UDC 004.9 **doi: 10.26906/SUNZ.2022.4.117**

Alevtyna Filonenko, Heorhii Molchanov, Oleksandra Bellorin-Herrera

National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine

THE METHOD OF CALCULATING THE CAPACITY OF THE CLOUD COMPONENT OF THE DISTRIBUTED MULTISERVICE NETWORK

Abstract. A method for calculating the bandwidth of the cloud component of a distributed multiservice network is proposed. Which takes into account the probabilistic characteristics of the link of the first and second orders. At the preliminary stages, the probability of packet loss within the network link is calculated. An analysis of the queues of switching nodes of communication with the cloud is also carried out. Consistent application of the method will allow to estimate the probability of losses for different network users. It will also make it possible to determine the rational loading of network links with the aim of optimal distribution of network resources. The obtained results can be applied directly in the design of a distributed multiservice network. In it, user access to the link resource can be either unlimited or limited with the introduction of link resource reservation for priority classes of users. And also for the design of a distributed multiservice network in which users are provided with fixed bit rates of information transmission. The direction of further research is the extension of the method for a distributed multiservice network, in which some of the links are dependent. It is planned to develop an algorithm in which calculations for the cloud component of the network can be performed in parallel.

Keywords: the distributed multiservice network, cloud component, cloud communication switching node, statistical multiplexer, bandwidth, probabilistic characteristics, network link, network resources.

Introduction

Statement of the problem and analysis of the literature. The problems of integrating various types of communication, including information, system and network aspects, have been the most relevant in the field of telecommunications for a number of years [1]. Today, there is an expansion of user needs in new types of communication with high speeds of information transfer. This requires significantly higher throughput values. Such requirements have led to the fact that at present many network operators have begun the transition to distributed multiservice networks that implement the second level of integration and users of distributed types of communication: audio, video, data transmission, multimedia, etc. To date, intensive research is being carried out in the field of the theory of construction and operation of a distributed multiservice network [2, 3]. Among the complex of problems solved in this area, one of the main ones is the problem of optimal, from the point of view of distribution of network resources, building a switching node.

The integration of various types of communication in an extensive multiservice network in the system aspect is based on ATM technology as a secure method of information transmission regardless of its source and uses the principle of asynchronous time multiplexing [4, 5]. All kinds of information are displayed in standard forms of fixed length packets (ATM cells). These packets in an asynchronous mode, by the method of statistical multiplexing, bequeath the transmission medium. In a divided multiservice network, packet switching is networked over virtual circuits. Information packets arriving via statistically compacted input lines for take-offs with a cloudy component should be redistributed on similar output lines. This operation is carried out in the mode of fast switching of packets using various means of switching: multilink switching systems, a common bus or memory, switching tori, etc. However, the fundamental problem in creating a distributed multiservice network is the problem of ensuring the rational use of its resources and the required quality of user service [6-11]. The solution of this problem requires, first of all, the development of a method for calculating the throughput of a distributed multiservice network and assessing the quality of user service. Given the structural complexity of a distributed multiservice network, it is advisable to first solve this problem with respect to the cloud component of a distributed multiservice network. With this method at hand, the throughput of the entire network can be calculated.

The purpose of the article is to develop a method for calculating the throughput of an individual link in a distributed multiservice network. Such a network takes into account the probabilistic characteristics of the link of the first and second orders, multiservice factors and traffic structure, as well as the effect of link resource reservation as the most promising method for managing the distribution of resources of a distributed multiservice network.

1. Loss probability calculation packets within a distributed multiservice network link

Depending on the selected means of switching of a distributed multi-service network, when using fast switching of packets (there are situations when a packet cannot be immediately transmitted to the outgoing line due to the current transmission of another packet on this line and must wait for the appropriate moment of transmission in the ferry, which is designed to store L packets. If the buffer overflows, then packet loss is possible, i.e. packets are serviced at the fast packet switching node by a lossy and waiting mass service system.

Consider the problem of estimating the probability of packet loss in a multistage fast packet switching system by defining its basic switching element as an element of n inputs with capacity buffers L_i and m outputs, which are inputs to the switching system of fast packet switching. The structure of the fast packet switching system can be different (Fig. 1), while if the system has S cascades, the virtual transmission channel will contain S buffers with $L_{\Sigma}^{(S)} = \sum_{i=1}^{n} L_i$ waiting places.

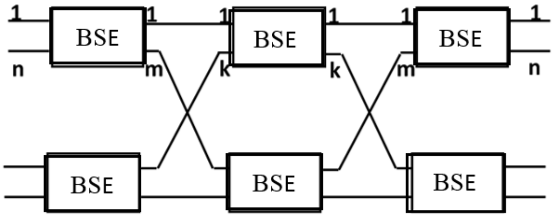

Fig. 1. Three-stage fast packet switching system

Packet losses in this model occur when any buffer overflows. Since packet losses are possible when the buffer is connected through the CS to one of the lines flowing from the basic switching element, the estimation of the probability of losses in the basic switching element can be reduced to the calculation of the probability of losses in a single-line mass service system with a buffer of finite capacity. Since all packets have the same length and the law of the arrival of packets in the mass service system is arbitrary, using Kendall's notation [5], the basic switching element of a link of a distributed multiservice network can be considered a mass service system of the form GI/D/1/$L_{\Sigma}^{(S)}$.

Let the incoming load be given by the load intensity λ and the dispersion coefficient k_D ($k_D = \sigma^2 / \lambda, \sigma^2$ – dispersion). Approximate this system with a single-channel system M/D/1/L_A with Poisson load and L_A waiting places, the probability of losses in which is defined as [2]:

$$P_A = 1 - \left(\lambda + P_0\right)^{-1}, \tag{1}$$

where the stationary probability of the absence of processed packets P_0 in the link is determined using the system of Kolmogorov equations [1]:

$$P_j = \left(k_0\right)^{-1}\left(P_j - k_j P_0 - \sum_{i=1}^{j}\left(k_{j+i-1}P_i\right)\right);\qquad k_j = \left(j!\right)^{-1}\lambda^j e^{-\lambda}. \tag{2}$$

Consider a Poisson unilinear system M/D/1/L_A. Let's select such a number of waiting places in it L_A, at which the system at the load intensity λ will have the same probability of loss as the system M/D/1/$L_{\Sigma}^{(S)}$. At the same time, the probability of losses in this system is equal [2]:

$$P_M = \lambda^{L_A}\left(P_{\lambda,1}^{-1} + \lambda\left(1 - \lambda^{L_A}\right)\left(1 - \lambda\right)^{-1}\right), \tag{3}$$

where $P_{\lambda,1}$ – the probability of loss, which is determined by the first Erlang formula at the capacity of the bundle of channels V = 1 and loads λ.

Accepted $P_M = P_A$ and using expressions (1) and (3), possible from the system M/M/1/L_A go to system M/D/1/L_A, which at the same intensity will have the same losses.

System M/M/1/L_A can be used to approximate the probability of losses in the system GI/M/1/L_A by using k_D - Witt's approximations [8], according to some imaginary behavior of the N-linear system GI/M/N/L_A at large loads can be roughly described by the Poisson system, which receives the load $\lambda_{eq} = \lambda / \omega^2$, servised $V_{eq} = V / \omega^2$ devices and has $L_{eq} = L_{\Sigma}^{(S)} / \omega^2$ waiting places, where:

$$\omega = \frac{k_D - 1 + C}{C}; \quad C = \frac{\lambda k_D - U}{\lambda k_D + U}, \tag{4}$$

and the values of the parameter U, which can be determined by the linear regression method, are in the range from 0,2 to 0,9 Erl. After transformations of expressions (1)–(4) we get

$$P_A = \left(\lambda_{eq}\right)^{L_A/\omega} \times \times\left(P_{\left(\lambda_{eq},V_{eq}\right)}^{-1} + \left(\lambda_{eq}\right)^{L_A/\omega+1}\left(1 - \lambda_{eq}\right)^{-1}\right), \tag{5}$$

where the probability of losses on the beam $V_{eq} = V / \omega^2$ $P_{\left(\lambda_{eq},V_{eq}\right)} = P\left(\lambda_{eq}, V_{eq}\right)$ calculate using the integral representation of Erlang's formula [8]:

$$P\left(\Lambda, V\right) = \Lambda^V\left(e^{\Lambda}\Gamma\left(V+1,\Lambda\right)\right)^{-1}, \tag{6}$$

Γ[V +1, Λ) – incomplete gamma function equal to the value of the definite integral

$$\Gamma\left(V+1,\Lambda\right) = \int_{\Lambda}^{\infty} e^{-t}t^{\Lambda}dt,$$

after representing which as a continued fraction and substituting the result into expression (6), we obtain the following recurrence relation for calculating the value of the beam loss probability:

$$P_i\left(\lambda_{eq}, V_{eq}\right) = \frac{\lambda_{eq} - V_{eq}/\left(1+\theta_i\right)}{\lambda_{eq}},$$

where $\theta_n = 0;\quad \theta_{i-1} = i\left(\lambda_{eq} + \left(i - V_{eq}\right)\left(1+\theta_i\right)^{-1}\right)^{-1}.$

The probability of losses defined in (3) can be used to approximate the probability of losses in the system GI/D/1/$L_{\Sigma}^{(S)}$, with load (λ, k_D), that comes.

Analysis of expressions (3) – (6) allows us to conclude that the probability of packet loss for a fast packet switching system significantly depends on the incoming load dispersion coefficient, especially with small losses. Even a relatively small increase in the dispersion coefficient k_D leads to a sharp increase in packet loss in the system (Fig. 2), and failure to take into account the second load moment when designing fast packet switching nodes can lead to a sharp deterioration in the quality of service in conditions where the incoming load is even slightly different from Poisson.

2. Analysis of queues of switching nodes of communication with the cloud of a link of a distributed multi-service network

A cloud communication node is a managed buffer that receives packets from packet generators. The application of statistical multiplexing is possible only when switching to an asynchronous method of information transmission, which allows you to take into account the statistical features of digital streams created by individual classes of users, and to carry out statistical compression of the transmission path. Organization of user access to a common communication channel through a multiplexer leads to the need to solve the following problems:

- building a classification of users according to the structure of the load created by them;
- evaluation of the transmission speed provided to a given group of users of different classes, necessary to guarantee the standardized quality of service, which is expressed in terms of the probability of packet loss and their delay;
- creation of a user load management procedure, which includes the following two parts:
  - individual management – monitoring the load of an individual user and checking its compliance with the declared class, as well as smoothing this load so that changing its parameters does not have an excessive impact on packet losses of other users;
  - integrated management - connection management, i.e. the procedure for making a decision on whether another connection can be established and whether it will not lead to an increase in packet losses and/or delays beyond the norm.

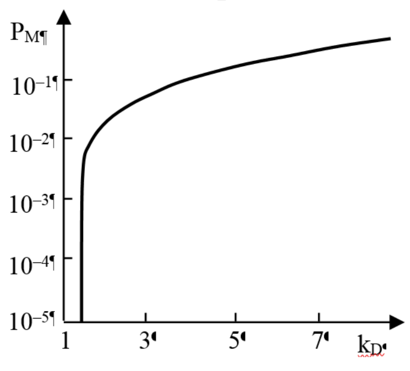

Fig. 2. The impact dispersion coefficient analysis of on the packet loss probability

If T – average active period of the source, Λ_p – intensity of packet generation in the active station S – average length of pause between periods of activity, D – the duration of deterministic service in a fictitious mass service system, then the burst rate (the ratio of the maximum intensity of the arrival of packets to the average) is calculated as

$$B = \left(T + \Lambda_p\right) / \Lambda_p .$$

With sufficient storage capacity, the average intensity of the flow of packets that have already been served by the mass service system can be taken as equal to the intensity of their arrival, i.e. equal to

$$\lambda\left(T + \Lambda_p\right) / \Lambda_p .$$

On the other hand, during busy periods, packets leave the system with an intensity of 1/D. By marking the relevant packet flow characteristics as $T_{L_\Sigma B}, \Lambda_{L_\Sigma B}, S_{L_\Sigma B}, B_{L_\Sigma B}$ and additionally accepting the condition that during periods of activity the source generates packets after a deterministic time, we obtain:

$$T_{L_\Sigma B} = T\lambda D;\ \Lambda_{L_\Sigma B} = 1/D;$$

$$S_{L_\Sigma B} = S + T(1-\lambda D),\ B_{L_\Sigma B} = B/\lambda D,$$

and the average waiting time is

$$W_{L_\Sigma B} = T(\lambda T - 1)2 .$$

From this it can be seen that the pulsating flow is smoothed out if the service rate is less than the packet generation rate in the user's active state.

If $\Pi(z) = a_0 + a_1 z + \ldots + a_n z^n$ (n – the maximum number of established connections) – the generating function (GF) of the number of packets arriving per clock (the time of sending one packet to the channel) is chosen so that at least the first moments of the total number of packets generated by all sources during the time period characteristic of the period are equal -dual changes in traffic intensity.

For the average time interval between the beginnings of two periods of activity, and if the storage for cells has unlimited capacity, then a fairly accurate estimate from above of the average waiting time W is equal to

$$W = \max_{z\in[0,1]} \left(\Pi(z)/\left(2\rho(1-\rho)\right)\right); \quad \rho = \Pi'(1);$$

GF distribution of packet waiting time is calculated as

$$W(z) = (1-\rho)\frac{\Pi(z)-1}{\rho\left(z-\Pi(z)\right)},$$

and GFdistribution of the queue length in the multiplexer has the following form:

$$Q(z) = (1-\rho)\frac{\Pi(z)z(-1)}{z-\Pi(z)} .$$

Substitute the Π(z) in the expression (4), it is possible to obtain a recursive procedure for calculating the queue length distribution ($q_i = P\{\zeta = i\}$, ζ – a random value characterizing the length of the queue k ≥ 1):

$$Q_0(z) = Q_0(z);\ q_0 = Q_0(0);$$

$$Q_k(z) = \left(Q_{k-1}(z) - q_{k-1}\right)/z;\ q_k = Q_k(0),$$

type of distribution of the number of packets in the queue with limited storage capacity L

$$q_k^{(L)} = q_k \left(\sum_{i=0}^{L} q_i \right)^{-1}, \quad q \in \overline{0, L}$$

and the probability of packet loss is equal to

$$P_{CM} = 1 - \left(1 - q_0^{(L)}\right) \Big/ \rho ,$$

and on the value P_{CM} are strongly influenced by the correlation properties of the incoming flow, and the best way to estimate the effect of the correlation of the flow on the probability value is to choose when defining Π(z) this duration interval as characteristic.

3. Bandwidth estimation links of a distributed multiservice network

Studies have shown that the bandwidth of a distributed multiservice network depends on many factors, the main of which are:

- number of user classes (number of load sources);
- the size of the bandwidth of the transmission bit rate, necessary to serve the requests of different classes of users;
- the nature of the change in the transmission bit rate band over time (load sources with CBR (constant bite rate) or VBR (variable bite rate));
- an adopted application access control procedure (in ATMs, this is Call Admission Control).

Only taking into account the entire set of factors allows you to estimate the network throughput, in particular, the probability of packet loss for certain classes of users, that is, to build a vector of packet losses. Due to the structural complexity of a distributed multi-service network, it is expedient to solve the problem of calculating the probability of losses first on one link. Analyzing a link of a distributed multiservice network, it is assumed that each class of users creates an incoming load, having M_i – the number of transmission bit rate bands required to serve users of the class K_i. The method of calculating the bandwidth of a link of a distributed multiservice network consists of two stages. At the first stage, all traffic sources are replaced VBR on sources of equivalent traffic CBR. Equivalence is understood relative to the preservation of value P_A – the probability of package losses. The replacement of sources is reduced to the recalculation of the bandwidth of the transmission bit rate. Equivalent transmission bit rate bandwidth for the i-th class of users with VBR traffic at a given probability P_A [6] defined as

$$k(P_A) = \chi(P_A) M[\varsigma] + \eta(P_A) D[\varsigma] / C_L \approx$$
$$\approx \chi(P_A) M[\varsigma] + \eta(P_A) M[\varsigma]\left(h - M[\varsigma]\right) / C_L ,$$

where C_L – transmission speed on the link; h – the maximum bit rate bandwidth value of the i-th user class for the normalized bit intensity of the load generated by this user class; $M[\varsigma], D[\varsigma]$ – respectively, the mathematical expectation and dispersion of a random variable characterizing the moments of the probability distribution of the bit rate bandwidth over time; $\chi(P_A), \eta(P_A)$ – coefficients depending on the probability of packet loss. The values are determined for different classes of users experimentally, while the mathematical expectation is determined not directly, but through burstiness B, which is the most important characteristic of the transmitted information, which significantly affects the bandwidth of the link and the entire network, case B = 1 corresponds to a constant rate of information transfer). In a link of a distributed multiservice network, the information transfer rate is a random process r(t). Due to physical reasons, there is always a limit to the maximum allowable transfer rate:

$$r_{\max} = \max_{t \in [0,T]} r(t) ,$$

where T – given time interval in which the average speed of information transfer is equal to

$$r_{cp} = \frac{1}{T} \int_0^T r(t) dt = r_{\max} / B .$$

The second stage of the method includes the calculation of the probabilistic characteristics of the link, taking into account the equivalent replacement of the transmission bit rate bandwidth performed at the first stage. According to the statement of the problem, the method is developed for two strategies of link resource management - in the absence of bandwidth reservation of the transmission bit rate and in its presence.

Let's consider strategy 1, when the access of users to the link resource is not limited and there is no bandwidth reservation of the transmission bit rate. The probability distribution of the number of simultaneously occupied transmission bit rate bands on the link has the form [7]

$$f\left(x \middle| x \in [0,V]; x \in \mathbb{Z}\right) = \frac{\frac{1}{x!} \prod_{j=1}^{x} \mu(j-1) \sum_{j=1}^{U} A_i}{\sum_{k=0}^{V} \frac{1}{k!} \prod_{j=1}^{k} \mu(j-1) \sum_{j=1}^{U} A_i} . \quad (8)$$

If all link users have unlimited access to the link resource, then the link is a fully accessible bundle and, therefore, $\mu(j) = 1 \quad \forall j \in [0,V]; j \in \mathbb{Z}$, that is, the probability of packet loss on the link for users of the i-th class can be calculated as

$$P_i = \sum_{x=V-M_i+1}^{V} f(x), \quad i \in \overline{1,U} . \quad (9)$$

The second strategy involves the introduction of reservation of link resources for some classes of users. The probability distribution of the number of simultaneously occupied bit rate bands on a link of a distributed multiservice network has the following form ($x \in \mathbb{Z}$, $x \in [0,V]$) [9]:

$$f_1(x) = \frac{\frac{1}{x!} \prod_{j=1}^{x} \mu(j-1) \sum_{j=1}^{U} A_i (T_i - x)}{\sum_{k=0}^{V} \frac{1}{k!} \prod_{j=1}^{k} \mu(j-1) \sum_{j=1}^{U} A_i \tau(C)(T_i - k)} , \quad (10)$$

where coefficient $\tau(C)$ takes on non-zero values ($\tau(C)$=1) with positive values C. Similarly, (9) the probability of packet loss on links for users of the i-th class is calculated on the basis of distribution (10):

$$\sum_{x=V-M_i+1}^{V} f_1(x), \quad i \in \overline{1,U} \,. \tag{11}$$

The proposed method allows constructing an algorithm for calculating the bandwidth of a distributed multiservice network link, taking into account the probabilistic characteristics of the first and second orders. Consistently applying this algorithm for implementing the calculation of throughput for individual network links, it is possible to estimate the probability of losses for different classes of network users from "point to point", as well as determine the throughput and allowable load of the links of a distributed multiservice network, that is, rationally distribute network resources.

Conclusions

The article considers a two-stage method of calculating the bandwidth of a separate link of a distributed multiservice network. At the previous stages, the probability of packet loss within the network link and the analysis of the queues of switching nodes communicating with the cloud are calculated. At the first stage, VBR traffic sources are replaced by similar CBR traffic sources, and the probabilistic characteristics of the link of the first and second orders are taken into account, the second stage is the final one. Consistent application of the method will make it possible to estimate the probability of losses for different network users, to determine the rational loading of network links with the aim of optimal distribution of network resources.

The obtained results can be applied directly in the design of a distributed multi-service network, in which the access of users to the link resource can be either unlimited or limited with the introduction of reservation of link resources for priority classes of users, as well as for the design of a distributed multi-service network , in which users are provided with fixed bit rates of information transmission. The direction of further research is the extension of the method for a distributed multi-service network, in which part of the links are dependent, and the development of an algorithm in which calculations for the cloud component of the network can be performed in parallel.

REFERENCES

1. Sviridov, A., Kovalenko, A. and Kuchuk, H. (2018), "The pass-through capacity redevelopment method of net critical section based on improvement ON/OFF models of traffic", *Advanced Information Systems*, Vol. 2, No. 2, pp. 139–144, DOI: https://doi.org/10.20998/2522-9052.2018.2.24
2. D. Reese Cloud computing [Text] / George Reese. - SPb .: 2011. - 288 p.
3. Google Cloud Platform [Electronic resource]. – access mode: http://cloud.google.com. – 12.04.2013.
4. Ruban, I., Kuchuk, H. and Kovalenko A. (2017), "Redistribution of base stations load in mobile communication networks", *Innovative technologies and scientific solutions for industries*, No 1 (1), pp. 75–81.
5. A. N. Singh, P. Shiva. Challenges and opportunities of resource allocation in cloud computing: A survey, 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015
6. N. Latha, S.T. Deepa, Cost Optimization in Cloud Services, International Journal of Computer Appl. Vol. 106, No.5, 2014
7. M. Malawskia, K. Figielab, J. Nabrzyskia, Cost Minimization for Computational Applications on Hybrid Cloud Infrastructures, Journal on Future Generation Computer Systems, Vol. 29 Issue 7, 2013
8. Q. Jia, Z. Shen, W. Song, etc. Supercloud: Opportunities and Challenges, ACM SIGOPS Operating Systems Review - Special Issue on Repeatability and Sharing of Experimental Artifacts, Vol. 49 Issue 1, 2015
9. G. Menaga, S. Subasree. Development of Optimized Resource Provisioning On-Demand Security Architecture for Secured Storage Services in Cloud Computing. *IJESIT*, Vol. 2, Issue 3, 2013
10. Donets, V., Kuchuk, N. and Shmatkov, S. (2018), "Development of software of e-learning information system synthesis modeling process", *Advanced Information Systems*, Vol. 2, No 2, pp. 117–121, DOI: https://doi.org/10.20998/2522-9052.2018.2.20.
11. A. Simak Transaction processing / A. Simak // DBMS. - 1997. - № 2. - p. 70 - 82.

Received (Надійшла) 12.09.2022
Accepted for publication (Прийнята до друку) 16.11.2022

Метод розрахунку пропускної здатності хмарної компоненти розподіленої мультисервісної мережі

А. М. Філоненко, Г. І. Молчанов, О. М. Бельорін-Еррера

Анотація. Пропонується метод розрахунку пропускної здатності хмарної компоненти розподіленої мультисервісної мережі, враховуючий імовірнісні характеристики ланок хмари першого та другого порядків. На попередніх етапах проводиться розрахунок ймовірності втрати пакетів у межах окремої ланки хмари та аналіз черг на комутаційних вузлах зв'язку із хмарою. Послідовне застосування методу дозволяє провести оцінку ймовірності втрат для різних класів користувачів мережі, визначити раціональне завантаження ланок хмари з метою оптимального розподілу хмарних ресурсів. Отримані результати можна застосувати безпосередньо при проектуванні хмарної компоненти мультисервісної розподіленої мережі, у якій доступ користувачів до хмарного ресурсу може бути або необмеженим, або обмеженим із введенням резервування ресурсів для пріоритетних класів користувачів, а також для проектування хмарної компоненти мультисервісної розподіленої мережі, у якій користувачам для обміну із хмарою надаються фіксовані бітові швидкості передачі інформації. Напрямок подальших досліджень – розширення методу для мереж, у яких частина ланок хмарних компонент є залежними, та розробка алгоритму, у якому розрахунки для окремих ланок хмари можуть виконуватися паралельно.

Ключові слова: розподілена мультисервісна мережа, хмарний компонент, комутаційний вузол зв'язку із хмарою, пропускна здатність, імовірнісні характеристики, ланка мережі, мережеві ресурси.