MATHEMATICAL MODELING AND COMPUTING, Vol. 9, No. 4, pp. 936–944 (2022)

Entropy calculation for networks with determined values of flows in nodes

Politanskyi R. L.[1], Bobalo Y. Y.[2], Zarytska O. L.[2], Kiselychnyk M. D.[2], Vistak M. V.[3]

[1] *Yuriy Fedkovych Chernivtsi National University, 2 Kotsyubynsky Str., 58012, Chernivtsi, Ukraine*
[2] *Lviv Polytechnic National University, 12 S. Bandera Str., 79013, Lviv, Ukraine*
[3] *Danylo Halytsky Lviv National Medical University, 69 Pekarska Str., 79010, Lviv, Ukraine*

(Received 20 January 2022; Revised 6 August 2022; Accepted 5 October 2022)

The paper analyses a network with given input and output flows in each of its nodes. The basis of this analysis is the algorithm for determining the set of solutions of the linear equations system, using the Gaussian method. The power of the set determines the structural entropy of the system. By introducing uncertainty into the value of part of the information flows, the deviation of the network from its equilibrium state is simulated. The set of potential solutions, as a part of the total set of the system solutions, determines the statistical entropy of the system. The probability entropy is calculated for a network with four nodes and a total flow of 10 erlangs with a sampling step of 1 erlang. Calculated entropy values for 1, 2, 3, and 4 uncertain flows out of a total of 16 flows that are transmitted between nodes of the fully connected network. As a result of the conducted statistical analysis of entropy values, the optimal number of statistical intervals for entropy values is determined: 4, 11, 24, and 43 intervals for 1, 2, 3, and 4 uncertain flows, respectively. This makes it possible to highlight the set of flows in the system that have the greatest influence on the entropy value in the system. The obtained results are of practical importance, as they enable the detection of deviations of the network from its equilibrium state by monitoring the passage of traffic on individual branches of a complex telecommunication network. Since, as shown in our previous works, the task of determining the complete set of solutions of the system for the number of nodes greater than 4 has a significant computational complexity, the application of the algorithm to such networks requires an increase in the discretization step of the values of information flows in the network. Another way to reduce computational complexity can be to reduce the set of analysed solutions to a subset of solutions close to the equilibrium state of the system.

Keywords: *entropy, solution of a system of linear equations, self-similarity, multifractality.*

2010 MSC: 15A03, 15A06, 68-04, 68R99 **DOI:** 10.23939/mmc2022.04.936

1. Introduction

Analysis of the processes occurring in telecommunication systems with given values of the total flows in the nodes of the system is a classic task that attracts the attention of many Ukrainian and foreign researchers. Traditionally, the most developed and widespread of them are the tensor method of network modeling [1,2], methods based on the Kerners theory of three phases [3], as well as numerous modifications of random graphs theory (Erdos–Reney, Watts–Strogats, Barabashi–Alberta models), percolation theory [4]. This is important for ensuring the security of information systems using telecommunication technologies [5,6], as well as other aspects of protected data transfer [7–9].

Nevertheless, most of them provide an opportunity to obtain only integrated indicators of the entire network. At the same time, the software and hardware used by the main telecommunication means of service providers are being improved, and the introduction of neural network technologies allows the use of complex algorithms to analyze the state of the network. In addition, the requirements for the complexity of the systems being studied to enable the obtaining of significant results in a practical sense are growing significantly. This creates an additional incentive for the study of existing and development of new methods and algorithms for the study of networks with information flows, which impose additional conditions dictated by the content of the applied problems to be solved.

One of the possible methods of detailed analysis of the network state with the specified operating conditions, due to both technical and economic factors of its operation, is to find all possible states. Identifying all possible states and predicting transitions between them is the most detailed and reliable information about the state of the network. Therefore, determining the set of all possible states and predicting their change over time makes it possible to detect any changes due to external influences: increase or decrease bandwidth between network nodes, increase or decrease traffic in its nodes and any other changes. This makes it possible to improve load balancing algorithms and increase security guarantees in information transmission systems.

A detailed analysis of the network makes it possible to apply the principle of maximum entropy. This algorithm has been traditional for a long time in the tasks of finding information flows in networks that arise in various fields of science and technology — from modeling of financial flows to tasks of load balancing in telecommunication networks [10–12]. At the same time, some authors emphasize that the maximum entropy method is more effective among the other traditional methods (fuzzy control and heuristic strategy, the center of mass of the cluster, round–robin scheduling) [13].

One of the most important areas where the principle of maximum entropy can be applied is the prediction of dynamic changes in networks, such studies are carried out in [14].

Another important aspect of the development of the maximum entropy method is its use in such modern branches of science as artificial intelligence and quantum computing. The prediction of dynamic changes in quantum mechanical systems, which are expected to be used in the latest technologies of quantum communication and the Internet, is given in [15]. Algorithms for determining equilibrium states in systems with flows between its nodes are used to develop methods of artificial intelligence and machine learning [16].

If a set of system states is determined, it is theoretically possible to predict system states at future observation points based on correlations between network states that form so-called Markov sequences.

For the least complex discrete systems, in which a correlation is observed only between adjacent observation points, the state of the system can be described by a simple model based on a matrix of transitions between system states:

$$P^n = \|A\|^n \times P^0 \tag{1}$$

where the P^0 denotes vector of probabilities of the initial state of the system, the P^n denotes probability vector at the n-th observation moment, the $\|A\|$ denotes a square matrix of transitions between system states.

The model based on the transition matrix is convenient to use if the dimension of the discrete system, or the number of its states is not very large.

2. Statement of the problem

2.1. Consideration of the system entropy

Previously, we have studied a system with specified values of flows in each of its nodes [17]. It has been shown that even with a small number of nodes and low power of a set of possible flow values in the system, the total number of its solutions is quite large, amounting to more than two million for systems with three nodes and total traffic of 20, and systems with four nodes and total traffic of 10. This makes it impossible to apply the classical model (1) because the dimension of the matrix of transitions between possible states of the system becomes huge:

$$\dim \|A\| > 2 \cdot 10^6 \times 2 \cdot 10^6.$$

Therefore, the paper considers a different approach to the analysis of changes in the system state over time, based on the concept of entropy and the most general pattern, it means that all processes in a closed system occur in the direction of its increase. Entropy is one of the numerical characteristics of the degree of complexity of systems that have different origins, can be both artificial and natural systems. Its definition is based on the number of possible states of the system and the probability of the system to be in one state or another. There are definitions of entropy for discrete and continuous systems. For discrete systems, entropy is determined by adding for all possible states the probability of this state multiplied by the logarithm of this probability taken with the inverse sign:

$$H = -\sum_i p_i \cdot \log(p_i),$$

where the index "i" means addition over all possible states of the system.

For continuous systems, the concept of differential entropy is introduced in many sources (as an example [18]), which is used in the theory of information transfer:

$$H = -\int_{\infty}^{\infty} p(x) \cdot \log[p(x)]\, dx.$$

where $p(x)$ is the probability density of x values that describe the state of the system.

The concept of entropy can be applied not only to static systems that are in one of the possible states, but also to processes that are determined by changes in the states of the system. Changes in the states of the system can occur both independently of each other and statistically correlated transitions from one state to another, such systems and processes are called Markov. A method based on the uncertainty of some system parameters has been developed to calculate the entropy of such systems. The entropy of such a system is determined on the basis of many possible values of uncertain parameters of the system, and this uses the classical concept of entropy of a discrete system with some set of parameter values that determine its state.

2.2. Determining the set of the system solutions

In [17] the algorithm for determining the complete set of solutions for a network of information flows with constraints on the value of the flow in each node of the system is described. By such constraints, we have chosen the values of the total values of input and output streams for each node of the system. In addition, there are restrictions on stream values: each value is a non-negative integer. The flow matrix of such a system in the most general formulation of the problem is formed from by n^2 flows between each pair of system nodes.

To determine the complete set of the system solutions, the Gaussian method was used for a system of $2 \times n$ equations, where one equation is linearly dependent on the other due to the condition of equality of flows in the input and output nodes of the system. Based on the structure of the system fundamental solution, it can be argued that the whole set of solutions can be determined by the set $(n-1)^2$ of non-negative integers, which we will define as a kernel of the system solutions set. The power of this set, can also be determined by the Gaussian method applied to the system of equations for a matrix of flows of dimension $n-1$.

Next, we present the basic relations that state the problem for the system of dimension n. Therefore, the matrix of flows in the system is a square $n \times n$ matrix:

$$I = \begin{pmatrix} i_{1,1} & \cdots & i_{1,n} \\ \vdots & \vdots & \vdots \\ i_{n,1} & \cdots & i_{n,n} \end{pmatrix}. \tag{2}$$

The system that defines the general set of solutions of the problem is :

$$\begin{cases} i_{1,1} + \cdots + i_{1,n} = s_1, \\ \dots\dots\dots\dots\dots\dots, \\ i_{n,1} + \cdots + i_{n,n} = s_n, \\ i_{1,1} + \cdots + i_{n,1} = d_1, \\ \dots\dots\dots\dots\dots\dots, \\ i_{1,n} + \cdots + i_{n,n} = d_n, \end{cases} \tag{3}$$

where $\{s_1, \dots, s_n\}$ is the set of input flows, and $\{d_1, \dots, d_n\}$ is the set of output flows.

3. The main results

3.1. Calculation of entropy by introducing indeterminate flows in the system

Consider an example of calculations for a system formed by four nodes with specified for each node values of input $\{s_1, s_2, s_3, s_4\}$ and output $\{d_1, d_2, d_3, d_4\}$ flows and a given condition of equality of sums of input and output flows: $\sum_{i=1}^{4} s_i = \sum_{i=1}^{4} d_i = S$. The algorithm defines a set of the system solutions, each of them represents 16 unknown values of the flow matrix x_i. Each such type solution can be expressed in 9 independent variables λ_j and therefore each set of variables expresses a unique solution of the problem. The analytical expression that defines the set of solutions of the problem has the following form:

$$\begin{cases} i_{11} = d_1 - s_2 - s_3 - s_4 + \sum_{i=1}^{9} \lambda_i, \\ i_{12} = d_2 - (\lambda_2 + \lambda_3 + \lambda_7), \\ i_{13} = d_3 - (\lambda_4 + \lambda_6 + \lambda_8), \\ i_{14} = d_4 - (\lambda_1 + \lambda_5 + \lambda_9), \\ i_{21} = s_2 - (\lambda_1 + \lambda_2 + \lambda_6), \\ i_{31} = \lambda_2, \\ i_{41} = \lambda_6, \\ i_{24} = \lambda_1, \\ i_{22} = s_3 - (\lambda_3 + \lambda_4 + \lambda_5), \\ i_{32} = \lambda_3, \\ i_{33} = \lambda_4, \\ i_{34} = \lambda_5, \\ i_{23} = s_4 - (\lambda_7 + \lambda_8 + \lambda_9), \\ i_{42} = \lambda_7, \\ i_{43} = \lambda_8, \\ i_{44} = \lambda_9. \end{cases} \tag{4}$$

This is the general solution of the system (3).

3.2. Determination of entropy by introducing indeterminate variables in the system flow matrix

Let us analyze the complexity of the problem associated with all possible ways to select independent variables for the same problem of calculating the states of a system with four nodes. Based on the general theory of the linear equations systems, it is obvious that one can choose from one to nine independent system variables that correspond to individual elements of the matrix of information flows (2). The number of possible choices from one to three independent variables is determined by the binomial coefficient, as they can all be independent:

$$C_{16}^{1} = 16, \quad C_{16}^{2} = 120, \quad C_{16}^{3} = 560.$$

When choosing 4 or more (up to 9, the maximum number of independent variables), keep in mind that not every combination of variables is a set of independent variables. Thus, for four variables that are in one row or column, one of the variables can be expressed through others, as well as for 5, 6, 7, 8 and 9 variables, of which 4 are in one row or column.

Next, let us look at an example of solving the problem for 5 levels of ranking the values of information flows, if the total traffic in the system is 10 erlangs,

$$\{s_1, s_2, s_3, s_4\} = \{1, 4, 3, 2\}, \quad \{d_1, d_2, d_3, d_4\} = \{4, 0, 2, 4\}. \tag{5}$$

The total number of solutions of the problem is 54795. Then the entropy of the problem, provided that all flows are defined, is equal to

$$H_0 = \log_{10}(54795) = 4.739.$$

The developed algorithm also enables to determine the entropy of system states, provided that the value of one or more variables in the system is unknown. As mentioned above, the total number of solutions in the system given by system (4) and boundary conditions (5) is 54795, and the values of the components of the matrix (2) belong to the set of integers $\{0,1,2,3,4,5,6\}$.

The following are the results obtained by calculating the entropy of the system for all possible combinations of unknown variables, namely the minimum, maximum, mathematical expectation and entropy variance for the number of independent variables from one to four. For more variables, the task becomes too difficult to perform on a personal computer. Its solution is possible only through the use of more powerful computing tools and with the use of algorithms for parallel calculations.

Table 1 shows the largest and smallest values of entropy determined on sets of 1, 2, 3 and 4 indeterminate system variables, as well as mathematical expectations and standard deviations of these values, the number of series members and the number of intervals in the statistical distribution.

Table 1. The value of the entropy of the matrix of information flows and its statistical characteristics.

The number of flows	$H_{\min}$	$H_{\max}$	H_{av}	H_{stdev}	The number of sets	The number of ranges
1	1.29	1.75	1.50	0.15	16	4
2	2.51	3.39	2.97	0.19	120	11
3	3.54	4.88	4.38	0.23	560	24
4	4.89	6.25	5.74	0.25	1812	43

Figures 1–3 show the entropy values calculated at the network nodes for one, two, three and four undefined flows respectively.

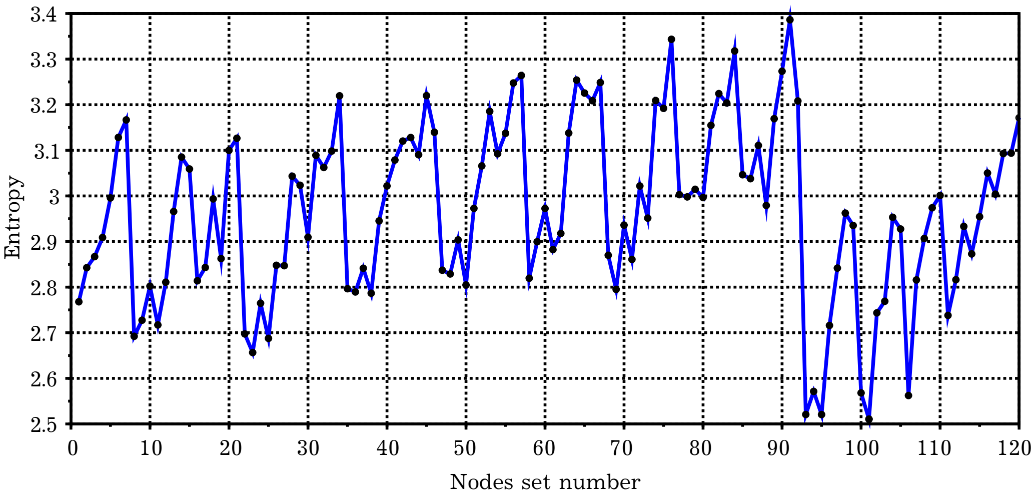

Fig. 1. The entropy of a network formed by four nodes is calculated by two indeterminate flows.

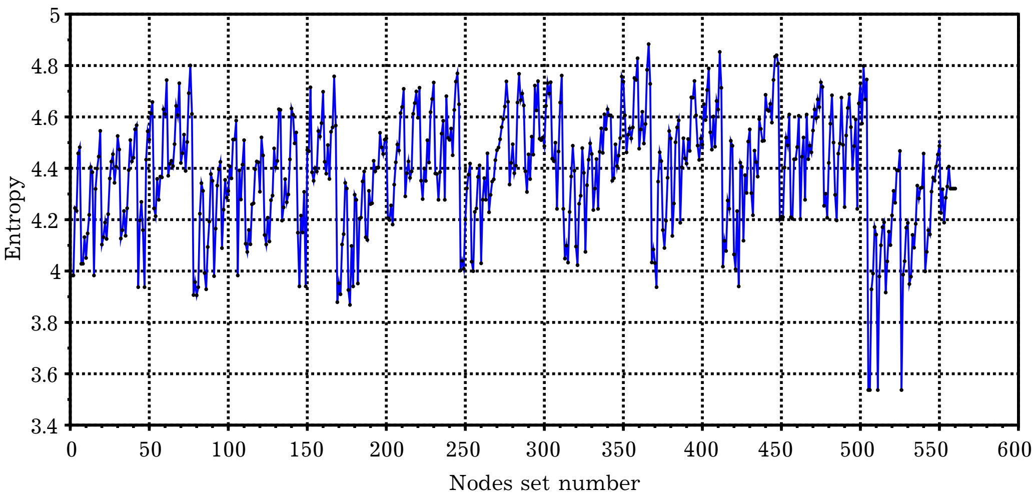

Fig. 2. The entropy of a network formed by four nodes is calculated by three indeterminate flows.

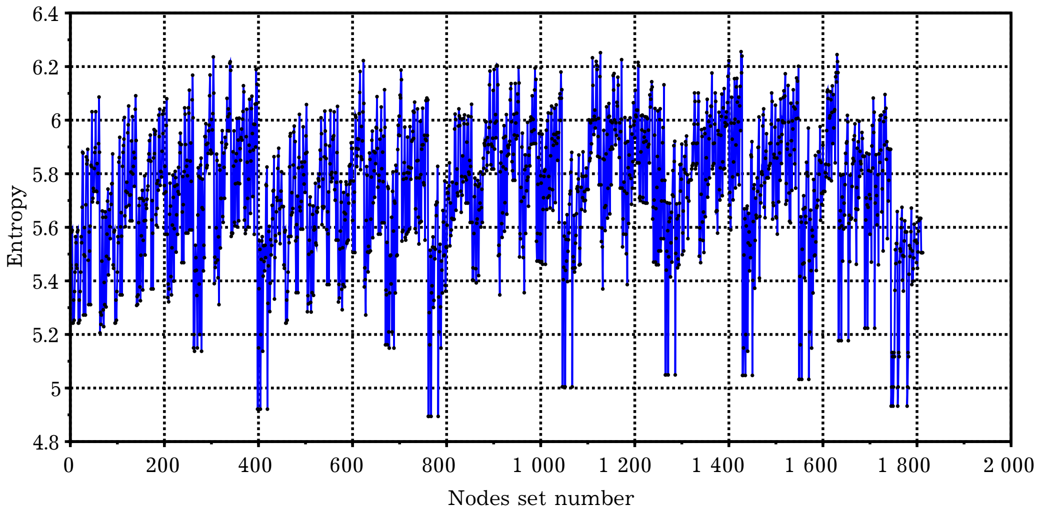

Fig. 3. The entropy of a network formed by four nodes is calculated by four indeterminate flows.

The form of the given entropy graphs allows us to conclude that the entropy of the system determined by us has the property of self-similarity on the set of tuples of independent variables that determine individual values of flows in the network.

In Figs. 4, 5 shows the histograms of the distribution of entropy values obtained for the studied system and for the entropy calculated for 3 and 4 undetermined nodes respectively, the number of entropy value intervals on these histograms is 24 and 43.

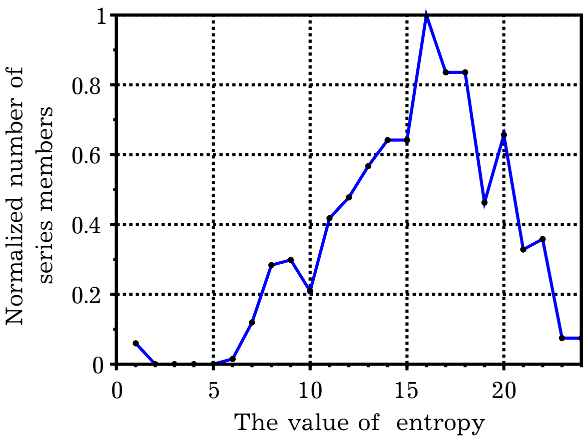

Fig. 4. Histogram of entropy values constructed for three selected nodes.

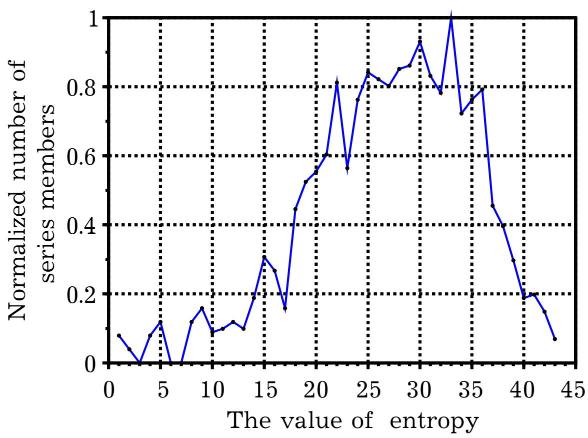

Fig. 5. Histogram of entropy values constructed for four selected nodes.

As can be seen from Figs. 4, 5, the histograms of the distribution of entropy values have heavy tails, from this it can be concluded that the entropy values have the property of self-similarity on a set of different sets of uncertain flows in the network [5].

4. Conclusions

The algorithm for determining the entropy of a network with given values of incoming and outgoing traffic in each node of the network is studied in the paper. The concept of structural entropy is used, as the basis of the most well-known definition of the entropy of a complex system and is determined on the basis of the number of the system states. The number of network states is calculated on the basis of the possible number of solutions of the linear equations system simulating a network of a given configuration with given values of the input and output flow in each node of the network.

Another achieved result of the conducted research is the application of the concept of statistical entropy, which is based on the known probabilities of the system being in one or another state. This is necessary in order to analyze the dynamic characteristics of the network, which determines its behaviour in case of deviation from the equilibrium state. The deviation of the network from the equilibrium state is modeled by introducing uncertain flows that are part of a set of system solutions, and statistical entropy is calculated based on the probability of an uncertain flow obtaining a certain value from a set of possible system solutions.

The algorithm is applied to a network of four nodes with a total value of flows in each node of 10 erlangs and a flow discretization value of 1 erlang. Thus, the conducted analysis is based on a set of the system solutions, the power of which is approximately equal to $2 \cdot 10^6$. The analysis is performed for network with 1, 2, 3, and 4 unknowns out of 16 flows that fully define the system. Calculated entropy values determined for each specific set of unknown flows from the 16 flows in the system. On the basis of calculated entropy values, entropy graphs are constructed for all possible sets of unknown flows in the system. Such data make it possible to predict the process of redistribution of information flows in the system, leading to an increase in entropy in the system.

A statistical analysis of the obtained entropy values is carried out. Based on the histograms of the entropy values distribution, the optimal number of entropy values intervals for the analysis is calculated: 4 for one unknown flow value, 11 for two, 24 for three, and 43 intervals for four unknown flow values. Statistical analysis simplifies the quantitative analysis of network entropy, while highlighting the sets of flows whose uncertainty leads to the largest increase in entropy. This enables to predict the most important areas in the network, highlighting them among the others. It also makes it possible to determine where significant changes in traffic occur, without conducting direct measurements, but only by analyzing changes in the entropy of information flows in the network.

The developed method can be used for more complex networks with more than four nodes. But at the same time, it is necessary to use discretization with a larger interval to reduce the possible number of values of information flows in networks.

The practical value of the developed method lies in the fact that for a network of a given configuration and with a given number of flows, it is possible to determine the branches of the network that lead to the greatest change in entropy, and therefore, deviations of the network from the equilibrium state. In such areas, additional traffic analyzers can be installed to monitor its values and dynamics of changes.

Conducting further research is associated with reducing the computational complexity of the problem. The studied set of the system solutions can be reduced, taking into account only the subset of solutions that is close to the equilibrium state of the system. That's why it is possible to carry out a detailed analysis of the system without increasing the step of discretization of the values of information flows in the system. Solutions that determine the equilibrium state of the system can be found by the method of maximum structural entropy.

[1] Lemeshko O., Yeremenko O. Enhanced method of fast re-routing with load balancing in software-defined networks. Journal of Electrical Engineering. **68** (6), 444–454(2017).

[2] Venugopal A., Leib H. A Tensor Based Framework for Multi-Domain Communication Systems. IEEE Open Journal of the Communications Society. **1**, 606–633(2020).

[3] Kerner B. S. Introduction to Modern Traffic Flow Theory and Control. Springer (2009).

[4] Andrushchak V., Maksymyuk T., Kaidan M., Andrukhiv T. Intelligent payload and energy balancing in optical transport network using graph neural network. 2021 IEEE 16th International Conference on the Experience of Designing and Application of CAD Systems (CADSM). 27–30 (2021).

[5] Kirichenko L., Abed S., Radivilova T. Generalized Approach to Analysis of Multifractal Properties from Short Time Series. International Journal of Advanced Computer Science and Applications. **11** (5), 183–198 (2020).

[6] Lemeshko O., Yeremenko O., Yevdokymenko M., Radivilova T., Ageyev D. Secure based traffic engineering model in softwarized networks. 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT). 143–147 (2020).

[7] Politanskyi R., Lesinsky V., Kachur V., Vistak M. Theoretical studies of the information capacity of complex signals generated by the method of their decomposition in two-dimensional space. 2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 01–06 (2022).

[8] Politanskyi R., Veryga A., Vistak M. Analyze of Scientific Problems Facing Developers of Modern Telecommunication Technologies. 2020 IEEE International Conference on Problems of Infocommunications, Science and Technology (PIC S&T). 533–536 (2020).

[9] Veryga A., Politanskyi R., Lesinskyi V., Ruda T. Analysis of Using of Fractal Signals for Noise Immune Information Transmission Systems. 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 162–165 (2020).

[10] Tanyimboth T., Templeman A. Calculation Maximum Entropy Flows in Networks. Journal of the Operational Research Society. **44** (4), 383–396 (1993).

[11] Almog A., Shmueli E. Structural entropy: monitoring correlation-based networks over time with application to financial markets. Scientific Reports. **9** (1), 10832 (2019).

[12] Lesinskyi V., Yemelyanov O., Zarytska O., Symak A., Petrushka T. Development of a toolkit for assesing and overcoming barriers to the implementation of energy saving projects. Eastern-European Journal of Enterprise Technologies. **5/3** (307), 24–38 (2020).

[13] Long C., Kehe W., Yi L. A Load Balancing Algorithm Based on Maximum Entropy Methods in Homogeneous Clusters. Entropy. **16** (11), 5677–5697 (2014).

[14] Niven R., Abel M., Schlegel M., Waldrip S. Maximum Entropy Analysis of Flow Networks: Theoretical Foundation and Applications. Entropy. **21** (8), 776 (2019).

[15] Pennini F., Plastino A., Ferri G., Plastino A. Energetic Cost of Statistical Order-Degree Change in a Fermions' Set. Entropy. **24** (6), 752 (2022).

[16] Hylton T. Thermodynamic State Machine Network. Entropy. **24** (6), 744 (2022).

[17] Politanskyi R., Zarytska O., Vistak M., Vlasenko V. Research of distribution of information flows in a network by methods of linear algebra. Mathematical Modeling and Computing. **8** (4), 821–829 (2021).

[18] Woods J. Multidimensional Signal, Image, and Video Processing and Coding. Elsevier (2011).

Розрахунок ентропії у мережах з інформаційними потоками

Політанський Р. Л.[1], Бобало Ю. Я.[2], Зарицька О. Л.[2], Киселичник М. Д.[2], Вістак М. В.[3]

[1] *Чернівецький національний університет імені Юрія Федьковича, вул. Коцюбинського, 2, 58012 Чернівці, Україна*
[2] *Національний університет "Львівська політехніка", вул. Бандери, 12, 79013 Львів, Україна*
[3] *Львівський національний медичний університет імені Данила Галицького вул. Пекарська, 69, 79010 Львів, Україна*

У роботі проводиться аналіз мережі із заданими вхідними та вихідними потоками у кожному її вузлі. Основою цього аналізу є алгоритм визначення множини розв'язків системи лінійних рівнянь, оснований на методі Гауса. Потужність множини визначає структурну ентропію системи. Шляхом введення невизначеності у значення частини інформаційних потоків моделюється відхилення мережі від її рівноважного стану. Множина потенційних розв'язків, які є частиною загальної множини розв'язків системи, визначає статистичну ентропію системи. Імовірнісна ентропія розрахована для мережі із чотирма вузлами та сумарним потоком, який становить 10 ерлангів із кроком дискретизації 1 ерланг. Обчислені значення ентропії для 1, 2, 3 та 4 невизначених потоків із загальної кількості 16 потоків, які передаються між вузлами повнозв'язної мережі. У результаті проведеного статистичного аналізу значень ентропії визначена оптимальна кількість статистичних інтервалів для значень ентропії: 4, 11, 24 і 43 інтервали для 1, 2, 3 та 4 невизначених потоків відповідно. Це дозволяє виділити множини потоків у системі, що мають найбільший влив на значення ентропії у системі. Одержані результати мають практичне значення, оскільки уможливлюють виявлення відхилень мережі від її рівноважного стану шляхом контролю за проходженням трафіку на окремих ділянках складної телекомунікаційної мережі. Оскільки, як показано у наших попередніх роботах, задача визначення повної множини розв'язків системи для кількості вузлів більше 4 має значну обчислювальну складність, то застосування алгоритму до таких мереж потребує збільшення кроку дискретизації значень інформаційних потоків у мережі. Іншим способом зменшення обчислювальної складності може бути зменшення множини аналізованих розв'язків до підмножини розв'язків, що наближені до рівноважного стану системи.

Ключові слова: *ентропія, розв'язки системи лінійних рівнянь, самоподібність, мультифрактальність.*