Use of samples for the choice of routes in networks of data transmission

A Purtov

Sobolev Institute of Mathematics, Siberian Branch of Russian Academy of Science, Omsk, Russia

Abstract. The article is connected with a problem of the organization of purposeful movement, actual for various spheres of activity. Described is one of ways of the decision of this problem on an example of imitation of purposeful movement of packages in a network of data transmission. For routing samples are used. Sample ST is of the parameters. These parameters are coordinates of a point in n-dimensional space. The state Si of each communication channel too has n the current parameters. The packet is transmitted over a channel for which the Euclidean distance from Si to ST is minimal. The simulation model is implemented on GPSSW. The results of the experiments showed the possibility and feasibility of using samples for routing. The samples can be used to select the routes of purposeful movement on the oriented graphs in systems for various purposes: technical, economic, social.

1. Introduction

Developed many formal and informal methods of decision-making in systems for various purposes, technical, economic, social, etc. Training and decision-making on samples often used in nature. The author of the article develops ways of using samples for the organization of purposeful movement on oriented graphs. At the current stage, data transmission networks (DTN) are used as objects for experiments.

In one of his last monographs [1] Zagoruiko N. G. gives various methods of data analysis and classification of objects. Some methods are based on the use of pillars (standards, samples). In this case, the database contains a set of research objects. Normalized parameters of objects are considered to be their coordinates in n-dimensional space. After processing the data by statistical methods, typical objects – samples are determined. Classification consists in attributing objects to those samples, the distance to which is minimal.

In another way, samples are used in the method of decision-making, which is developed by the author of the article. The current situation in the system is represented by a set of parameters. The values of some parameters are random and change during the operation of the system. Therefore, many situations are not defined in advance. Samples are specified with the same set of parameters that are used to describe situations. The values of the sample parameters are selected heuristically and using simulation. Depending on which of the samples is closer to the current situation in the system, a decision is made.
The first experiments of the author of the article on the use of samples were carried out in the framework of a comparative analysis of algorithms for controlling the flow of cars at intersections [2]. The conceptual model consisted of four intersecting streams, 12 directions, four traffic lights. Four traffic light control algorithms were compared:

- traffic lights change state at a given time interval;
- traffic lights change state in the absence of cars in the open direction;
- traffic permit receives the flow with the largest queue of cars;
- samples are used to make decisions about switching traffic lights.

The criterion for comparing the algorithms was the average waiting time of cars in the queue before the intersection.

For comparison, we used a simulation model, programmed on GPSSW (General Purpose Simulating System World) [3]. When using samples, the current state of a car intersection is determined by a set of parameters $S = \{Q_a, Q_c, Q_e, Q_g\}$, where $Q_a, Q_c, Q_e, Q_g$ - the number of cars before the intersection in each flow (a,c,e,g). The parameters $Q_a, Q_c, Q_e, Q_g$ specify a point in four-dimensional space. Each stream before modeling was given a sample $ST = \{Q_a, Q_c, Q_e, Q_g\}$, which also represents a point in four-dimensional space. The Euclidean distances of the ST to each of the four points S were calculated. The flow for which the distance from ST to S is minimal was allowed to move.

The results of simulation experiments showed that algorithms based on samples have the right to exist. They are often no worse than the known algorithms for controlling vehicles at an intersection in terms of time delays. The use of samples has great potential for solving the problems of distribution of time resources between directions.

Checking the possibility of using samples for routing in the DTN is convenient to carry out on the simulation model. A set of analytical and simulation methods for the analysis of large networks is proposed in [4], [5].

2. Problem statement, theoretical and practical background

It is necessary to evaluate the possibility of using samples when choosing routes on oriented graphs, in particular, in DTN. For this it is necessary to compare the management of packet on the basis of samples with known methods of routing. It is necessary to determine a set of sample parameters, select a DTN section, build a simulation model of its operation, conduct experiments. The DTN fragment should not be very large, but it should reflect the main routing problems in branched networks.

The first computer networks used the X. 25 Protocol for routing. Procedures X. 25 allow for the transfer of the data how through the permanent virtual connection, and in the mode of adaptive routing datagrams. The expansion of DTN and the increase in channel speed led to the replacement of X. 25 with more efficient protocols, such as RIP (information routing Protocol), OSPF (Open Shortest Path First). When the networks were small, the decision was made on the basis of the minimum time to achieve the goal. With the growth of networks, it has become easier to use the shortest path in terms of the number of intermediate routers. Therefore, despite the great capabilities of the RIP and OSPF protocols, routers usually receive information about the shortest paths in terms of the number of nodes. Recently, the MPLS (Multiprotocol label switching) Protocol has gained popularity [6]. The Protocol is located between the network and channel layers. Usually on top of IP. Under it can be Frame Relay, ATM, PPP, Ethernet. MPLS can use fixed routes. But the main advantage is the support of traffic engineering technology. MPLS is used by both large operators such as Rostelecom and small operators.

This article describes how to use the $ST = \{Q, V, N, T\}$ pattern for routing. It is assumed that the parameters of the sample are the coordinates of a point in four-dimensional space.

Parameter description:

- $Q$ - the queue length to communication channel;
- $V$ - outgoing channel speed;
- $N$ - the minimum distance from the node making the decision on the direction of the package to the destination. Metric - the number of segments between nodes;
- $T$ - the transfer time from the start node to the end node.

The selection of a set of sampling parameters took into account the theoretical and practical experience of routing in DTN. The state of each $i$-th channel is defined as $S_i = \{Q_i, V_i, N_i, T_i\}$.

Parameter description:
- $Q_i$ - the current queue length to a communication channel;
- $V_i$ - outgoing channel speed (Mbps);
- $N_i$ - the minimum distance from the node making the decision on the direction of the packet to the destination;
- $T_i$ - the current transfer time from the start node to the end node.

When deciding on the direction of transmission of the packet, the outgoing communication channel from the node for which the distance from $S_i$ to $ST$ is minimal is selected.

To analyze the possibility of using samples, a simulation model of the DTN fragment is constructed.

3. Simulation model description

The simulation model is developed on GPSSW. Figure 1 shows the structure of the DTN fragment. Circles with numbers indicate DTN nodes (routers, switches). Lines with numbers and arrows indicate communication channels.

![Figure 1. The structure of the fragment DTN](image)

The following main parameters were set at the input of the simulation model:
- the average time between the arrival of packets in the DTN ($T_1$). The time interval between packets arrival is a random variable distributed exponentially with the mean $T_1$;
- the average time between the receipt of packets for each pair of nodes DTN;
- packet length-576 bytes;
- speed of communication channels (table 1);
- minimum distances from the current node to the end node $N_i$ (table 2);
- simulated time interval of work the DTN;
- set and values of sampling parameters $ST = \{Q, V, N, T\}$. Various combinations of sample parameters were used in simulation experiments.

The criterion for comparing the samples was the average time of transmission of packets by DTN. The comparison was made in the ordinal scale ($>$, $=$, $<$). Therefore, in this case not have much value specific values of the speeds of communication channels.
Table 1. Communication channel speeds.

| Channel number | Node transmitter | Node receiver | Channel speed (Mbit/s) |
|----------------|------------------|---------------|-----------------------|
| 1              | 1                | 2             | 1                     |
| 2              | 1                | 3             | 2                     |
| 3              | 1                | 5             | 1                     |
| 4              | 2                | 1             | 1                     |
| 5              | 2                | 3             | 2                     |
| 6              | 2                | 4             | 1                     |
| 7              | 3                | 1             | 2                     |
| 8              | 3                | 2             | 2                     |
| 9              | 3                | 4             | 2                     |
| 10             | 3                | 5             | 2                     |
| 11             | 4                | 2             | 1                     |
| 12             | 4                | 3             | 2                     |
| 13             | 4                | 6             | 1                     |
| 14             | 5                | 1             | 1                     |
| 15             | 5                | 3             | 2                     |
| 16             | 5                | 6             | 1                     |
| 17             | 6                | 4             | 1                     |
| 18             | 6                | 5             | 1                     |

Table 2 shows the numbers of communication channels for fixed shortest routes. The first line of the table shows the numbers of node of transmitters. The first column of the table shows the receivers node numbers. At the intersection of a row and a column, the number of the channel from which the shortest route begins is indicated. The length of the route is indicated in brackets.

Table 2. Shortest routes.

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 4, (1) | 7, (1) | 12, (2) | 14, (1) | 18, (2) |
| 2 | 1, (1) | 8, (1) | 11, (1) | 15, (2) | 17, (2) |
| 3 | 2, (1) | 5, (1) | 12, (1) | 15, (1) | 18, (2) |
| 4 | 1, (2) | 6, (1) | 9, (1) | 16, (2) | 17, (1) |
| 5 | 3, (1) | 5, (2) | 10, (1) | 12, (2) | 18, (1) |
| 6 | 3, (2) | 6, (2) | 10, (2) | 13, (1) | 16, (1) |

The following sequence of actions is used to simulate the transmission of a packet in the DTN.
1. The generation of the packet.
2. Determination the number of the starting node.
3. Determination of the number of the end node.
4. The choice of communication channel.
5. The packet is queued to the communication channel.
6. Seize of the communication channel for the packet, exit the queue, and send the packet to the destination node.
7. If the packet has not reached the destination node, go to step 4.
8. Formation of statistics.
9. The terminate of the packet.

Packets arrive at the model through time \( t = T_1 [\ln(1 - R_N)] \), where \( R_N \) is a uniformly distributed random variable from 0 to 1. The numbers of the start and end nodes are discrete random variables distributed evenly between 0 and 6. Is played first the number of the starting node \( i \). Then played the number of the target node \( j \). If \( i = j \), the number of the target node is played back. For transmission selected the communication channel for which the distance from \( S_i \) to \( S_T \) is minimum. The packet transfer time through the communication channel is defined as \( T_p = T_q + \frac{L}{V} \), where \( T_q \) is the waiting time of packet in the queue before the communication channel, \( L \) is the length of the packet, \( V \) - velocity of the communication channel.

The results of simulation experiments are given in table 3.

| Number of experiment | ST          | \( T_1 \) (\( \mu s \)) | \( T_2 \) (ms) |
|----------------------|-------------|--------------------------|---------------|
| 1                    | \{-, -, 0, -\} | 300                      | 4.64          |
| 2                    | \{0, -, 0, -\} | 300                      | 3.46          |
| 3                    | \{-, -, 0, 0\} | 300                      | 3.16          |
| 4                    | \{0, 4.0, 0, -\} | 300                      | 4.56          |
| 5                    | \{0, 3.0, 0, -\} | 300                      | 3.46          |
| 6                    | \{0, 3.1, 0, -\} | 300                      | 3.04          |
| 7                    | \{0, 3.1, 0, -\} | 225                      | 4.28          |
| 8                    | \{0, 3.5, 0, 0\} | 300                      | 2.91          |
| 9                    | \{0, 3.5, 0, 0\} | 225                      | 3.60          |

Table 3. The basic data of the experiments.

Table column assignment:
- first column- the experiment number;
- second column- the values parameters \( S_T \);
- third column- the average time between packets arrival between each pair of nodes;
- fourth column- the average time of transmission of packets in DTN.

The values of sample parameters and packet arrival rates were chosen empirically with subsequent testing on a simulation model. The criterion for the efficiency of the set and the values of the sample parameters was the average time of packet transmission by DTN. The packet intensities were supposed to provide the maximum but acceptable load of communication channels.

A dash in the sample means that the parameter was not used when running the simulation model.

The sample first line is defined by a single parameter that minimizes the number of transmission sections. In fact, it is the transmission on the shortest routes.

In the second experiment, a parameter is added to the sample that takes into account the current packet queue length into the outgoing channel from the node. This algorithm can be classified as adaptive. This option is directly available to the router and reduces network transfer time.

The sample lines 3, 8, 9, use a parameter that takes into account the time of transmission of packets to the destination. This criterion can reduce the time \( T_2 \), but its application requires additional time and resources for the transmission of SPD service messages. Therefore, this parameter is currently rarely used.

Three parameters are used in the sample lines 4, 5, 6, 7. Lines 4, 5, 6 show that the specific value of the sample parameter can greatly affect the transmission time of packets.
Lines 6, 7, 8, 9, show that taking into account the current time of transmission of packets to the destination can reduce the time $T_2$.

Figure 2 shows the dependence of the average transmission time of the packet by DTN ($T_2$) on the value of the parameter $V$ of the sample $ST=\{0, V, 0, 0\}$. The average time between the arrival of packets in the DTN = $250 \mu s$. The minimum value of $T_2 = 3.26$ ms is reached at $V = 3.35$.

![Figure 2](image_url)

**Figure 2.** The dependence of $T_2$ on the value of the sample parameter $V$

Figure 3 shows the dependence of the average time of transmission of packets by DTN ($T_2$) on the intensity of arrivals of the packets in the DTN for samples $ST1 = \{ -, -, 0, -\}$ and $ST2 = \{0, 3.5, 0, 0\}$.

![Figure 3](image_url)

**Figure 3.** The dependence of $T_2$ on the intensity for $ST1$ and $ST2$

Figure 3 shows that the greater the load on the DTN, the more important are the parameters $Q$, $V$, $T$ to reduce $T_2$. 
4. Conclusions
The idea of using samples for selecting the direction of purposeful movement on oriented graphs was tested. Testing was carried out on the processes of choosing routes in the DTN. For this purpose we have chosen a set of parameters of the sample. GPSSW developed a simulation model of a fragment of the DTN, the values of the parameters of the sample of the conducted, simulation experiments. The results of the experiments showed the possibility and feasibility of using samples to select routes.

Samples can be used to select routes for targeted traffic on oriented graphs in systems for various purposes: technical, economic, social. In the future, it is planned to test samples in the organization of traffic in road networks.

5. References
[1] Zagoruiko N D 2013 Cognitive data analysis (Novosibirsk: Academic publishing house "geo") p 186
[2] Purtov A M 2016 Simulation of control systems by streams of cars at the intersection Omsk scientific Bulletin, 3 (147) pp 92-96
[3] Boev V D 2004 Modeling of systems. Tools GPSS World: textbook manual (Petersburg: SPb.: Peter) p 368
[4] Zadorozhnyi V N, Dolgushin D Y.and Yudin E B 2013 Analytical and simulation methods for solving actual problems of system analysis of large networks: monograph (Omsk: Publishing house OmSTU) p 324
[5] Zadorozhnyi V N and Yudin E B 2016 Dynamic equations of node degrees in growing networks with connection losses Dynamics of Systems, Mechanisms and Machines, Dynamics 2016 art 7819111 DOI: 10.1109/Dynamics.2016.7819111
[6] Oliver V G and Oliver N A 2011 Computer networks. Principles, technologies, protocols. Textbook for universities (Petersburg: SPb.: Peter) p 944

Acknowledgments
«The work was supported by the program of fundamental scientific researches of the SB RAS № 1.5.1., project № 0314-2019-0020»