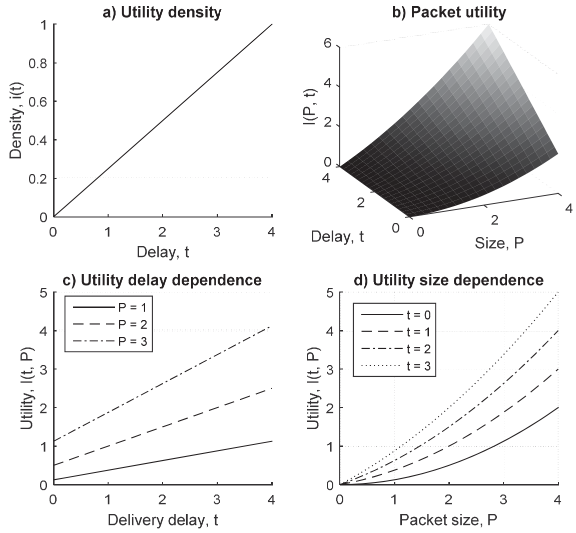

Fig. 4. Average delay (linear utility) estimation of the packet depending on the packet duration P and the transmission packet delay t (for $t_0 = 4$)

D. *Single meeting analysis (utility gain from a single event)*

A model of a single meeting allows preliminary estimating the possibilities of gain through the use of a mesh network. Subsequent meetings (that could increase the gain significantly) are not considered.

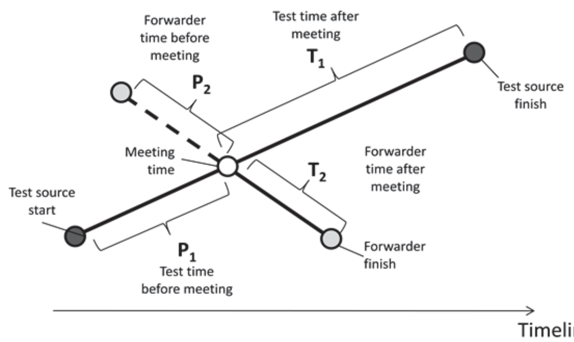

Fig. 5. Estimation model visualization: "test" object (data generating source) is moving from start to finish during the time $P_1 + T_1$, and in time P_1 after the start meets the "forwarder", who will finish earlier, in time $T_2 < T_1$

We compare the relative change in the aggregated delay (for both objects) during the transfer of the accumulated information to the one who comes first to the delivery area.

Test object (1) is defined as the object with longer delivery time (after the meeting): $T_1 > T_2$; objects have travel times P_1 and P_2 respectively before the meeting (see Fig. 5). Parameters $x = \frac{P_1}{P_1+T_1} \in [0..1]$ and $y = \frac{T_2}{T_1} \in [0..1]$ show parts of the paths where the meeting took place. Parameter $z = \frac{P_2+T_2}{P_1+T_1} \in [0..+\infty)$ sets the ratio of total route times.

For the case of average delay (linear utility density function) utility ratio factor (4):

$$\alpha_{agg} = \frac{\Delta I}{I} = \frac{2}{z+1}\big[\max\big(0, x(1+z(1-y)-x)\big) + \max\big(0, zy(1-z(1-y)-x)\big)\big] \tag{4}$$

For different ratios of route time before and after the meeting there will be different ranges of possible gain (due to the different data volumes accumulated before the meeting), one example is shown in Fig. 6.

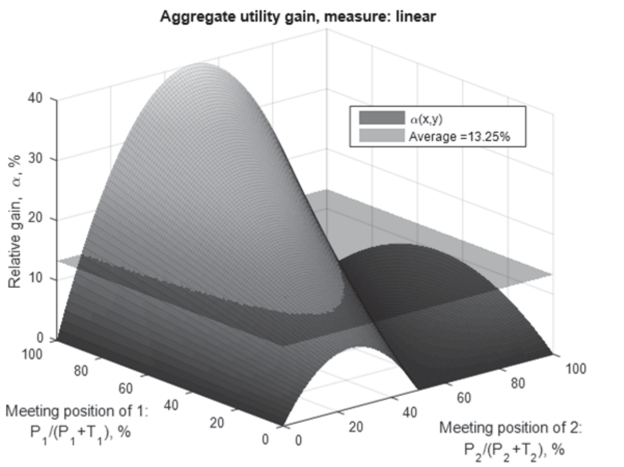

Fig. 6. Aggregate utility gain (linear utility density function): travel time of objects differs by 2 times $(T_2 + P_2) = 2(T_1 + P_1)$

A change in the average expected data delay is significant for the system of two objects even for one single meeting.

Of course this analysis does not take into account a lot of limitations and practical aspects, including: random/arbitrary connection and meeting times and durations (in practice, they need to be predicted that cause possible inaccuracies and errors), limited data transmission rate between objects (forwarding and uploading) and their dependence on signal/noise ratio, various overhead costs for transmission, data delivery fragmentation and others. This should be taken into account for practical use.

Next, consider a simulation of the delivery of telemetry data on the set of generated maps of railway lines and schedules. The effectiveness of the DTN network will depend on the selected protocol for exchanging data between trains at their meetings. With various protocols, the average efficiency will vary significantly depending on how much telemetry is generated by trains compared to the data rate.

Each moving object can have multiple encounters. Therefore, data transmission schemes in order to gain maximum utility can be complex.

To explore some of the decision-making approaches computer simulation was done. A simple model of moving objects between control points (some of them are BS) on a two-dimensional surface according to a strict schedule was considered. The movement was assumed to be uniform and straight. Based on the schedule, it is possible to accurately calculate the time and duration of communication with BS and other objects. The simulation is implemented on specially created software using PHP Version 5.3.3 (visualization and reports were put out in Plain text or HTML + CSS + javascript formats).

E. Simulation generator

The simulator generates a lot of different railway scenarios (the coordinates of the railway stations, railroad tracks, train schedule and coordinates of base stations). All random parameters for models are evenly distributed in their range (shown in brackets) using Mersenne Twister method [30].

For each simulation, the coordinates of nodes (2-20 items) and their connections in the model space of 560x360 km are generated. These nodes are railway stations. Each node must be connected to one other node at least, as a maximum - to each other node. No intersection allowed. Each node can be a BS of outer network with a certain probability for the scenario (10-90%), but at least one BS must exist.

The RWMM schedule is created as follows: each object (of 2-60 pieces that represent trains) starts moving from one of the BSs in its initial time (0-100 minutes) with random speed (20-120 km/h). It moves to one of the nodes with which the current node is connected. After arrival, the object stays there for some time (0-20 minutes). Then it repeats the move procedure again.

Example of model state for one of the scenarios at a specific point in time is shown in Fig. 7.

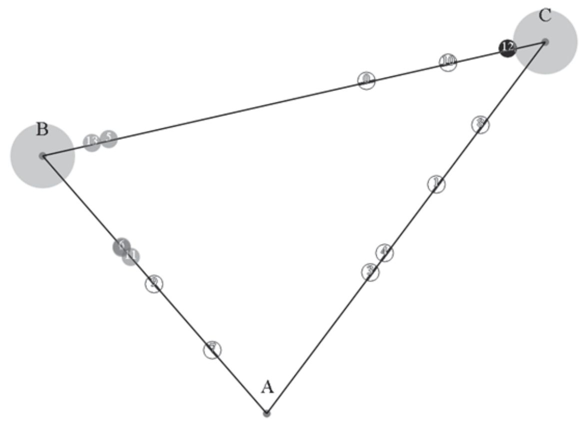

Fig. 7. An example of visualization of nodes location (denoted by letters), their connections, BS at some nodes (coverage is marked with big light gray circles around the nodes), objects (indicated by numbers) and TS coverage (small circles: normally without filling, gray while meeting TS, dark gray if connected to BS) at some time.

After each node, the object decides to finish movement with a certain model probability (10-90%), or if the number of steps has already exceeded the threshold (10-50 steps). If this decision is made, then the movement is terminated at the nearest arrival at a BS. This ensures the finiteness of the object movement and its start and finish online (at railway stations with BS).

F. Simulation analyzer

Scheduled movements were calculated using 1 minute time steps. At each of them, the objects moved, generated a portion of the data, uploaded the data (in the BS coverage area) or transferred them to another object (in the TS coverage area) with specified data rates using several decision making algorithms. During uploading, the total delay was determined - the difference between the generation time and the current one.

5000 different generated scenarios were stored and analyzed (see schedule example in Fig. 8).

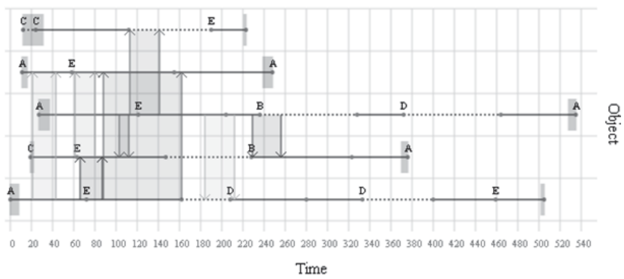

Fig. 8. An example of a visualization of the schedule for moving objects (objects are arranged vertically, the horizontal axis corresponds to the observation time). The horizontal lines correspond to the movement of the object between the control nodes. Vertical arrows visualize meetings and possible transmissions between objects.

G. Decision-making/routing protocols

1) Earliest Direct Delivery (EDD)

All the objects at a meeting (2 or more) compare their next expected connection times (next time when they will arrive at BS) with each other. Data is transferred to the object with the best result. This method is very simple and practical because it uses the minimum number of predicted parameters.

2) Earliest Direct Delivery with Local Queuing (EDDLQ)

As both uploading time and uploading data rate are limited, some of the accumulated data may not be uploaded during the nearest communication session with BS.

To take this into account, the nearest available uploading times of the objects are compared based on the accumulated data amount at the current moment (its own and received), the expected volume of later generated data and the duration of the future connections with BSs.

3) Earliest Delivery (ED)

EDD and EDDLQ approaches find the local maximum utility gain. But if all subsequent meetings with other objects are also known (from the schedule), following forwarding could be taken into account. Decision criterion is this "shortest possible time" (found with Dijkstra algorithm). See comparison of EDD and ED example in Fig. 9.

But if the subsequent planned meetings will not take place due to schedule deviations, or objects cannot transfer the entire accumulated amount of data (including from other objects that did not take these plans into account), then ED may not reach global utility maximum and even be less successful than EDD.

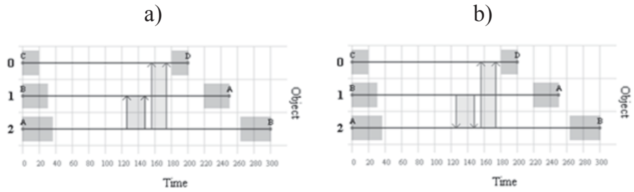

Fig. 9. Transmission direction examples for EDD (a) and ED (b) approaches.

4) Epidemic

Objects while meeting transmit to each other a copy of their data. Uploading is made only for data that was not uploaded earlier. This approach eliminates the wrong choice of transmission direction, but requires a large amount of memory, and only part of the transfer between objects may be useful.

H. Decision-making/routing protocols

In the simulations made, all approaches gave utility gain (total delay reduction) in most cases. Average utility gain (delivery delay) and its deviation are calculated in % of the utility of the scenario without retransmissions (no DTN between moving objects).

TABLE I. MULTIPLE RETRANSMISSIONS SIMULATION RESULT FOR LOW UPLOADING RATE

Decision-making approach	Simulation average results upload/generate data rates ratio <50		
	Average gain	*Standard deviation*	*Zero gain cases*
EDD	16,40%	16,09%	16,49%
EDDLQ	22,11%	16,90%	6,89%
ED	10,18%	11,72%	26,39%
Epidemic	7,07%	9,91%	40,38%

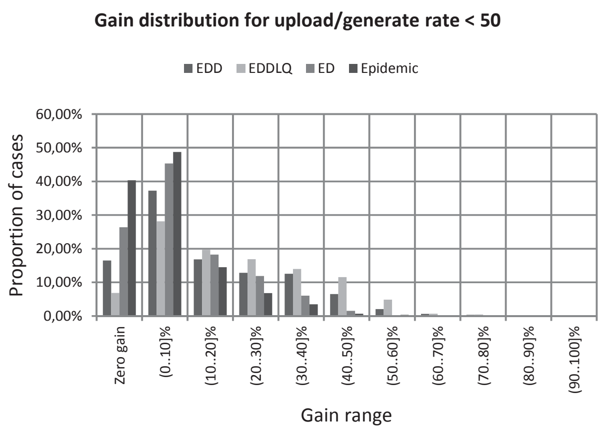

Fig. 10. Gain distributions in simulation for different decision-making algorithms for low uploading rate.

TABLE II. MULTIPLE RETRANSMISSIONS SIMULATION RESULT FOR HIGH UPLOADING RATE

Decision-making approach	Simulation average results upload/generate data rates ratio ≥50		
	Average gain	*Standard deviation*	*Zero gain cases*
EDD	37,59%	20,23%	3,57%
EDDLQ	37,64%	20,25%	3,57%
ED	38,93%	21,95%	3,57%
Epidemic	24,88%	16,59%	6,84%

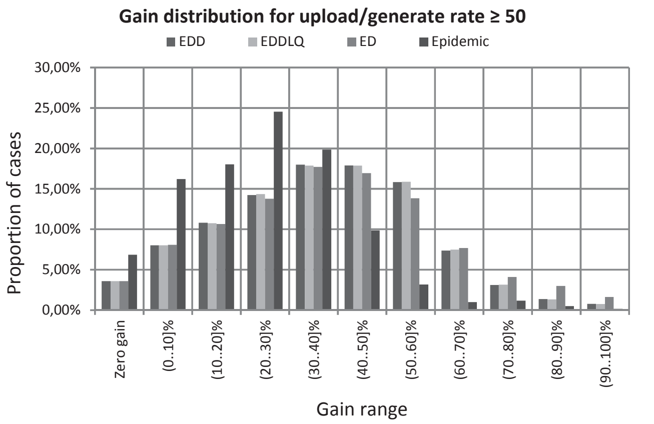

Fig. 11. Gain distributions in simulation for different decision-making algorithms for high uploading rate

As can be seen from the presented simulation results (see Fig. 10, Fig. 11 and Tables I and II) of various DTN protocols, for telemetry a gain of 20% or more can be achieved. An important feature of the above simulations is the scenario, in which trains have much more than one meeting for each area without coverage. Also, the times of states of a radio link of train with base stations and with other trains are values of the same order.

Further, when analyzing real or near real maps of railway lines, schedules and mobile network coverage, the importance of this ratios for the effectiveness of DTN will be shown.

IV. DTN SIMULATION ON THE REAL RAILWAY SECTIONS

Analysis of real railway scenarios was performed with specially created software in C++, PHP Version 5.3.3 (visualization and reports were put out in Plain text or HTML + CSS + javascript formats).

Consider the section of the railway line of the Ural Federal District between the cities of Tyumen and Surgut (see Fig. 12). These cities as of today are daily connected by 11 pairs of trains (22 trains) with a fixed schedule. The line is covered by a mobile network of three major carriers (see coverage example in Fig. 13).

Inside the total coverage of the mobile network, according to the modeling conditions, we assume a data rate ten times the date rate between trains, and the radio link maximum range between two trains is 5 km.

For the simulation the actual schedule was used, including arrival and departure times. Trains in the simulation move with an individual constant speed between the stations according to the schedule.

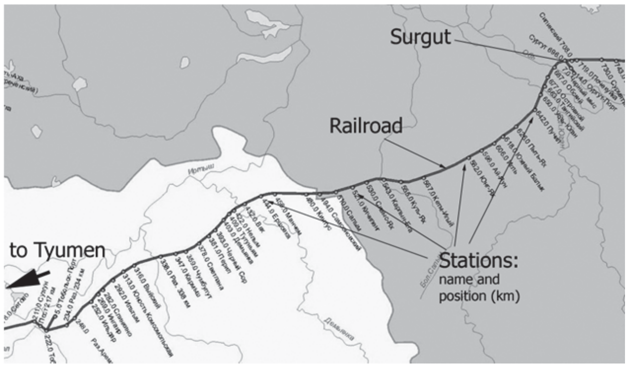

Fig. 12. Part of Tyumen-Surgut section official railroad scheme used for simulation.

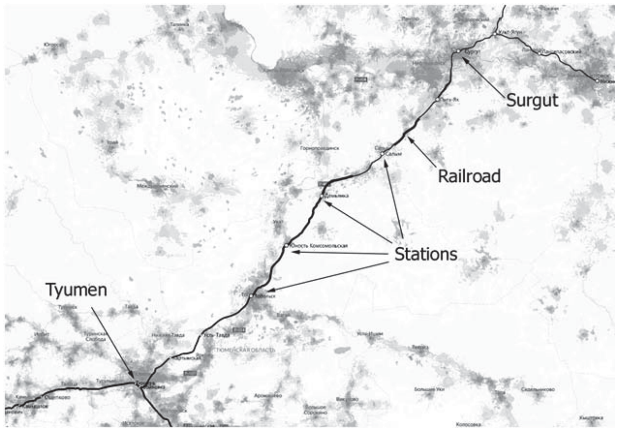

Fig. 13. Tyumen-Surgut railroad coverage map (for one mobile network carrier, MTS). Color intensity corresponds to quality of coverage.

Generating messages on trains at arbitrary points in time, we obtain the following efficiency charts for message delivery time (see Fig. 14). The average single message delivery time with DTN and Direct delivery and the maximum message delivery time in the corresponding systems are also shown. DTN is using the epidemic distribution protocol [7] estimated with Dijkstra's algorithm.

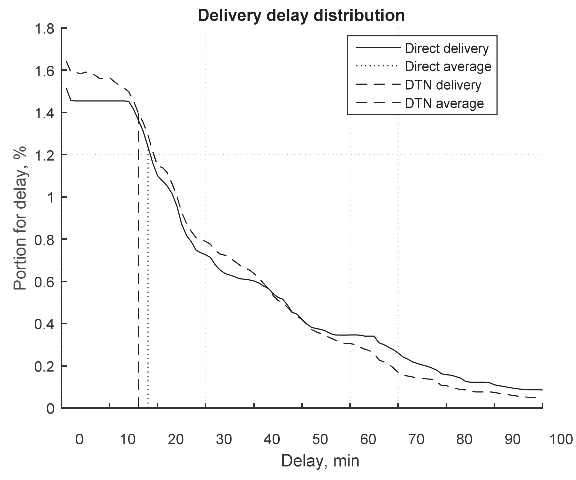

Fig. 14. Average delivery delay distribution with and without DTN (except immediate delivery with zero delay).

Most of the subscribers' communications in messengers, e-mail, etc. are not a single message to another subscriber, but dialogs with replies and forwarding. A subscriber generates a response or logical continuation to a previously received message. Suppose the response generation time is distributed normally (with $\sigma=\mu$). Then the graphs of the average and maximum message delivery times will change depending on the parameters of the response time distribution (see Fig. 15, Fig. 16, Fig. 17, Fig. 18).

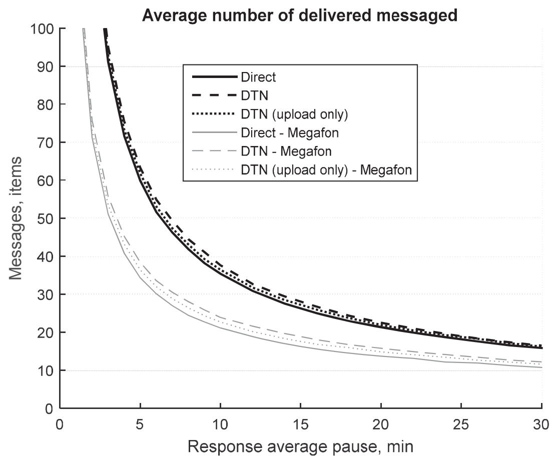

Fig. 15. Average number of generated messages with response-to-request message scheme depending on average response time (normally distributed random value) for total mobile carriers available, and for only one of them ("Megafon")

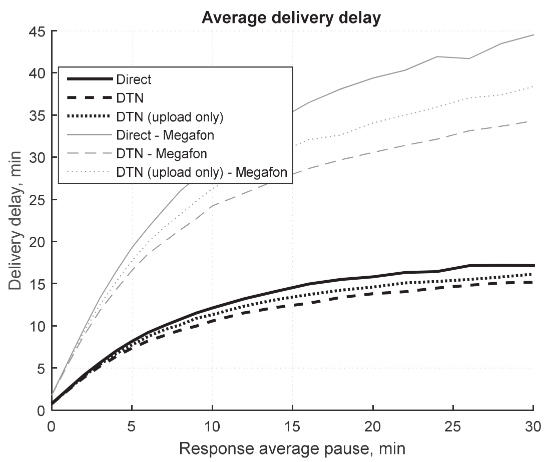

Fig. 16. Average delivery time with response-to-request message scheme depending on average response time (normally distributed random value) for all mobile operators available, and only one of them ("Megafon").

Note that in the above simulation, the ratio of train time being in a coverage of mobile network is from 21% (for one mobile carrier "Megafon") to 38% (if total coverage of all three available networks are used by trains), and the time when trains are linked with each other is 4%. These values differ by an order of magnitude. The average message delivery time shows a decrease up to 15-23% in this cases relatively.

The message maximum delivery time in this simulation is reduced in DTN by 15-25%. Significant reducing of the maximum message delivery time can be crucial for telemetry of certain goods that require the maintenance of certain storage/transport conditions.

Fig. 17. Maximum delivery time with response-to-request message scheme depending on average response time (normally distributed random value) for all mobile operators available.

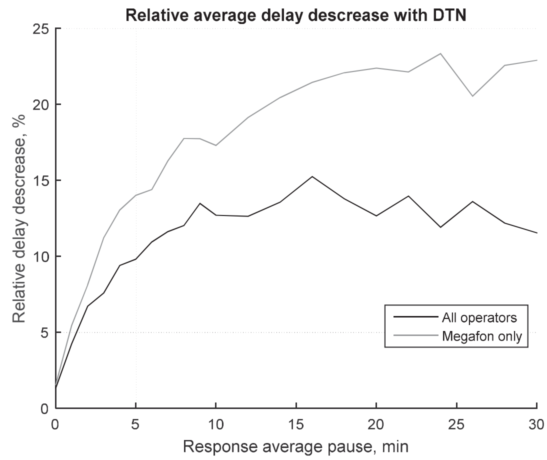

Fig. 18. Relative average delay decrease with response-to-request message scheme depending on average response time (normally distributed random value) for total mobile carriers coverage, and only one of them ("Megafon").

Note that the simulated section of the railway line has a well-established mobile network of about 27% coverage of the distance and about 38% of the time the trains are in the coverage area (trains spend at stops that are covered, more of the time on the way). Previously, this network was smaller, and coverage began growing from large cities, then from large intermediate stations/towns. Fig. 19 shows the effectiveness of the DTN network with the expected growth of coverage (in % of the current one).

As can be seen in Fig. 20, the number of messages "supported" by the DTN system does not increase with the coverage growth - this number of messages is determined primarily by the trains' schedule, the presence and number of trains' meetings, the position of these meetings with respect to the location of the base stations. The number of messages with

Direct delivery alone increases with coverage growth. With the ratio of the time spent by trains within the coverage of base stations of the same order to the time of radio link between trains during meetings, the result of the DTN simulation on a real schedule is in good agreement with the DTN results on the generated timetables and railway maps from section III.

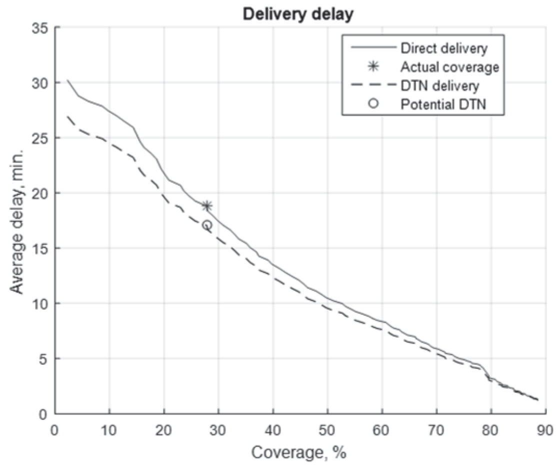

Fig. 19. Average delivery delay depending on coverage area (scaled from existing now)

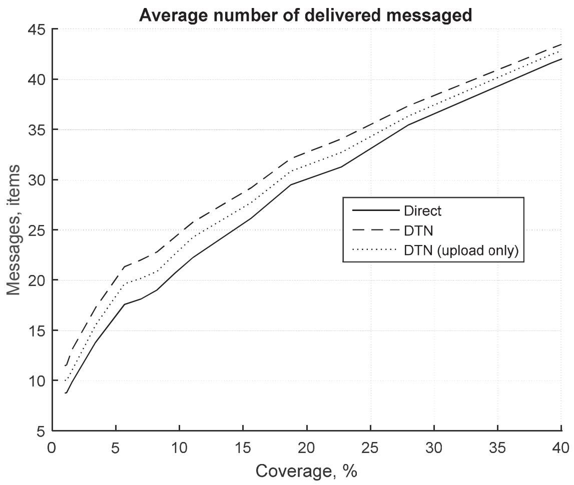

Fig. 20. Average number of delivered messages depending on coverage.

The efficiency of DTN in reducing the average message delivery time and in increasing the total number of messages fades with the growth of the mobile network coverage (to 0 with full coverage). This indicates the highest efficiency of DTN when starting a new network on a section of a railway line and when migrating to it from the old network.

V. DTN EFFICIENCY DURING MIGRATION TO A NEW NETWORK

Due to the obsolescence of the GSM-R standard and the refusal of its maintenance in the near future (2030), among other alternatives, the 5G mobile standard and/or Satcom can be used as a railway communication standard. Consider two sections of the railway line: the previously considered section of Tyumen-Surgut and the section with a high concentration of trains Moscow - St. Petersburg. From the point of view of the feasibility of satellite-connected trains implementation, the first section belongs to the category of "Greenfield" [31] - therefore, the use of satellite terminals there is more preferable. Since Moscow - St. Petersburg line with a coverage of up to 90% or more of the way falls into the category of "Brownfield", it is more advisable to develop a ground-based railway communication network and a higher-speed communication network based on existing infrastructure.

Consider the introduction of satellite terminals for trains on the section Tyumen - Surgut. As can be seen in Fig. 21, the increase in the number of trains connected to an external network via a satellite channel has practically no effect on the number of messages "supported" by the DTN network. In this sense, the increase in the number of trains with a satellite terminal is similar to the increase in the coverage of a ground-based network (see Fig. 22).

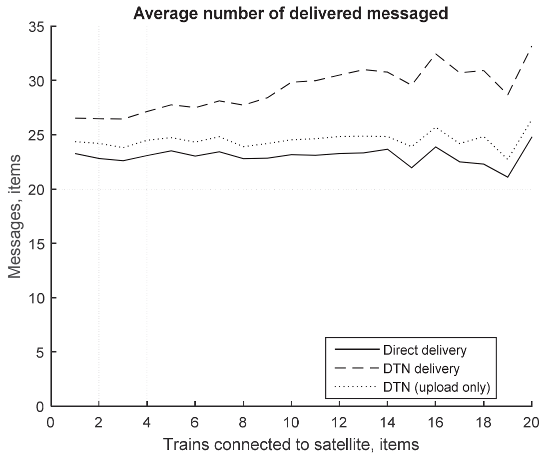

Fig. 21. Average number of delivered messages per train (only for the trains that are not connected to a satellite) depending on number of trains connected to satellites

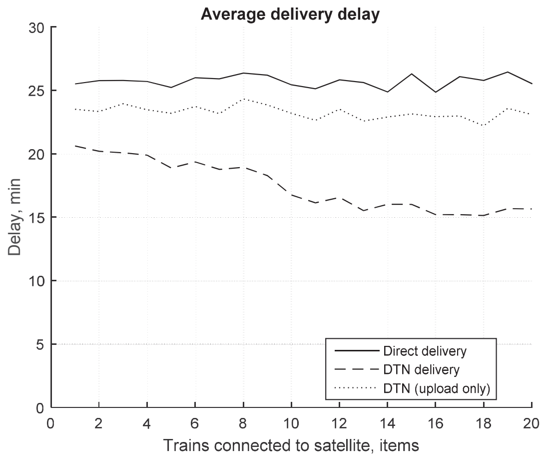

Fig. 22. Average delivery delay (only for the trains that are not connected to a satellite) depending on number of trains connected to satellites.

Now consider a section of the Moscow-St. Petersburg railway. For this section, we used a close to the real model of passenger and commuter trains movement. This section also has some high-speed trains. A large number of trains and a large difference in their speed indicate the great potential of

DTN for this railway section. Since this section of the railway line is one of the busiest, it is very likely that a network with a higher data rate (5G) will be implemented on it. A 5G network with coverage of about 5-10% of an existing network can increase traffic relative to an existing network by 2 or more times.

The next Fig. 23. shows the dependences of the average message delivery time with increasing coverage of the new network.

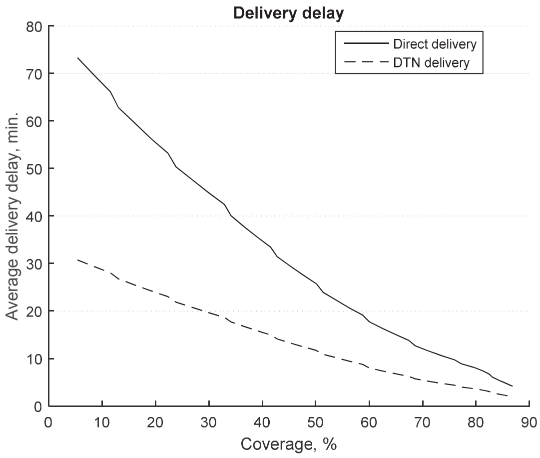

Fig. 23. Average delivery delay depending on suggested 5G coverage of the Moscow-St. Petersburg Railway.

VI. Conclusions

The total number of messages "supported" by the DTN network primarily depends on the train schedule: the number of meetings and the ratio of train speeds at meetings. On railway sections that have long intervals out of mobile network coverage, DTN makes a significant contribution (15-25%) to reducing the maximum delay of message delivery. This may be crucial for some types of telemetry. With an increase in the coverage of the mobile network, as well as with an increase in the number of trains with a permanent satellite connection, the number of messages "supported" by DTN does not drastically change. Thus, DTN is most effective at the initial stages of network implementation and migration to it from a previously existing system. DTN efficiency on established coverage in rural areas (about 30-40% coverage) is 10-23% reduction in message delivery time. DTN efficiency on newly deployed networks can be significantly more than 20% reduction in message delivery time depending on the schedule of trains.

References

[1] D. Namiot, "On mobile mesh networks", *International Journal of Open Information Technologies*, vol. 3.4, 2015, pp. 38-41.

[2] I.F. Akyildiz and X. Wang, "A survey on wireless mesh networks", *IEEE Communications magazine*, vol. 43.9, 2005, pp. S23-S30.

[3] D. Namiot and M. Sneps-Sneppe, "Proximity as a service", *Proceedings of the 2nd Baltic Congress on Future Internet Communications,* IEEE, 2012, DOI 10.1109/BCFIC.2012.6217947.

[4] R. Ramanathan and J. Redi, "A brief overview of ad hoc networks: challenges and directions", *IEEE communications Magazine*, vol. 40.5, 2002, pp. 20-22.

[5] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets", *Computer Communication Review*, vol. 33, 2003, pp. 27-34, DOI 10.1145/863955.863960.

[6] S. Jain, K. Fall, and R. Patra, "Routing in a Delay Tolerant Networking", *Computer Communication Review – CCR*, vol. 34, 2004, pp. 145-158, DOI 10.1145/1030194.1015484.

[7] A. Vahdat and D. Becker, "Epidemic Routing for Partially-Connected Ad Hoc Networks", *Technical Report*, 2000.

[8] A. Lindgren, A. Doria, and O. Scheln, "Probabilistic Routing in Intermittently Connected Networks", *Mobile Computing and Communications Review*, vol. 7, 2003, pp. 19-20.

[9] E. Jones, L. Li, and P. Ward, "Practical Routing in Delay-Tolerant Networks", *SIGCOMM05-DTN Workshop 05,* 2005, pp. 237-243, DOI 10.1145/1080139.1080141.

[10] S. Nelson, M. Bakht, and R. Kravets, "Encounter-based routing DTNs", *Proceedings - IEEE INFOCOM*, 2009, pp. 846-854, DOI 10.1109/INFCOM.2009.5061994.

[11] J. Burgess, B. Gallagher, D. Jensen, and B.N. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks", *IEEE INFOCOM*, April 2006.

[12] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation forwarding", *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '08)*, Association for Computing Machinery, 2008, pp. 251–260.

[13] X. Chen and A. Murphy, "Enabling disconnected transitive communication in mobile ad hoc networks", 2001.

[14] H. Kang and D. Kim, "Vector Routing for Delay Tolerant Networks", Vehicular Technology Conference, 2008, pp. 1-5, DOI 10.1109/VETECF.2008.21.

[15] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Spray and wait: An efficient routing scheme for intermittently connected mobile networks", *Proceedings of ACM SIGCOMM workshop on Delay-tolerant networking*, 2002, DOI 10.1145/1080139.1080143.

[16] T. Spyropoulos, K. Psounis, C. Raghavendra, "Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility", *Proceedings of Fifth Annual IEEE International Conference on Pervasive Computing and Communications - Workshops (PerCom Workshops 2007)*, 2007, pp. 79-85, DOI 10.1109/PERCOMW.2007.108.

[17] S. Burleigh, "Contact Graph Routing", *Internet-Draft*, July 2010, available: https://tools.ietf.org/html/draft-burleigh-dtnrg-cgr-00.

[18] G. Araniti, N. Bezirgiannidis, E. Birrane, I. Bisio, S. Burleigh, C. Caini, M. Feldmann, M. Marchese, J. Segui, and K. Suzuki, "Contact Graph Routing in DTN Space Networks: Overview, Enhancements and Performance", *Communications Magazine*, 2015, pp. 38-46, IEEE 53, DOI 10.1109/MCOM.2015.7060480.

[19] R. Handorean, C. Gill, G. Roman, "Accommodating transient connectivity in ad hoc and mobile settings", 2nd International Conference on Pervasive Computing, vol. 3001, 2004, pp. 305-322.

[20] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research", *Wireless communications and mobile computing*, vol. 2, issue 5, 2002, pp. 483-502, DOI 10.1002/wcm.72.

[21] A. Sharma, Dr. Gurpreet, and Er. Jaswinder Singh, "Mobility models for MANET: Mathematical perspective", *International Journal of Advanced Research in Engineering and Applied Sciences*, vol. 2.5, 2013, pp. 59-68.

[22] P. Dwivedi and R. S. Pippal, "Performance Evaluation of Spray and Wait DTN Routing Protocol under Different Mobility Models", *Journal of The Gujarat Research Society*, vol. 21-16, 2019, available: http://www.gujaratresearchsociety.in/index.php/JGRS/article/view/1820.

[23] B. Cho and I. Gupta, "New algorithms for planning bulk transfer via internet and shipping networks", *IEEE 30th International Conference on Distributed Computing Systems*, 2010.

[24] D. Jea, A. Somasundara, and M. Srivastava, "Multiple controlled mobile elements (data mules) for data collection in sensor networks", *Proceedings of the First IEEE international conference on Distributed Computing in Sensor Systems (DCOSS'05)*, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 244–257, DOI 10.1007/11502593_20.

[25] R.C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: Modeling a three-tier architecture for sparse sensor networks", *Ad Hoc Networks*, vol. 1, 2003, pp. 30-41, DOI 10.1109/SNPA.2003.1203354.

[26] J. Lebrun, C. Chuah, and D. Ghosal, "Knowledge Based Opportunistic Forwarding in Vehicular Wireless Ad Hoc Networks", *Proceedings IEEE Vehicular Technical Conference*, vol. 4, 2005, DOI 10.1109/VETECS.2005.1543743.

[27] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with ZebraNet", *Proceedings of the 10th international conference on Architectural support for programming languages and operating systems (ASPLOS X)*, 2002, pp. 96–107, Association for Computing Machinery, New York, USA, DOI 10.1145/605397.605408.

[28] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, "Delay-Tolerant Networking Architecture", *RFC 4838*, April 2007, available: https://tools.ietf.org/html/rfc4838.

[29] A. Balasubramanian, B.N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem", *Proceedings of ACM SIGCOMM*, August 2007.

[30] M. Matsumoto, T. Nishimura, M. Hagita and M. Saito, "Cryptographic Mersenne Twister and Fubuki Stream/block Cipher", *IACR Cryptology ePrint Archive*, 2005.

[31] *Study on feasibility of SATCOM for railway communication*, Final Report, ERA 2016 01 SC, VERSION 1/1, European Union Agency for Railways.