

#3



On the Regularity of the Bias of Throughput Estimates on Traffic Averaging

Victor A. Rusakov, National Research Nuclear University MEPhI (Moscow Engineering Physics Institute) 31 Kashirskoe shosse, 115409 Moscow, Russia. VARusakov@mephi.ru



ABSTRACT

The interaction of intelligent agents implies the existence of an environment to support it. The usual representations of this environment are graphs with certain properties. Like reliability, throughput is one of the most important characteristics of such graphs. When evaluating throughput in the analysis and synthesis of graphs, a reasonable combination of heuristic and strict approaches is used. In practice, this leads to the use of graph metrics. Usual shortest paths are widely used as part of the various multi-colour flow distribution procedures. The analytical capabilities of the Euclidean metric can achieve much more than just obtaining such a distribution. Such a metric allows us to introduce an abstract measure of the (quadratic) proximity of an arbitrary graph to a complete graph. This measure can be used as a single indicator of the reliability and throughput of the graph. Other conditions for tasks on graphs can be attributed to restrictions. A traffic matrix is one of these conditions. Non-stationarity of traffic when averaging over time can significantly reduce the accuracy of the estimates of throughput. The described dependencies of the throughput on the traffic's non-stationarity can be used in the analysis and synthesis of the communication environment when organizing the structure of the interaction of intelligent agents in the conditions of limited resources. These dependencies are verified by the results of numerical experiments.

INTRODUCTION

In the real world it takes random amounts of time for any packet to go through the network. The simplest probabilistic models are often used to get the average amount of such time estimated for the whole network. A number of supplementary assumptions and terms are incorporated into these models. Some of these additions are quite unrealistic, but in practice the simplest models give not bad results. Using this heuristic approach, the average packet delay τ during transmission over the network was obtained [1]. In practice, the selection of channel capacities must be made from a small finite set [2]. In many procedures either these capacities are assigned before the routing is performed, or a routing is first performed and then the capacities assigned [3]. Following the second approach, we will further assume that channel capacities are the same. Large amounts of accurate data are very rare [4, 5]. The situation is the same for the network's draft stage [6]. At this stage throughput estimates are usually obtained under the condition that equal opportunities are provided to all pairs of nodes [2, 7]. That's why the matrix of the so-called uniform traffic is often used. Such a matrix has all equal entries except the zero ones on the main diagonal. These entries are time-average intensities of the input flows. Such entries are equal to 1 in a normalized traffic matrix. There is a huge number of graphs even for modest number k of vertices and h of edges. So the problem of finding the distribution of a multi-coloured flow has to be solved many times when synthesizing networks that have the best throughput index. Using an averaged traffic matrix significantly reduces the complexity of obtaining estimates of graph throughput. However, this approach leaves open the question of the accuracy of such estimates. The importance is emphasized in [8, 9] of the robustness of the design solutions with respect to the low certainty and non-stationarity of traffic. An ARPANet insensitivity property was established that justifies the use of traffic averages for network design. A supposition was made that this property is typical for distributed networks [2, 10]. Numerical simulation of non-stationary traffic is described in [1] using ARPANet examples. A slight decrease (10-13%) in the estimate of the throughput for non-stationary traffic is noted compared to the case of an averaged traffic matrix. There were no justifications or generalizations of such behavior of these estimates.

METHOD

Let X and Y be vector spaces with dimensions h and k correspondingly over the field of real numbers. The arbitrarily introduced orientations of the edges in the absence of limits to their capacity turns the incidence matrix A of an undirected graph having ± 1 as non-zero elements in each column into a useful instrument for the linear transformation Axy of the edge flows xy into the vertex flows ye . From now on the incidence matrix A will be understood to be such a matrix. Let $\{x \in X | Ax = y\}$ where $y = y(t)$ has an entry equal to 1- k , and the remaining entries are equal to 0. Any of the vectors $x = x(t)$ represent the transfer of a flow with a magnitude of 1 and of the colour t from $s_1, \dots, (k-1)$, to k along the edges of the graph. Then, without additional designations one should search for the extreme norm along all such x . Let C designate the incidence matrix of a complete graph, NH and $R(H)$ – the H kernel and the H image for any matrix $H = \pm 1$ and $\pm =$ a transposition and the Moore-Penrose pseudo inverse, (u, w) – the scalar (dot) product of u and w . The subscript indices $\sim C$ and $\sim A$ together with $x(t)$ are used to indicate the following: $x(t) \sim C$ and $x(t) \sim A$. Let any $y \in R(A)$. Clearly, $y \in R(C)$. Lemma 1 [11, 14, 16]. $(C^*y_0, A^*y_0 - C^*y_0) = 0$. Proof. [11, 14, 16]. Fig. 1 illustrates Lemma 1. The vectors

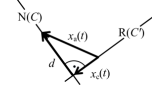


Fig.1. The spaces and vectors of Lemma 1 for $y_0 = y(t)$

$d = A^*y_0 - C^*y_0$ and C^*y_0 for $y_0 = y(t)$ are the legs of a right-angled triangle with the hypotenuse A^*y_0 for the same y_0 . Obviously, for a given k and $y_0 = y(t)$, $t=1, \dots, k$, the vector $x_0(t) = C^*y(t)$ is constant. Then, according to the Pythagorean theorem, the Euclidean lengths of vectors d and $x_0(t) = A^*y(t)$ are minimal simultaneously for an arbitrary graph with the incidence matrix A . According to a well-known [17, 18] property of the Moore-Penrose pseudo inverse, the Euclidean norm of the vector $x_0(t) = A^*y(t)$ will be minimal among $\{x(t) \in X | Ax(t) = y(t)\}$.

This means that the distribution along the edges of the flow $x_0(t)$ of any colour t in an arbitrary graph is the best quadratic approximation to the similar flow $x_0(t)$ in the complete graph [11, 14].

ENTER FINDING #1

Under conditions of significant error there is no point in achieving a neat passage of the curve almost through all the points obtained, in fact going over to interpolation. For another set of measurement results, the curve will have to change significantly. But if the power of the approximation is relatively small, then its description and properties will practically not change. And this will continue, despite significant changes occurring again and again in other sets of measurement results.

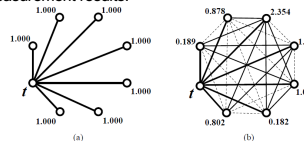


Fig.2. The flow of the colour t in the complete graph ($k=8$) for (a) uniform averaged and (b) random traffic

ENTER FINDING #2

Each of the three graphs has the same number of vertices and edges, and all edges have the same capacity. The first graph is a star with 6 edges added in a random way, Fig.3. The second graph is the result of a synthesis having throughput as the object function and taking into account traffic non-stationarity. And the third one is a given, Fig.4. It's a well-known version of one of the stages of the ARPA network.

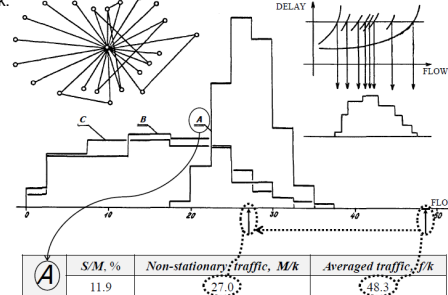


Fig.3. The scheme of experiments, graph and results of its tests

Three types of random traffic A, B and C were generated based on uniformly distributed pseudo-random numbers. The values of throughput were calculated, and their simple statistical processing was carried out.

ENTER FINDING #3

The results for all three graphs are combined on the flow axis in Fig. 4. The arrows of the first graph from Fig.3 are moved to this axis for a clear comparison. They are ringed with the same dashed lines as in Fig. 3. The figures clearly demonstrate the fulfillment of both rules (Conclusions 1 and 2). Note there is an almost twofold decrease in the throughput for random traffic of A type compared to the use of time-averaged traffic for the graph in Fig. 3. Random traffic of types B and C gives an even greater decrease.

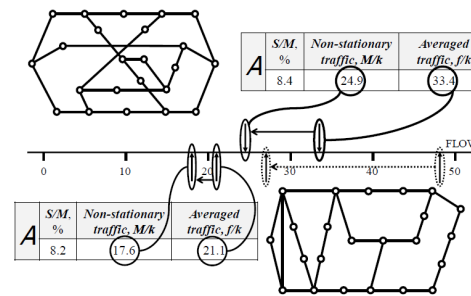


Fig.4. Graphs and their test results

CONCLUSIONS

- Let us take a graph with greater throughput (with a better flow distribution) for the time-averaged traffic matrix. Then we get greater changes in this index (in this distribution) with random traffic.
- For any graph throughput is *no less* for a time-averaged traffic matrix than the averaged throughput for random traffic.
- The reckless desire to achieve greater throughput when using averaged traffic leads to a significant overrating of the graph throughput since the real traffic is non-stationary.

REFERENCES

- Kleinrock, L.: Analytic and simulation methods in computer network design. Proc. AFIPS Conf., SJCC, v.36, pp. 569-579 (1970).
- Frank, H., Kahn, R.E. and Kleinrock, L.: Computer communication network design – experience with theory and practice. AFIPS Conf. Proc. SJCC, pp. 255-270 (1972).
- Frank, H. and Chou, W.: Topological Optimization of Computer Networks. Proc. IEEE, 60, 11, pp. 1385-1397 (1972).
- Hamming, R. W.: Numerical Methods for Scientists and Engineers. McGraw-Hill, N.Y. (1962).
- Voevodin, V.V.: Computational foundations of linear algebra. Nauka, Moscow (1977).
- Rusakov, V.A.: Synthesis of computer network structures and the problem of small certainty of initial values. USSR AS's Scientific Council on Cybernetics. Proc. 5th All-Union School-Seminar on Computing Networks, 1, VINITI, Moscow-Vladivostok, pp. 112-116 (1980).
- Frank, H., Frisch, I.T. and Chou, W.: Topological considerations in the design of the ARPA computer network. AFIPS Conf. Proc. SJCC, pp. 581-587 (1970).
- Van Slyke, R., Frank, H. and Chou, W.: Avoiding simulation in simulating computer communication networks. Proc. AFIPS Conf., June 4-8, NOCES, pp. 165-169 (1973).
- Frank, H. and Chou, W.: Network Properties of the ARPA Computer Network. Networks, 4, pp. 213-239 (1974).
- Gerla, M., Kleinrock, L.: On the Topological Design of Distributed Computer Networks. IEEE Trans. on Comm., COM-25, No.1, pp. 48-60 (1977).
- Rusakov, V.A.: Using Metrics in the Throughput Analysis and Synthesis of Undirected Graphs. Antipova T. (Ed.), ICIS 2019: Lecture Notes in Networks and Systems, vol. 136. Cham, Switzerland: Springer, pp. 277-287 (2021). DOI: 10.1007/978-3-030-49264-9_25
- Hu, T.C.: Integer Programming and Network Flows. Addison-Wesley, Menlo P.-London (1970).
- Chou, W. and Frank, H.: Routing Strategies for Computer Network Design. Proc. Symp. on computer communications networks and teletraffic, Polytech. Inst. of Brooklyn, April 4-6, pp. 301-309 (1972).
- Rusakov, V.A.: Analysis and Synthesis of Computer Network Structures. Part 1. Analysis. Moscow Engineering Phys. Inst. Report: All-Union Sci. Tech. Inform. Center No 5796153 (1979).
- Rusakov, V.A.: On Markov Chains and Some Matrices and Metrics for Undirected Graphs. Antipova T. (Ed.), ICIS 2019: Lecture Notes in Networks and Systems, vol. 136. Cham, Switzerland: Springer, pp. 340-348 (2020). DOI: 10.1007/978-3-030-22493-6_30
- Rusakov, V.A.: Using Metrics in the Analysis and Synthesis of Reliable Graphs. Samsonovich, A.V. (Ed.), BICA 2019: Advances in Intelligent Systems and Computing, vol. 948. Cham, Switzerland: Springer, pp. 438-448 (2020). DOI: 10.1007/978-3-030-25719-4_58
- Albert, Arthur E.: Regression and the Moore-Penrose pseudoinverse. Academic Press, New York (1972).
- Beklemishev, D.V.: Additional chapters of linear algebra. Nauka, Moscow (1983).
- Horn, R., Johnson, C.: Matrix Analysis. 2nd edn. Cambridge University Press, New York (2013).
- Voevodin, V.V.: Linear Algebra. Nauka, Moscow (1980).
- Gantmakher, F.R.: Matrix Theory, 3rd edn. Nauka, Moscow (1967).
- Rusakov, V.A.: Matrices, shortest paths, minimal cuts and Euclidean metric for undirected graphs. Procedia Computer Science, 145, pp. 444-447 (2018). DOI: 10.1016/j.procs.2018.11.104
- Kemeny, J. and Snell, J.: Finite Markov Chains. University series in undergraduate mathematics. Van Nostrand, Princeton NJ (1960).
- Rusakov, V.A.: The impact of traffic changes in the tasks of graphs optimization of the backbone data communication network. USSR AS's Scientific Council on Cybernetics. Proc. 5th All-Union School-Seminar on Computing Networks, 3, VINITI, Moscow-Pushchino, pp. 90-95 (1984).
- Rusakov, V.A.: On the regularity of the displacement of the mean estimate for the throughput with non-stationary traffic. USSR AS's Scientific Council on Cybernetics. Proc. 5th All-Union School-Seminar on Computing Networks, 1, VINITI, Moscow-Pushchino, pp. 48-52 (1984).
- Gelfand, I.M.: Lectures on Linear Algebra. 4th edn. Nauka, Moscow (1971).
- Rusakov, V.A.: On the Moore-Penrose Pseudo Inverse of the Incidence Matrix for Weighted Undirected Graph. Procedia Computer Science, 169, pp. 147-151 (2020). DOI: 10.1016/j.procs.2020.02.126
- Rusakov, V.A.: A technique for analyzing and synthesizing the structures of computer networks using Markov chains. Computer networks and data transmission systems. Znaniye, Moscow, pp. 62-68 (1977).
- Rusakov, V.A.: Non-stationary flows in computer networks: a linear basis for their study. USSR AS's Scientific Council on Cybernetics. Proc. 7th All-Union School-Seminar on Computing Networks, 1, VINITI, Moscow-Breun, pp. 110-124 (1980).