

SIMULATION-OPTIMIZATION MODEL FOR DYNAMIC DISTRIBUTION OF INFORMATION FLOWS IN A HETEROGENEOUS DATA TRANSMISSION NETWORK

Malika Baxadirovna Mirzaeva

Doctor of Philosophy in Technical Sciences (PhD), Associate Professor, “Tashkent Institute of Irrigation and Agricultural Mechanization Engineers” National Research University. Tashkent International University
malikamirzaeva01@gmail.com

Abstract:

This paper considers a heterogeneous data transmission network (DTN) comprising several mobile information sources and a single central receiver, jointly using radio and wired communication channels. Each channel is characterized by its throughput, service intensity, and probability of successful packet delivery. For every newly arrived packet, the delivery time is estimated along all admissible routes, and the route with the smallest expected delay is selected, thereby providing adaptive traffic distribution among alternative paths. The simulation-optimization model of the network is built on a discrete-event approach. Its effectiveness was assessed through five experiment series examining the effects of source intensity, packet size, alternative-route usage, delivery probability, and channel capacity on the network's time characteristics. The results show that as the delivery probability increases from 0.1 to 1.0, the maximum delay decreases by approximately 90.6%, while beyond a certain packet-size threshold the transmission time begins to dominate over the waiting time, accompanied by a marked decline in the frequency of alternative-route usage. These results can serve as a basis for practical recommendations on justifying channel parameters and improving adaptive routing mechanisms when designing heterogeneous DTNs.

Keywords: heterogeneous network, multi-source traffic, adaptive routing, delivery probability, discrete-event simulation, alternative routes, simulation-optimization model.

Introduction

The growth of the volume of transmitted information, the increasing requirements for the speed of message delivery, and the need to ensure high reliability are further increasing interest in methods for mathematical and simulation modeling of the operation of data transmission networks (DTNs) [1]. In many practical systems, monitoring, telemetry, control of moving objects, and distributed information systems, a network structure in which information is transmitted from several information sources to a single central receiver is widely used. A distinctive feature of such networks is the combined use of transmission channels of different nature (radio channels and wire channels) [2]. While radio channels are characterized by a high probability of errors due to interference, variability of radio wave propagation conditions, and source movement, wire channels, although more stable, are also subject to random packet loss and temporary overloads. As a result, the quality of data transmission is determined not only by the channel capacity, but also by the probability of successful packet delivery [3]. In such conditions, the network under study is a hybrid MUT with multi-source traffic, multi-route transmission, and probabilistic channel operation. When there are several alternative routes between the information source and the receiver, the problem of choosing an efficient path arises. In traditional networks, routing is often carried out based on fixed metrics (number of hops, route length), but this approach does not take into account the current state of the network and can lead to overloading of some links [4]. Modern routing methods are aimed at using dynamic characteristics such as the current load of channels, queue sizes, and the predicted value of the delay. The purpose of the study is to develop a simulation-optimization model that takes into account the dynamic distribution of information flows between alternative routes in a heterogeneous MUT with several moving information sources and one central receiver, and to evaluate its effectiveness based on computational experiments [5].

Methodology

The extensive use of probability theory, mathematical statistics, and queueing theory methods in describing data flows is justified by the works of Kleinrock and Harchol-Balter. Regarding the issues of channel-level reliability and delivery probability, Gao et al. analytically modeled the end-to-end delay in mobile ad-hoc networks using a quasi-birth-death process, which provides a deep understanding of the impact of transmission errors and retransmissions on network delay. Stepanov et al. studied the model of traffic service via a dedicated resource in an LTE cell, and Osei Kofi and Ahene studied the load balancing algorithm in software-defined networks, both of which offer similar solutions to the problem of resource allocation in a multi-resource environment. However, most of the existing research either focuses only on the processes within the node (local) or only on the routing issues at the network level. The need for a single simulation-optimization model that simultaneously integrates multiple source traffic, probabilistic nature of channels, and adaptive route selection has not been adequately addressed in existing research.

The theoretical foundations of simulation modeling have been extensively covered in the framework of discrete-event systems theory by Zeigler, Muzy, and Kofman and Cassandras and Lafortune. The formalisms proposed in these works form the methodological basis of the model in this study. In the direction of accelerating discrete-event simulation, Hong, Song, and Wang proposed a method based on the Euler approximation, which may be useful in the future for extending this model to larger networks. Modern algorithmic solutions for load balancing and routing are reviewed in the works of Balseiro et al. and Yao, Ding, and Clausen.

Thus, the analysis conducted shows the need to study heterogeneous MUT with multi-source, multi-route and probabilistic nature within the framework of a single simulation-optimization model. Filling this gap is the main task of the article. Issues related to this area were also considered in the author's previous works. Research was conducted on the evaluation of queueing-theoretic time characteristics at the pre-design stage of computing systems as well as comparative analysis and simulation of increasing the efficiency of data transmission in telecommunication systems using neural networks. The article logically continues those studies, focusing on the issue of dynamic resource allocation at the network level. The methodological basis of the proposed model is

described below.

The object of study is a data transmission network consisting of several moving information sources and one central receiver. Each source is a moving object that forms an independent packet stream in the process of performing its functional task. Information transmission from the source to the main information and computing center is carried out in two stages. First, each moving object transmits information via a radio channel to the nearest moving radio center, and then via a repeater to a stationary radio center. Then the packets are delivered to the central receiving point via intermediate nodes (switching centers) along the wired network. There may be several alternative routes between the source and the receiver, differing in the number of intermediate nodes, channel characteristics, and current load level.

Each packet is generated at a random time interval according to a given distribution law. Three common types of distributions are used to conduct computational experiments:

Poisson distribution - used in modeling random and independent events;

Erlang distribution - describes relatively regular flows, taking into account the correlation between events;

Uniform distribution used in the study of almost deterministic regimes.

The size of the packet is also randomly generated and directly affects the transmission time over the channel, and therefore the overall delivery delay. The multi-source nature of traffic leads to streams from different objects competing for shared channel and intermediate node resources, resulting in a dynamic network state.

Unlike most classical MUT models, in this work communication channels are not assumed to be perfectly reliable — their probabilistic nature is taken into account. Each channel connecting nodes i and j is assigned the following set of parameters: channel capacity, service intensity, and probability of successful packet delivery. Capacity is the maximum amount of data that can be transmitted over the channel per unit of time, service intensity is the average number of requests that the channel can process per unit of time, and the delivery probability reflects the impact of transmission errors, data loss, and device rejections. The developed model estimates the delivery time for each newly arrived packet along all allowed routes. This takes into account the channel capacity, packet size, current channel idle times, and retransmission probabilities. As a result of the estimation, the route with the smallest expected delay is selected. This approach provides adaptive distribution of traffic between alternative routes and helps to more evenly use network resources. This mechanism is especially important in conditions of high traffic intensity, since in such cases local overloads in some sections of the network can significantly degrade the quality of service. The transmission of a packet over a communication channel includes the following stages: waiting for the channel to become free, transmitting the data directly, receiving confirmation of delivery, and retransmitting the packet if necessary. For a packet of size Q and a channel with capacity C , the base transmission time is given by $t_{tr} = Q / C$. However, the real service time depends on the current state of the channel: if the channel is busy at the moment of packet arrival, a waiting queue is formed. The model stores its idle time for each channel; based on the packet arrival time, the real start and end times of transmission are calculated, and then the channel idle time is updated. This approach allows you to model the formation of queues directly on communication channels, without introducing separate service objects.

Based on the formalization, the input parameters of the model include: the number of information sources, the intensity of packet formation, the distribution laws of input flows, the minimum and maximum packet sizes, the bandwidth of the channels, the probability of packet delivery, and the number of transmitted messages. The output characteristics of the model include the average and maximum message delivery time, the average waiting time in the network, the loading of individual channels and routes, the flatness coefficients of the load distribution, the number of retransmissions, and the share of using alternative routes. Analytical research of the operation processes of such networks faces great difficulties due to a number of factors. The network has several random flow-forming sources, the use of multi-route transmission (in which the choice of route depends on the current state of the network), the probability of packet delivery leads to the randomness of the

number of retransmissions, and the dynamic change of queues in the channels and their load levels. This set of factors complicates obtaining accurate analytical solutions even for relatively small networks, therefore, discrete-event simulation modeling, which allows for iterative modeling of network performance over time, becomes the most effective research tool.

The modeling process begins with generating a packet stream for all sources. For each packet, its time of appearance in the network and its size are determined, then a transmission route is selected and the process of passing the packet through the sequence of channels included in the selected route is modeled. The main types of events are: packet generation by the source, packet arrival on the channel, start and end of channel service for the packet, packet arrival at the next node, packet delivery to the recipient, and retransmission of the packet if necessary. The result of the modeling is a set of statistical data characterizing the network performance: the average and maximum message delivery time, the level of channel load, and the frequency of using alternative routes.

Results

To evaluate the effectiveness of the model, a network consisting of four information sources, one central receiver, a set of intermediate communication nodes, radio transmission channels, and wired channels of the backbone network was selected as the object of study (Figure 1) [6]. A series of five computational experiments were conducted to study the impact of input traffic characteristics, channel parameters, and routing mechanisms on network efficiency.

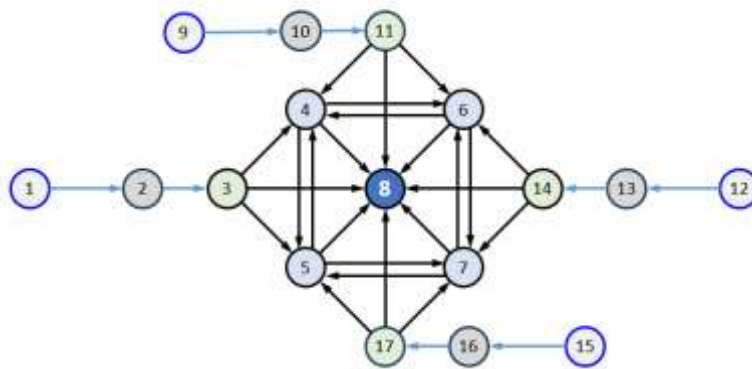


Figure 1. Graph of the data transmission network under study

The first set of experiments investigated the effect of the intensity of input flows on the network performance. A symmetric regime was considered, in which all sources generate messages with the same intensity [7]. Table 1 presents the average and maximum delay values obtained various values of the network's total input-flow intensity λ .

Table 1

Delay values (s) at different input current intensities

λ , pkt/s	200	400	600	800	1000	1200	1400	1600	1800	2000
$T_{\max}$	0.015	0.017	0.024	0.023	0.03	0.029	0.026	0.035	0.032	0.042
T_{aver}	0.013	0.014	0.017	0.018	0.019	0.019	0.022	0.029	0.025	0.032

The relationship in the table is not monotonic, because in different experiments the intervals of arrival of packets and their size are generated based on different distributions. This variability affects both the transmission time and the waiting time in the queue, so the delay value may temporarily decrease at some points as the load increases [8]. However, as a general trend, the delay values increase with increasing load: $\lambda = 200$ in the pkt/s $T_{\max}$ if only 0.015 s, whereas at $\lambda = 2000$ pkt/s it reaches 0.042 s. This result indicates the need to control the overall network load to ensure the specified delay limit. In the second series of experiments, the effect of the size of the transmitted

messages was studied. Table 2 shows the packet size Q . The maximum (obtained at different values of $T_{\max}$), minimum ($T_{\min}$) and average (T_{aver}) delay values, as well as their difference (T_d) are presented [9].

Table 2

Delay characteristics depending on packet size (s)

Q , bytes	100	300	500	700	900	1100	1300	1500	1700	1900	2000
$T_{\max}$	0.391	0.368	0.421	0.400	0.390	0.405	0.443	0.411	0.475	0.532	0.609
$T_{\min}$	0.383	0.359	0.409	0.386	0.380	0.393	0.429	0.401	0.432	0.473	0.522
T_d	0.009	0.009	0.012	0.015	0.010	0.012	0.014	0.011	0.043	0.059	0.087
T_{aver}	0.387	0.363	0.417	0.393	0.384	0.398	0.435	0.405	0.464	0.517	0.596

From the table data, it can be seen that approximately $Q = 1500$. At sizes up to 1500 bytes, the maximum delay remains relatively stable (in the range of 0.37–0.44 s), and at this stage, waiting and routing processes play a major role. However, $Q > 1500$ starting from the byte, the delay grows significantly faster: $Q = 1700$ bytes $T_{\max}$ reaches 0.475 s, and at $Q = 2000$ bytes it reaches 0.609 s. In this case, the main part of the delay begins to be the time of direct transmission of the packet over the channel. Also, in large packets dT ($T_{\max}$ and $T_{\min}$ difference) also increases sharply ($Q = 1900$ bytes 0.059 s, $Q = 2000$ bytes 0.087 s), which indicates increased load unevenness [10].

The third set of experiments analyzed the use of alternative routes. When the packet size is relatively small, the frequency of using alternative routes is high, because in this case all routes have approximately the same expected delay. When the packet size increases and exceeds a certain threshold (approximately $Q > 1500$ bytes), the direct transmission time begins to dominate the total delivery time. As a result, the frequency of using alternative routes with more communication channels (i.e., longer ones) is significantly reduced, since the total transmission time on such routes naturally increases. This result confirms the tendency of the adaptive routing mechanism to prefer shorter routes for large packets. In the fourth set of experiments, the effect of the successful packet-delivery probability P_{deliv} on the network's time characteristics was studied. The same value of P_{deliv} was assigned to each channel and varied from 0.1 to 1.0 (Table 3, Figure 2).

Table 3

Time characteristics depending on the probability of delivery (s)

P_{deliv}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$T_{\max}$	4.084	1.190	0.576	0.422	0.409	0.376	0.404	0.408	0.395	0.383
$T_{\min}$	3.346	1.015	0.505	0.391	0.394	0.350	0.390	0.394	0.384	0.376
T_d	0.738	0.174	0.072	0.031	0.016	0.026	0.014	0.014	0.011	0.007
T_{aver}	3.461	1.056	0.534	0.405	0.400	0.364	0.395	0.401	0.390	0.379

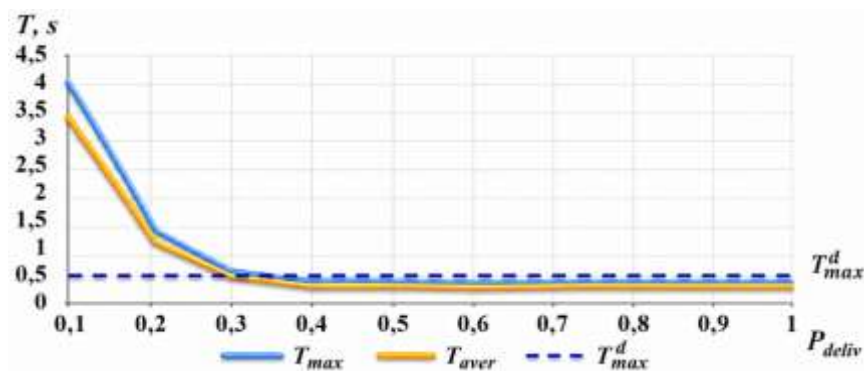


Figure 2. Dependence of data transmission time characteristics on delivery probability

The results clearly demonstrate a strong inverse relationship between delivery probability and delay: at $P_{deliv} = 0.1$, $T_{max} = 4.084$ s, whereas at $P_{deliv} = 1.0$. This value decreases to 0.383 s, i.e., it decreases by $(4.084 - 0.383) / 4.084 \approx 90.6\%$. Such a sharp difference is explained by the increase in the number of repeated transmissions at low delivery probability: the smaller the value of P_{deliv} , the more times a packet must be retransmitted to ensure successful delivery, resulting in the transmission time becoming dominant in the overall delivery time [11]. The diagram also shows that approximately $P_{deliv} \geq 0.4$ when T_{max} value approaches the limit permissible level and stabilizes; therefore, for the network to operate efficiently, each channel must have $P_{deliv} \geq 0.4$. In the fifth experimental series, the channel capacity C . The effect of C on network timing characteristics was investigated. For a packet of size Q , the transmission time is $t_{tr} = Q / C$. As can be seen from the expression, an increase in bandwidth leads to a decrease in packet transmission time and, as a result, a decrease in the overall message delivery delay [12]. The results confirmed the expected inverse (close to hyperbolic) relationship between channel bandwidth and delay: at low values of bandwidth, the delay is sharply higher: as bandwidth increases, the delay decreases rapidly and then reaches a nearly stable level. A set of five series of computational experiments has shown that the proposed simulation-optimization model allows: (1) determining the overall network load limits, (2) assessing the relationship between packet size and route selection, (3) quantitatively analyzing the impact of channel reliability (delivery probability) on delay, and (4) drawing practical conclusions for bandwidth planning [13].

Discussion

Comparison of the obtained results with existing studies confirms a number of general patterns. In particular, the strong inverse effect of the probability of delivery on the delay is consistent with the analytical end-to-end delay model for mobile ad-hoc networks presented by Gao et al. There, it was also shown that with a decrease in channel reliability, the number of retransmissions and, therefore, the total delay increases sharply. In this sense, the 90.6% reduction value obtained in this paper is a quantitative confirmation of the theoretically expected effect [14]. The decrease in the frequency of using alternative routes with an increase in packet size is consistent with the “longer path — higher cost” principle noted in classical routing theory and is consistent with the load balancing approaches proposed by Stepanov et al. and Osei Kofi and Ahene: in both studies, it is shown that resource selection depends not only on the current load, but also on the expected service duration. In contrast, the distributed gradient descent algorithm proposed by Balseiro et al. proves that stable equilibrium can be achieved even in a decentralized environment. This trend suggests the possibility of using decentralized routing mechanisms in the future development of this model.

The main limitation of the study is that the computational experiments were conducted on a relatively small (four-source) network topology [15]. The accelerated discrete-event modeling methods proposed by Zeigler, Muzy, and Kofman and Hong, Song, and Wang may provide significant computational savings when testing the model on larger, multi-node networks in the future.

Conclusion

As part of the research, a simulation-optimization model was developed that takes into account the dynamic distribution of information flows between alternative routes in a heterogeneous data transmission network consisting of several moving information sources and one central receiver. The model is based on a discrete-event approach, in which each channel is characterized by its capacity, service intensity, and delivery probability, and route selection is based on the principle of minimizing the expected delay.

Based on the five series of calculation experiments conducted, the following conclusions were reached:

- as the network's total input-flow intensity increases, delay values tend to rise: over the range $\lambda = 200$ to 2000 pkt/s $T_{\max}$ increase by almost 2.8 times;
- when the packet size exceeds approximately 1500 bytes, the transmission time begins to dominate the latency, and at the same time, the frequency of using alternative (longer) routes decreases significantly;
- as the packet delivery-success probability increases from 0.1 to 1.0, the maximum delay decreases by approximately 90.6%, for efficient operation, a delivery probability of $P_{\text{deliv}} \geq 0.4$ is recommended for each channel;
- increasing channel capacity leads to a decrease in latency, but this relationship is not linear — after a certain value, latency becomes almost constant.

The results obtained serve as a basis for justifying channel parameters (throughput, delivery probability) in the design of heterogeneous data transmission networks, improving adaptive routing mechanisms, and developing practical recommendations for the efficient use of network resources. As a next step in the research, it is planned to combine this network-level model with the author's previously developed node-level simulation-optimization model to form a comprehensive multi-stage efficiency assessment of MUT.

REFERENCES

- [1] A. Elahi and A. Cushman, *Computer Networks: Data Communications, Internet and Security*. Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-031-42018-4.
- [2] L. Kleinrock, *Teoriya Massovogo Obsluzhivaniya* [Queueing Theory], I. I. Grushko, Trans., V. I. Neyman, Ed. Moscow, USSR: Mashinostroenie, 1979.
- [3] V. M. Vishnevsky, *Teoreticheskie Osnovy Proektirovaniya Komp'yuternykh Setey* [Theoretical Foundations of Computer Network Design]. Moscow, Russia: Tekhnosfera, 2003.
- [4] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
- [5] J. Gao, Y. Shen, X. Jiang, O. Takahashi, and N. Shiratori, "End-to-end delay modeling for mobile ad hoc networks: A quasi-birth-and-death approach," *Ad Hoc Sens. Wireless Netw.*, vol. 29, no. 1–4, pp. 265–285, 2015.
- [6] M. S. Stepanov, M. G. Kanishcheva, E. E. Malikova *et al.*, "The development and analysis of a service model for the traffic of a surveillance system operator by a dedicated resource of an LTE cell," in *Proc. 2022 Syst. Signals Generating Process. Field On Board Commun.*, 2022, pp. 1–6.
- [7] E. Osei Kofi and E. Ahene, "Enhanced network load balancing technique for efficient performance in software defined network," *PLOS ONE*, vol. 18, no. 4, Art. no. e0284176, 2023, doi: 10.1371/journal.pone.0284176.
- [8] B. P. Zeigler, A. Muzy, and E. Kofman, *Theory of Modeling and Simulation: Discrete Event & Iterative System Computational Foundations*, 3rd ed. London, U.K.: Academic Press, 2018.
- [9] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*, 3rd ed. Cham, Switzerland: Springer, 2021.
- [10] L. J. Hong, Y. Song, and T. Wang, "Fast discrete-event simulation of Markovian queueing networks through Euler approximation," 2024, *arXiv:2402.13259*.

- [11] S. R. Balseiro, V. S. Mirrokni, and B. Wydrowski, "Load balancing with network latencies via distributed gradient descent," 2025, *arXiv:2504.10693*.
- [12] Z. Yao, Z. Ding, and T. Clausen, "Reinforced cooperative load balancing in data center," 2022, *arXiv:2201.11727*.
- [13] M. B. Mirzayeva, "Queuing-theoretic temporal characterization of computing systems at the pre-project stage of IPTS design," *Educ. Sci. Innov.*, vol. 6, pp. 12–17, 2026.
- [14] M. B. Mirzayeva, "Telekommunikatsiya tizimlarida ma'lumotlarni uzatish samaradorligini neyron tarmoqlar yordamida oshirishning qiyosiy tahlili va simulyatsiyasi" [Comparative analysis and simulation of improving data transmission efficiency in telecommunication systems using neural networks], *Raqamli Transformatsiya va Sun'iy Intellekt*, vol. 4, no. 4, pp. 106–110, 2026, doi: 10.5281/zenodo.22107751.
- [15] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. New York, NY, USA: Wiley, 1991.