

The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters

Mariia Skulysh

National Technical University of Ukraine
«Ihor Sikorsky Kyiv Polytechnic Institute»
Kyiv, Ukraine
mskulysh@gmail.com

Abstract — a method of optimizing the use of server resources that can reduce the energy costs of a network while preserving service quality parameters is proposed in this paper. This method involves the use of resource involvement schedule based on long-term statistics. Unlike existing methods, proposed method simultaneously considers the losses of requests, expected delay and energy efficiency of processing applications in telecommunication nodes. The proposed method will allow planning a queuing system for telecommunication systems, both at the access level, in the organization of SDN networks, and in planning the work of data centers.

Keywords — data network, network resources optimization, energy efficiency of telecommunications network, quality of service, queuing system.

I. INTRODUCTION

To provide the requirements of today's users in global communications to mobile operators raises the problem of providing access to high performance online media services (digital telephony and video telephony, online video surveillance, high-definition digital TV, etc.). All this leads to higher performance requirements of telecommunications equipment. As a result, there are problems associated with the continuous increase in energy resources use by mobile operators. Therefore, problems of energy efficiency and conservation in the telecommunications environment are very relevant today.

Today there are various methods of optimizing the use of network resources [1 – 3], but none of them takes into account both the parameters of service quality and energy efficiency of the network. The method of resources involvement scheduling based on the long-term statistics simultaneously considers the losses requests, expected delay and energy efficiency of processing applications in telecommunication nodes. This method allows improving the efficiency of network resources usage, while ensuring the required quality of service.

II. THE MAIN PROBLEMS OF SERVICE EFFICIENCY IN TELECOMMUNICATIONS NETWORKS

Today there is integration of telecommunication systems (TCS) and distributed computing environments (CE). As the

result, there is a single heterogeneous environment for telecommunication services maintenance, which has the ability to control information flow maintenance at each stage and to ensure compliance with high quality standards. At the same time, there is still no single concept of interactive models and methods that would take into account the peculiarities of the computing environments servicing a large number of queries generated during the provision of telecommunications services. This leads to inefficient use of resources that provide heterogeneous environment for maintenance of telecommunication services.

Here is a list of problems that hinder efficient maintenance of telecommunication services in TCS, which do not take into account the peculiarities of the computing environment:

- The need for methods and approaches for interaction between TCS and CE that classify the computing resources boot process;
- The need for conceptual solutions and unified tools for energy control system of TCS and OS;
- Lack of flexible methods of telecommunications services management that take advantage of heterogeneous telecom environment, and take into account the energy costs of maintaining system performance.

Introduction of SDN technologies requires the introduction of new models of queuing theory to estimate the parameters of the system and timely implementation of computing operations to meet the needs of TCS.

A. Application of SDN technology in TCS

Leading scientists of the world work to resolve the problem of network functions implementation in a virtualized computing environment, while the main problems is the development of algorithms and separation of functions for the organization of telecommunications and computing environment as a single whole mechanism.

The problem of calculating the capacity to service the growing needs of users was examined in the dissertation. It should be noted that the functions of the SDN controller

includes the maintenance of service flows, namely the connection, the organization of secure channels, updating information on network connectivity, and many others. On the correct calculations for the service system depends the quality of operations. The work formed the requirements for key indicators of service traffic quality of service in SDN controllers.

SDN controllers are designed in a way to share the functionality of the control system and operation of a plurality of applications. The work developed structural model of controller that combines a set of machines perform their prescribed steps, plurality of functions that are independent and perform necessary operations, the results of which are fed to the machines and system of orchestration of telecommunication services flows in the SDN controller.

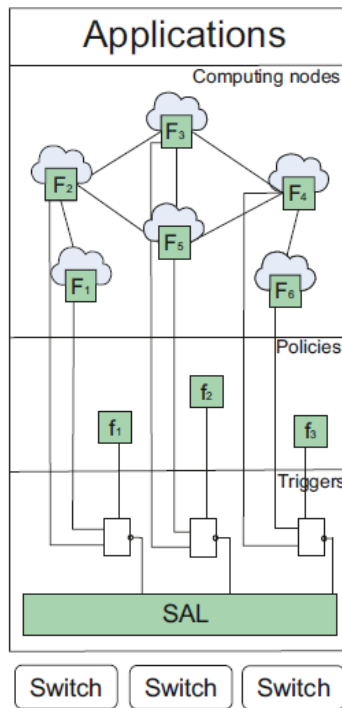


Fig. 1 Structural model for the controller

Based on the structural model the work of SDN controller serving as a unit can be planned. Accordingly, it is possible to collect statistics on the operation of controller subsystems and to calculate resources needed to ensure the increasing flow of applications for service data flows.

For the proposed structural model of the SDN controller, control algorithm of service system based on the proposed method was developed. The peculiarity of the controller subsystems operation is technical possibility to use the cloud-computing environment to perform tasks related to maintenance of telecommunications service. To ensure the efficient use of system resources it is necessary to plan the process of services maintenance. Analysis of computer networks planning methods has shown that static models are mainly used in load control. To the main models of the incoming flow can be attributed Poisson and self-similar flow. As for the level of network management, which corresponds to

SDN controller, maintenance of service network traffic is typical more details are considered of mathematical applications flow forecasting models where intervals between groups of applications distributed according to Poisson. The analysis showed that there are no mathematical models that take into account the gradual change in flow parameters.

That is why the developed method of constructing a resource schedule for long-term statistics with ensuring quality and efficiency indicators to organize the resource usage schedule gives better results in the analysis of service quality indicators and energy consumption.

III. METHOD OF EXPERT ASSESSMENT OF QUEUING SYSTEM POSSIBLE STATES

One of the stages in method of system resources involvement scheduling is to determine the serial number of the system state with setting quality of service parameters. For this purpose, the method of expert assessment of queuing system possible states in which telecommunications network presented as a queuing system with one or more servicing units and limited queue is recommended. This system can be described by a set of possible states, for which ergodic distribution probability that system stays in the selected state is calculated. This task is described as follows.

Input: $1/\mu$ - average time of request service in the node; t - request time delay desired value in the service node (determined by experts); R - the percentage of requests that are served by the system within maximum acceptable delay (determined by experts).

Output: s - number of the system, for which the following condition: the time of request staying in the system does not exceed t (fig. 1)

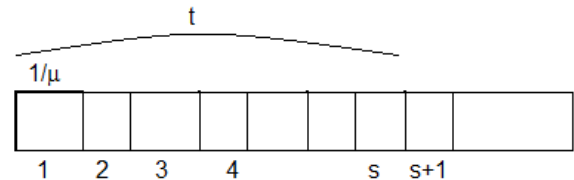


Fig. 1 Representation of telecommunication network as a set of system states

Suppose that the time of request servicing is independent identically distributed random value distributed according to the Poisson law with parameter $1/\mu$, and the time of one request servicing is small compared to t , respectively s - a large number. Based on the assumption there is a central limit theorem:

$$\frac{S_n - (1/\mu)s}{\sqrt{(1/\mu)s}} \sim N(0,1)$$

where S_n - the sum of random variables,

Then, using statistical tables of Laplace function we can find s from the equation:

$$P\left(\frac{S_n - (1/\mu)s}{\sqrt{(1/\mu)s}} \leq t\right) = R$$

$$P(S_n \leq t\sqrt{(1/\mu)s} + (1/\mu)s) = R$$

The solution of this task allows determining the number of state with given configuration of service system.

IV. METHOD OF RESOURCES INVOLVEMENT SCHEDULING BASED ON THE LONG-TERM STATISTICS

Long-term statistics stored in the statistics database, sampling interval time is 5 min.

Incoming data:

- μ^k – request service intensity provided k-th service device configuration that provides quality of service parameters at a necessary level, i.e. the number of requests per unit of time that can be serviced, $k = 1, \dots, K$, K – number of possible configurations of the system.
- $\vec{\lambda}$ – vector of number of applications statistical values ($t \in 0, \dots, T$, $T = 2016$ – number of vectors is 7 (days/a week)*24 (hours)*12 (measurements per hour)). Vector $\vec{\lambda}_t = (\lambda_t^1, \lambda_t^2, \dots, \lambda_t^{50})$ – a set of number of requests values submitted to the service unit and recorded in the i-th time, over the past 50 weeks. This method operates with the average value in the vector

$$\bar{\lambda}_t = (\lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^{50})/50$$

- ε – a small number
- Δ – permissible time of request delay in the service process.
- l – threshold number of requests in the queue
- m – the maximum number of requests in the queue.
- P_{lost} – permissible percentage of lost packets.

Output data:

(t, μ^k), where $t \in [0, T]$ – recommended moments of service intensity changing, μ^k – recommended system configuration, which should jump at time t .

The proposed method

Step 1 consists of the time line smashing into segments, which corresponds sustainable rate approximating of statistics line slope ($\vec{\lambda}_t, t$), found by the least squares method. The main stages of the algorithm:

- To divide set T into subsets,
- To find approximating lines (parameters \hat{a} and \hat{b}) for each of the subsets of the least squares method.
- To compare the parameters \hat{a} and \hat{b} for adjacent subsets.
- If the parameters differ by less than a given small number ε , then subsets combined and again calculated parameters approximating line for the combined set.

- The process of combining subsets of T ends when parameters of approximating line different from a subset related to an amount exceeding ε . As a result of the combination of the sets get set pieces $T_1 \cup T_2 \cup \dots \cup T_n = T$, each i-th interval corresponds to a pair of approximating line parameters \hat{a}_i and \hat{b}_i .

All straight lines defined time intervals are collected in a curve:

$$f(t) = \{x | x = \hat{a}_i t + \hat{b}_i, t \in T_i\}$$

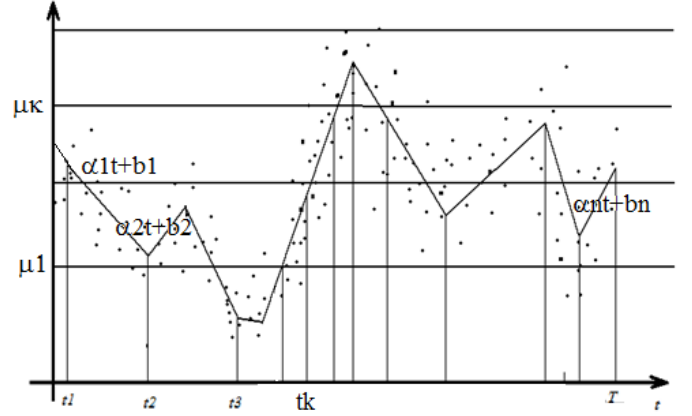


Fig. 2 A characteristic curve

Step 2. As base model assessment of quality indicators chosen model K. Zhernovyi [5], who proposed universal formula for calculating the ergodic distribution of probability logged different number of requests in the system (valid states of Markov chain). Analysis ergodic probability distribution will determine the probability of applications loss due to limits on the length of the queue and will allow assessing the expected requests delay in the system.

Zhernovyi model described as follows: the single-channel system of queuing incoming stream applications (distributed by the Poisson law with parameter λ). The intensity of services distributed by the Poisson law with parameter μ . Queue length is fixed at m , if the application enters the system at the time when in queue there is more applications than l ($l = 0, \dots, (m-1)$), it is discarded.

p_k – ergodic probability distribution of staying service system in different k states of the system ($k = \overline{1, K}$) previously calculated according to the model of queuing, K – set of possible states is organized so that the k -th position corresponds to staying k applications services in the system.

Ergodic distribution of applications number in the system in the case where $\alpha = \frac{\lambda}{\mu} \neq 1$:

$$p_0 = \frac{(1-\alpha)(1-\alpha^{m-l})}{A_l(\alpha)},$$

$$A_l(\alpha) = 1 - \alpha^{m-l} - (m-l)(1-\alpha)\alpha^{m+2}$$

$$p_k = \alpha^k p_0 = \frac{(1-\alpha)(1-\alpha^{m-l})\alpha^k}{A_l(\alpha)}, (k = \overline{1, l+1})$$

$$p_k = \frac{(1-\alpha)(\alpha^k - \alpha^{m+1})}{A_l(\alpha)}, (k = \overline{l+2, m})$$

$$p_{m+1} = \frac{(1-\alpha)^2 \alpha^{m+1}}{A_l(\alpha)} \quad (1)$$

Thus, the probability of request loss in case of the application of above subjects will be calculated by equation (1). The contents of the parameter l means the limit value of applications number in the system when needed queue unloading to reduce service latency. This service discipline meets the early overloads determination method, which is widely used in telecommunications practice.

Step 3. Using the method of expert evaluation of possible queuing system states to calculate the value s^k , corresponding to the k -system configuration.

Stage 4 Soling of the optimization task to find the optimal λ^k , which satisfy the constraints:

$$\begin{cases} \frac{\lambda^k}{\mu^k} \leq \rho \\ p_{m+1} \leq P_{lost} \\ s^k \\ \sum_{k=0} p_k \leq R \end{cases}$$

where R - the percentage of requests that are served by the system no longer than the allowable delay time;

ρ - optimum load of the system meets the energy efficient node operation scheme.

Step 5: Returning to the problem of finding switching points (t, μ^k) , corresponding to possible service system configuration, it is necessary to solve optimization problem for k times, i.e. for each possible system configuration:

$$f(t) \rightarrow \max$$

under restriction $f(t) + 3\sigma \leq \mu^k$

where σ - standard deviation, which is defined by the formula (2).

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (\lambda_i - (\hat{a}i + \hat{b}))^2}$$

As a result, we obtain a set of t^k - switching points in k -one configuration.

So you can build a schedule of network switching to the desired configuration, which allows more efficient use of network resources, supporting quality of service at a given level.

V. APPLICATION OF THE METHOD OF RESOURCES INVOLVEMENT SCHEDULING

The use of the proposed method is possible for various tasks of managing the resources of the queuing system, in

particular when solving the problem of saving energy in the radio access system. The resource of base stations, which provides access to the network for mobile subscribers, is calculated to a large extent, with the aim of providing services at peak times. However, within a day the amount of traffic transmitted through the communication network changes to almost two. Calculation of the necessary configurations of the base station service system will save millions of kilowatts of energy.

The example of load distribution between base stations is shown in Fig. 3.

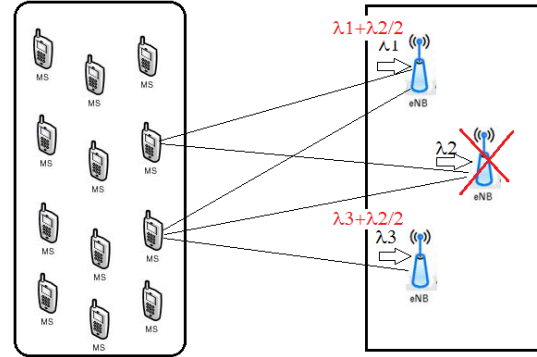


Fig. 3 Distribution of load between base stations on a given territory

CONCLUSIONS

The paper proposed the method of resources involvement scheduling based on the long-term statistics, which also considers the losses of requests, expected delay and efficiency of processing applications in telecommunication nodes. This method can improve network efficiency while preserving the necessary parameters of quality of service.

REFERENCES

- [1] A.Yu. Pilipenko. On the Skorokhod mapping for equations with reflection with the possibility of a jump-like exit from the boundary // Ukrainian Mathematics Journal - 2011, т.63, №9, с.1241-1256.
- [2] Globa L., Novograduska R., Oriekhov O. Method of heterogeneous information resources structuring and systematizing for Internet portals development // EUROCON 2013 IEEE, 1-4 July 2013, PP. 319-326
- [3] Skulysh M., Sulima S. Management of Multiple Stage Queuing Systems // CADSM 2015 : 13-th International conference, 24-27 February 2015 : conference proceedings. — Lviv-Polyana, 2015. — P. 431-434.
- [4] A. Koval, L. Globa, R. Novograduska The approach to web sevice composition // Hard and Soft Computing for Artificial Intelligence, Multimedia and Security. — 2016. — Volume 534 of the series Advances in Intelligent Systems and Computing. — Springer international publication. — P 293-304
- [5] Ya Yeleiko, Yu. Jernovyi. Universal formula for maintenance of M/M/n/m to block incoming flow // Visnyk Lviv Univ. Ser. Appl. Math. Inform. 2009. Is. 15. P. 234-239.
- [6] L. Globa, M. Skulysh, S. Sulima. Method for resource allocation of virtualized network functions in hybrid environment // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6-9 June 2016 : IEEE Digital Library. — Varna, Bulgaria, 2016. — P. 1-5.