Описується метод перерозподілу ресурсу пам'яті програмних модулів багатосерверної модульної системи обробки інформації, що дозволяє враховувати характеристики оброблюваних підпотоків даних та приймати рішення про факт перевантаження програмного модуля. Отримані за допомогою імітаційної моделі результати підтверджують придатність розроблених методу і комплексу алгоритмів для мінімізації локальних втрат в умовах змінної інтенсивності надходження заявок

Ключові слова: багатосерверна модульна система обробки інформації, програмний модуль, управління ресурсом пам'яті

Описывается метод перераспределения ресурса памяти программных модулей многосерверной модульной системы обработки информации, позволяющий учитывать характеристики обрабатываемых подпотоков данных и принимать решение о факте перегрузки программного модуля. Полученные с помощью имитационной модели результаты подтверждают пригодность разработанных метода и комплекса алгоритмов для минимизации локальных потерь в условиях изменяющейся интенсивности поступления заявок

Ключевые слова: многосерверная модульная система обработки информации, управление ресурсом памяти

UDC 004.67
DOI: 10.15587/1729-4061.2018.123257

DEVELOPMENT OF THE METHOD OF DISTRIBUTION OF MEMORY IN NONSTATIONARY LOADED INFORMATION PROCESSING SYSTEMS

A. Minaychev
Adjunct
Department of Informatics and Computing Technologies
Academy of Federal Security Guard Service of the Russian Federation
Priborostroitelnaya str., 35, Oryol, Russian Federation, 302034
E-mail: han57rus@gmail.com

1. Introduction

Over the past decade, there has been an increase in the capacity of modern transport infrastructures, as well as the growing popularity of network services in which they operate. These facts contribute to increasing requirements for multi-server information processing systems (MSIPS), designed to work with such services.

At the same time, the systems of IP telephony, video conferencing, instant messaging and similar services are essentially multiservice software and hardware-software systems. In them, along with priority (depending on functional purpose) service, processing of customer requests of some other services is performed. To improve the security of personal data, such systems are based on the client-server models (client-server encryption) and peer-to-peer models (end-to-end encryption). The latter approach means that the system should provide the decentralized management of the stream of multiservice messages. In general, there is a combined use of the specified models of stream-oriented network services.

Due to the high reliability and cost-effectiveness, the systems of fiber-optic communication lines (FOCL) take leading positions as transmission channels for complex streams of multiservice messages. Over the past two years, only in the Russian Federation, the total length of communication lines of this type has increased by 150 %, and their capacity has tripled. Modern stream-oriented MSIPS should support the reception and processing of such high-speed, complex data streams. An important condition for their functioning is the minimization of the delay in the data stream of the network service they process [1]. To ensure this condition, data processing in MSIPS is carried out using pre-split substreams. Splitting of the input stream into substreams is usually performed either on the basis of the stream number, for example, for talkspurts, or the address information present in the headers of network packets.

The input stream splitting step precedes the substream processing step and is mandatory, since the split substreams can be processed independently. This allows developing MSIPS that implement the processing step in the form of distributed computing systems consisting of N nodes, where N is the number of processed substreams. A node of such a system can be either a computing unit – a server, or a subsidiary distributed computing system, consisting, in turn, of M computing nodes. A general diagram of MSIPS for stream-oriented network services is shown in Fig. 1.

Obviously, the effective operation of the substream allocation server requires a special implementation, supporting the algorithms for the high-speed stream splitting into substreams with a minimum delay. Options of such implementation can be hardware-software solutions based on programmable logic devices (PLD) [2]. In the framework of the research, such servers are considered as generators of load on the MSIPS nodes, which are servers for processing the substream of a particular service (groups of services with similar functionality).

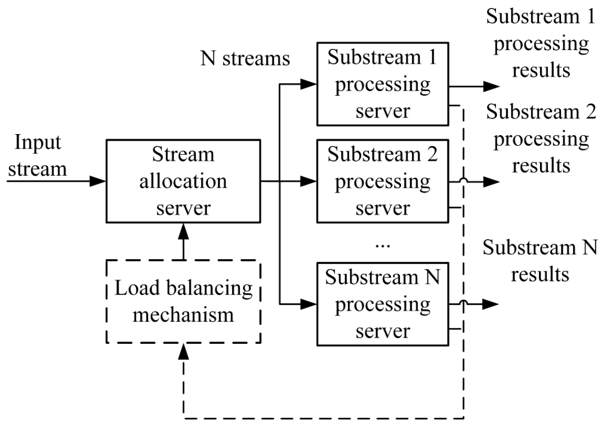

Fig. 1. General diagram of MSIPS for stream-oriented network services

Today, virtualization and cloud computing technologies allow implementing the concept of modular MSIPS, in which the substream processing server is associated not with the computing system, but with the software module (SM) running on this computing system. Such representation of the substream processing server allows introducing the concept of a universal processing server of MSIPS that contains processing SM for all classes of substreams, but specialized to handle substreams of requests of one class. Such a concept allows implementing the modular MSIPS (MMSIPS) in the form of a set of specialized SM. In this case, SM are dynamically distributed over the processors of the substream processing server, depending on the type and complexity of its tasks. Queues of processing requests are placed in the buffer memory (BM) of the SM, implemented in a single memory space of the substream processing server. In the simplest case, in the absence of mechanisms of interaction between SM, MMSIPS can be represented by the multi-server queueing system with a limited queue capacity and failures (Fig. 2).

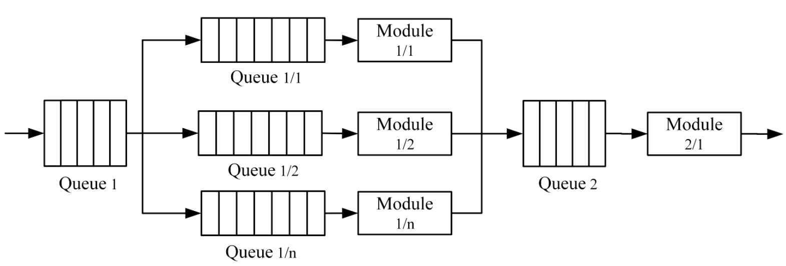

Fig. 2. Diagram of a modular substream processing server

Such an MMSIPS diagram is widely used in commercial implementations of data processing centers (DPC). The functioning of a set of specialized SM on a single platform allows minimizing the time for solving the problem posed to the MMSIPS. In the long term, this significantly reduces the capital and operating costs for the construction, modernization and operation of DPC.

The MMSIPS architecture is widely used in the systems for operative investigative activities (SORM) [3]. In accordance with Federal Law No. 374-FZ of July 6, 2016 (Yarovaya package), the task of such systems is to process and store voice and text traffic of subscribers. The feature of such MMSIPS is functioning on a round-the-clock basis, and the rate of processed information streams depends on such factors as time of day, day of the week, as well as incidents occurring on a global/regional scale. The excess rate of the processed substream over the SM performance leads to an abrupt increase in the queue size in the SM BM. Overloads of the SM (or the substream processing server as a whole) result in loss of the processed information and reduce the MMSIPS performance. Information losses in DPC lead to financial losses, while in case of SORM used by intelligence agencies in communications networks, such failures can endanger state security.

Given the above, it can be concluded that the problem of increasing the MMSIPS performance under non-stationary load is important. Development of the method that minimizes losses of processed information without changing the system architecture is a more cost-effective solution compared to other modern studies.

2. Literature review and problem statement

Research works to determine the load parameters of sections of the GSM network [4] and network devices of new-generation networks (NGN) [5] prove the existence of "bottlenecks" in these networks, which undergo local overloads during peak hours. Such exceptional cases are most likely to occur on days of mass events or extraordinary events (major accidents, terrorist acts, etc.). The study of the load parameters for such situations requires consideration of all characteristics of the real flow to improve the accuracy of calculations in determining the requirements for the on-site equipment. The same conclusions are reached by the authors of the work on simulation of multiservice traffic on the Internet [6]: the integrated serviced stream is characterized by the grouping of traffic packets with random frequency and duration of the peak load.

To prevent overloads in nodes of packet-switched networks, queue management mechanisms have been widely used in recent years [7, 8]. The most common of them are Active Queue Management (AQM) algorithms. In the model of the service system node using this algorithm, each arriving request can be discarded with a certain probability, regardless of the buffer fullness. In such models, this probability is determined by the length of the queue at the time of request receipt and is called the queue dropping function [9]. In [10], it is concluded that the queue dropping function is a powerful regulator of the service node parameters. With its help, the authors control not only the probability of losing a request in the system, but also the parameters of delay jitter and queue length variance. MMSIPS do not allow the presence of losses, so the group of AQM algorithms is not suitable in cases of local overloads.

As a rule, the authors use the queuing theory to simulate the processes with group receipt of requests and analyze active queue management algorithms. A null receiver, where the intervals between moments of receipt of requests in the system are described according to a predetermined distribution law acts as a stream generator with group receipt of requests.

This way of request processing simulation allows reproducing cases of the rate "outburst" beyond expectation with the interval variation coefficient substantially greater than unity. This behavior of the stream of requests implies its self-similar properties (Hurst exponent of 0.5<H<1). Despite all advantages of stream description, the fractal apparatus is quite complex, and its use for constructing a dynamic model requires a large amount of computing resources. To simulate real traffic with similar properties, a number of distributions can also be used: Weibull [5], hyperexponential, log-normal [11], also known as heavy-tailed distributions [12].

The simplest distribution with similar properties is the hyperexponential distribution, which is a special case of multiexponential one, and its distinctive feature is the possibility of approximation of time intervals with the variation coefficient greater than unity.

On the basis of the methodological apparatus of queuing systems, analytical dependencies, reflecting the relationship between the system characteristics and load parameters, by obtaining probability characteristics under the constraint that the process is Markovian were constructed in [13, 14]. In [15], approximation of time intervals between requests of the input stream, the second-order hyperexponential distribution, which allowed reducing the random process to Markovian one and reproducing "packed" traffic properties was used.

As a result, the above authors have estimated the delay time of requests in the system to justify the forward-looking requirements for the MMSIPS on-site equipment. When using the $H_k/M/1$ simulation model, the influence of the variation coefficient on the probability of request processing denial was not evaluated. This is due to the fact that their MMSIPS node is represented by an independent physical server, for which hardware upgrade of components without stopping its functioning is not provided.

On the other hand, this is necessary within the present research, since its object is the MMSIPS, constructed by the Software Defined Networking technology [16, 17], and the amount of SM BM can be changed dynamically. Thus, identification of "packed" properties of the processed SM of the data stream will allow taking timely preventive measures to increase its BM.

3. The aim and objectives of the study

The aim of the paper is to develop a method of inter-module memory resource distribution, which ensures minimization of losses of processed information under varying load on its SM.

To achieve this aim, it is necessary to accomplish the following objectives:

– to develop an MMSIPS SM functioning model for determining the influence of an abrupt increase in the rate of the processed substream on its performance;

– to develop an SM BM resource redistribution method using the mechanisms of distribution of the shared memory resource of MMSIPS;

– to develop a set of SM BM control algorithms, which minimizes losses of processed information through the application of the decisive function of the developed method;

– to develop a simulation model of queue processing in MMSIPS for performing computing experiments that confirm the efficiency of the proposed method and algorithm.

4. Materials and methods for research on data stream processing under varying load

4. 1. Mathematical model of software module functioning under varying load

To simulate the SM functioning process, the first stage of the present research was to use the $H_2/M/1$/n-type queuing model of processing of the substream with the hyperexponential distribution by the SM under the constraint that the request processing time in the processing device is T_{proc}= =1/μ=1 (Fig. 3).

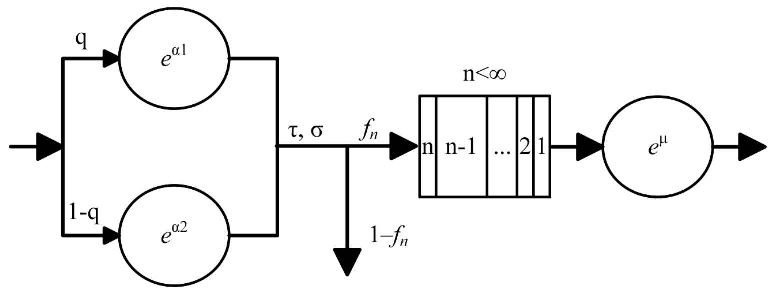

Fig. 3. $H_2/M/1/n$ queuing model of processing of traffic with the hyperexponential distribution by the software module

The figure shows the queuing system with the null receiver, one processing device and BM with a limited capacity $0 \le n < \infty$. The input stream of similar requests is conventionally divided into two parallel phases with different rates, at which the requests enter the system. The request that has completed the process of external generation on the null device with the f_n probability joins the queue in the BM. Otherwise, it leaves the system with the $1-f_n$ probability.

The use of the Kolmogorov equations for such queuing system requires the development of a state graph, which, in turn, leads to the necessity of using the phase method for reducing the non-Markovian system to Markov one.

To this end, the $H_2/M/1/3$-type queuing system was considered, the parameters used to encode the state of the Markov process were determined and the calculation of the number of states for the given queuing system was performed in the research.

The next step in the analysis of the model was to develop a system of differential equations for describing the specified states:

$$\begin{aligned}
&-\alpha_1 p_{01} + \mu p_{11} = p'_{01},\\
&-\alpha_2 p_{02} + \mu p_{12} = p'_{02},\\
&-(\alpha_1+\mu)p_{11} + q\alpha_1 p_{01} + q\alpha_2 p_{02} + \mu p_{21} = p'_{11},\\
&-(\alpha_2+\mu)p_{12} + (1-q)\alpha_1 p_{01} + (1-q)\alpha_2 p_{02} + \mu p_{22} = p'_{12},\\
&-(\alpha_1+\mu)p_{21} + q\alpha_1 p_{11} + q\alpha_2 p_{12} + \mu p_{31} = p'_{21},\\
&-(\alpha_2+\mu)p_{22} + (1-q)\alpha_1 p_{11} + (1-q)\alpha_2 p_{12} + \mu p_{32} = p'_{22},\\
&-(\alpha_1+\mu)p_{31} + q\alpha_1 p_{21} + q\alpha_2 p_{22} + \mu p_{41} = p'_{31},\\
&-(\alpha_2+\mu)p_{32} + (1-q)\alpha_1 p_{21} + (1-q)\alpha_2 p_{22} + \mu p_{42} = p'_{32},\\
&-\mu p_{41} + q\alpha_1 p_{31} + q\alpha_2 p_{32} = p'_{41},\\
&-\mu p_{42} + (1-q)\alpha_1 p_{31} + (1-q)\alpha_2 p_{32} = p'_{42},
\end{aligned} \qquad (1)$$

where

$$p_{ij} = f(t); \quad i,j = 0,1,2; \quad \frac{dp_{ij}}{dt} = p'_{ij}.$$

Upon receipt of the request, the transition to a state with a large number of requests necessarily takes place, and request receipt phases may also change (Fig. 4).

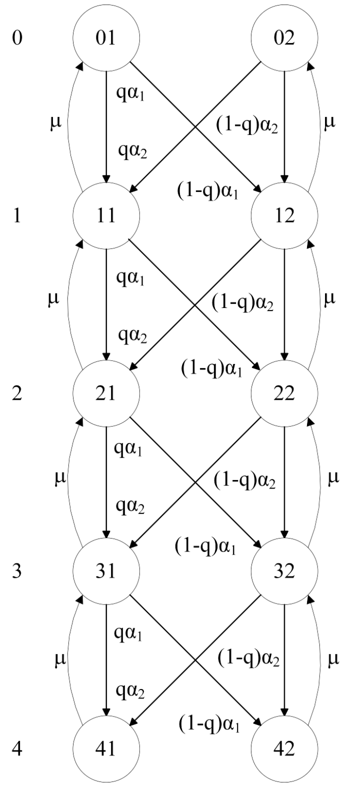

Fig. 4. State transition diagram for the $H_2/M/1/3$ system

The analysis of the state graph of the $H_2/M/1/3$ model and its system of equations showed the influence of the variation coefficient ν on the probability of processing denial of this system. This allowed distinguishing four groups of states necessary for drawing up a universal system of equations in a general form:

$$-\alpha_j p_{0j} + \mu p_{1j} = p'_{0j},$$

$$-(\alpha_j+\mu)p_{1j} + \mu p_{2j} + q_j\sum_{m=1}^{k}\alpha_m p_{0m} = p'_{1j},$$

$$-(\alpha_j+\mu)p_{nj} + \mu p_{n+1j} + q_j\sum_{m=1}^{k}\alpha_m p_{n-1m} = p'_{nj},$$

$$-\mu p_{N+1j} + q_j\sum_{m=1}^{k}\alpha_m p_{Nm} = p'_{n+1}, \qquad (2)$$

where

$$p_{nj} = f(t);\quad j=\overline{1,k};\quad n=\overline{2,N};\quad \frac{dp_{nj}}{dt} = p'_{nj}.$$

It is necessary to add the normalization condition to the system (2):

$$\sum_{n=0}^{N+1}\sum_{j=1}^{k} p_{nj} = 1. \qquad (3)$$

The system of equations (2) and the normalization condition (3) are a universal tool for studying the $H_2/M/1/N$ systems. To simulate the processing module, state probabilities considering the number of requests processed by it, the N value and the value of its load r are required.

4. 2. Functioning of the system using the two-level diagram of inter-module memory resource distribution

The next stage of the research was to develop a diagram of the SM BM resource. This was done by means of the memory paging diagram, also used for migration of virtual machines [18]. The existing paging methods and algorithms are well suited for application to the SM BM elements. Herewith, the two-level SM BM resource management diagram, based on a set of local managers (LM) and global managers (GB) is proposed (Fig. 5). The developed memory resource management algorithm according to this diagram is shown in Fig. 6.

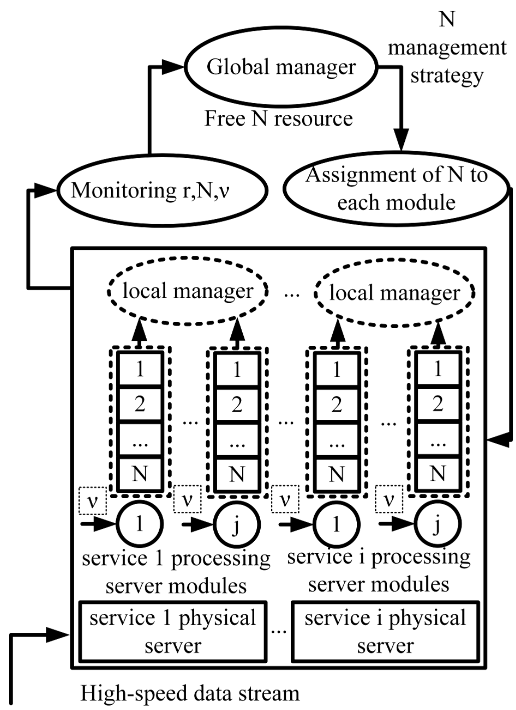

Fig. 5. Diagram of two-level buffer memory resource management

According to the proposed diagram, the LM is hosted on each MMSIPS physical server and performs the function of monitoring the server characteristics (r, N) and the values of the substreams at the SM inputs associated with processing a particular network service.

In turn, the GB implements the following functions:

1. Tracking the rate "outbursts" of substreams to determine their anomaly.

2. Gathering information about the values of the characteristics (r, N) of SM of all physical servers to determine the degree of their load.

3. Based on the data obtained (paragraphs 1 and 2), determines the SM status and, in the case of SM overload, implements the SM BM resource redistribution mechanism.

To determine the SM operation mode, the hysteresis load control algorithm [19], having proved itself well in solving the problems of overload control in DPC is used.

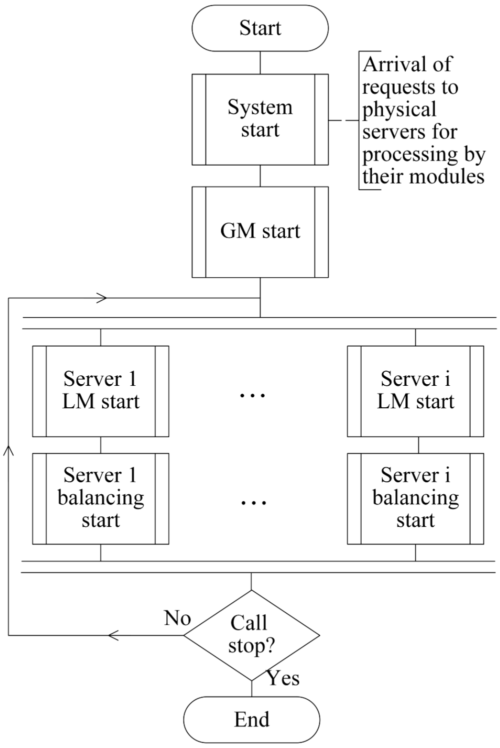

Fig. 6. Diagram of the buffer memory resource management algorithm based on the two-level control mechanism

With regard to the developed algorithm, the following modes of hysteresis control were set (Fig. 7):

1. Norm – the number of requests in the system does not exceed the threshold H, but exceeds the threshold L.

2. Overload – the number of requests exceeds the H value. In this case, the LM registers this fact, and the GB implements the SM BM resource redistribution mechanism.

3. Memory lockout – the number of requests does not exceed the threshold L. This event is registered by the GB and the amount of the SM BM resource will not decrease below the threshold L.

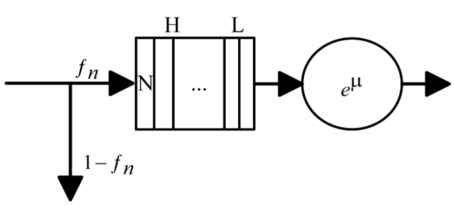

Fig. 7. Simplified scheme of the $M|M|1|\langle L, H\rangle|$N-type queuing system using hysteresis

The schemes and the algorithm described were laid in the simulation model of request processing in the MMSIPS. Multiple repetition of the experiment by means of this model has allowed evaluating their effectiveness.

5. Results of the research

5. 1. Result of the simulation of the software module functioning under varying load

The results of estimating the decrease in the substream processing quality of the SM according to the P_{otk} parameter depending on the r and v values are shown in Fig. 8.

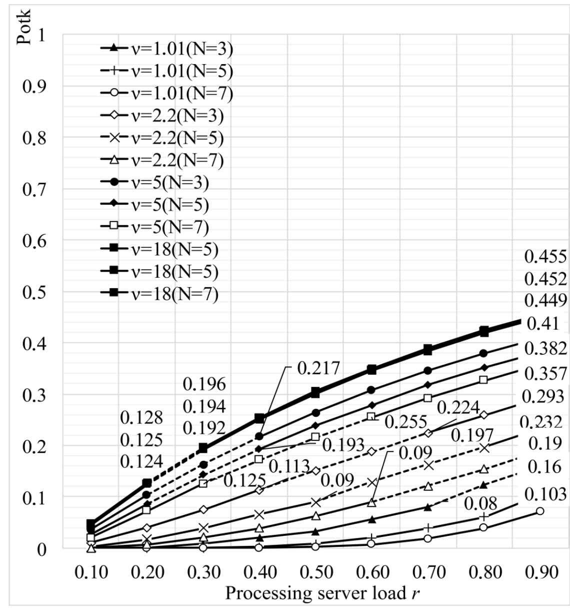

Fig. 8. Dependence of P_{otk} of the $H_2/M/1/N$ system (where N={3, 5, 7}) under increase in load r

When analyzing the results obtained, we note that the v value of the substream is a key indicator of SM overload and can determine the following zones of its functioning depending on the r and N values (the valid value of P_{otk}=0.2 is taken as an example):

– the zone of SM normal functioning (0; 0.1);

– the zone of SM load, dwell time (0.1; 0.2);

– the zone of SM overload and unacceptable losses of processed requests (0.2, 1).

Definition of these zones depends on the processing quality requirements for requests of the appropriate type.

5. 2. Results of algorithmization of functioning schemes of managers

The task of the LM is to monitor the SM characteristics: the current volume of the virtual resource assigned to it – N, the degree of load – r, the value of the variation coefficient – v. After that, the LM compares these characteristics with the threshold values obtained from the load table. The result of the comparison is the value of the flag (PiFlag) – the key parameter required for the GM. The characteristics and the PiFlag value obtained are rewritten in the load table on the GM side (Fig. 9).

The central link in the two-level management of MMSIPS is the GM (Fig. 10).

Functioning of the GM begins with the formation of the system load table, containing the performance parameters of all SM and key characteristics of the streams they process. Information in the load table is updated due to the information obtained from the LM, located directly on the nodes. In the developed algorithm, the main "marker" to activate the server balancing mechanism is the Z flag, which is assigned the "1" value in case the LM detects an anomalous behavior of at least one substream processed by the SM and/or if the hysteresis threshold H is exceeded on one of the SM. Server balancing involves defining the P_{otk} functional zone for each SM and presenting them in the form of a ranked list the procedure will work with. Ranking is based on the "PiFlag" value and the degree of their load ρ, from the most loaded SM to the less loaded one. This operation is necessary to determine the "exchanging" pairs of SM. "Exchange" means an increase in the BM of the top N in the SM pair due to the lower N in the SM pair. The SM pair is selected according to the rule: the first top SM in the list and the last lower SM in the list, the second SM in the list and the "penultimate" SM, etc. The "exchange" process at this stage looks like the generated resource distribution strategy, which is applied if the average P_{otk} of the SM does not change. If this value decreases, the strategy will be improved after going through another cycle of "exchange", starting with ranking.

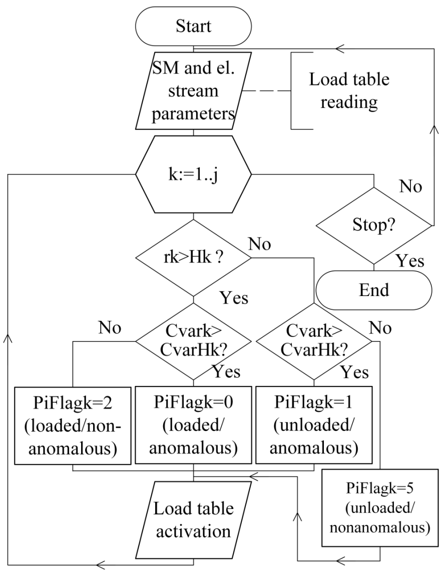

Fig. 9. Diagram of the local manager functioning algorithm

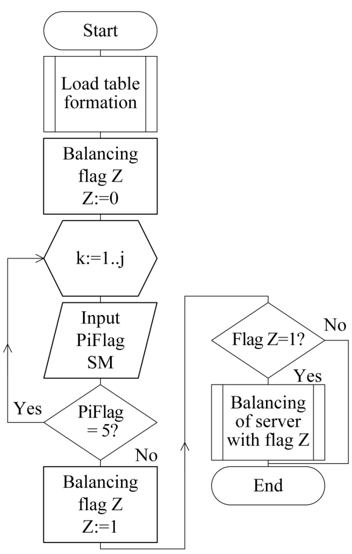

Fig. 10. Diagram of the global manager functioning algorithm

It should be noted that after applying the balancing strategy, the GM introduces new values of thresholds and SM parameters into the load table.

5. 3. Results of the experimental research on the efficiency of the inter-module memory resource distribution mechanism

In the course of the research, the simulation model of request processing in MMSIPS with the implemented set of algorithms for the BM resource distribution between the set of SM was developed.

The software implementation of the simulation model is the author's development, confirmed by the certificate of state registration of the computer program No. 2017617207. On its basis, the computing experiment was planned according to the method of scenario planning. The scenario refers to a sequence of events that simulates the functioning of the current version of MMSIPS and the version of MMSIPS with a set of BM resource redistribution algorithms.

Design of the computing experiment was carried out on the basis of a full factorial experiment, for which, in accordance with the GOST, the corresponding planning matrix was compiled using the method of selection and optimization of the controlled process parameters.

Proceeding from the aim of the research, the response variable was the probability of SM processing denial of the current request.

The experimental factor was the number of "problem" SM (characterized by the "growth" or "bursts" of the arrival rate of the input substream). The factor levels ranged from 1/4 ("a" letter) to 4/4 ("g" letter) of all SM, for which the BM resource redistribution was implemented.

During the experiment, four types of scenarios were used:

1. Gross increase in the arrival rate of requests ("Collapse" scenario);
2. Time-spaced "bursts" of growth in the arrival rate of requests ("Event" scenario);
3. Simultaneous long "bursts" of growth in the arrival rate of requests ("Attack" scenario);
4. Time-spaced, group "bursts" of growth in the arrival rate of requests ("Holiday" scenario).

An example of the obtained experimental results without and with resort to the developed set of algorithms is shown in Fig. 11.

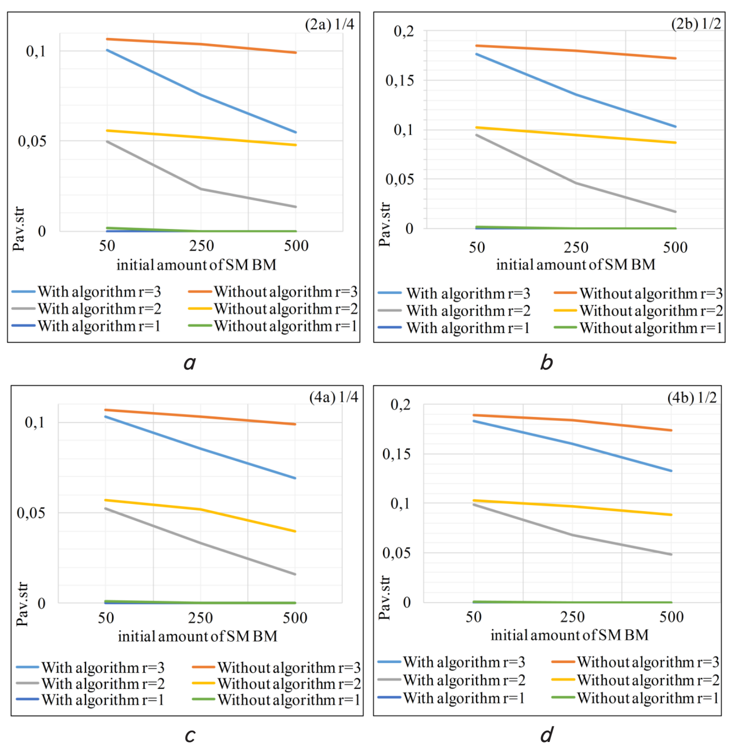

Fig. 11. Results of the experiment for scenarios: *a* – "2a"; *b* – "2b"; *c* – "4a"; *d* – "4b"

The results of the computing experiment show that the gain of the proposed set of algorithms is especially significant for 2 and 4 scenarios. This is particularly true for MMSIPS configurations with high initial amount of SM BM.

6. Discussion of the results of simulation and computing experiment

In the simulation of the SM functioning under varying load, the following assumptions were introduced:

1. For the approximation of time intervals between adjacent requests, the second-order hyperexponential distribution based on a similar approach to the simulation of multiservice traffic in infocommunication systems is selected.
2. Request processing time is distributed according to the exponential law.
3. Values of $\nu \in [1.01; 18]$, corresponding to multiservice traffic in real infocommunication systems.
4. Processing module load $r \in [0.1, 0.9]$.
5. Amount of BM $N = \{3, 5, 7\}$.

The graphs obtained during the simulation show that the value of the variation coefficient determines the zones of the SM functioning.

Upon transition of the ν value (for $\nu < 0$) from the normal functioning zone to the load zone, measures should be taken to improve the substream processing quality of the SM. For this purpose, it is necessary to implement the procedure of increasing the N value of this SM, due to the resource of free SM BM with low load (located in the normal functioning zone).

The main advantage over other approaches is the preventive adoption of measures to minimize losses on the processing module, which is especially critical for cases of arrival rate “bursts”. This approach is applicable to any MMSIPS configurations and allows a more rational use of all system resources. In turn, the process of “exchange” can be used in any shared-memory system, which makes it universal.

The developed simulation model, in which this approach is laid, proves the efficiency of the method and algorithms in the two-level BM resource management system, based on the hysteresis load control method.

The results of the computing experiment by means of the simulation model have led to the conclusion that the developed method and the set of algorithms allow reducing local losses of processed information to 10 %, without the need for hardware upgrade of the MMSIPS components.

The drawbacks of the method include its inefficiency in cases of gross increase in arrival rate. Consideration of such scenarios can be the subject of further research. In the future, it will be necessary to improve the mechanisms for exchanging service information between agents and conduct additional experiments.

7. Conclusions

1. The model of the MMSIPS SM functioning, adequately describing the functioning of the software module under varying input load is developed. Using the prescribed distribution law, analytical expressions for estimating the probabilities of processing denial of requests of input substreams are found. The feature of the model is the presence of SM functioning zones, which are necessary for the correct operation of the whole BM resource redistribution method.
2. The method of the MMSIPS SM BM resource redistribution, allowing to consider the characteristics of processed data substreams and to decide on the fact of the software module overload is developed. The use of the two-level diagram based on managers is proposed.
3. The set of algorithms for the BM resource distribution in MMSIPS, based on the hysteresis load control method is developed. They allow implementing a rational version of MMSIPS operation and reducing losses of processed information.
4. The simulation model of request processing in MMSIPS, which differs from the known ones in the possibility of generating the samples of processed substreams is developed. It allows simulating an abrupt increase in arrival rate, associated with real incidents.

The results obtained using the simulation model confirm the conclusions about the suitability of the developed method and the set of algorithms. The scenario-based experiment method showed that the proposed method allows reducing losses of processed information to 10 %. It is worth noting that the efficiency of the algorithms and method in comparison with the existing MMSIPS solutions mainly depends on the embedded scenario. However, the solutions obtained contribute to the improvement of the methodology of resource management of computing systems.

References

1. Minaychev A. A. Podhody k modelirovaniyu potokovo-orientirovannyh setevyh servisov s blizkim k real’nomu vremenem // Informatsionnye tekhnologii modelirovaniya i upravleniya. 2015. Vol. 92, Issue 2. P. 164–171.
2. Alphonsus E. R., Abdullah M. O. A review on the applications of programmable logic controllers (PLCs) // Renewable and Sustainable Energy Reviews. 2016. Vol. 60. P. 1185–1205. doi: 10.1016/j.rser.2016.01.025
3. Elagin V. S. Inzhenernye problemy SORM // Vestnik svyazi. 2017. Issue 9. P. 9–14.
4. Investigation into the parameters of the traffic of the gsm network section under the effect of local overflows / Smetanin I. N., Lozhkovsky A. G., Piza D. M., Verbanov O. V. // Telecommunications and Radio Engineering. 2013. Vol. 72, Issue 10. P. 893–905. doi: 10.1615/telecomradeng.v72.i10.50
5. Spline-approximation-based restoration for self-similar traffic / Strelkovskaya I., Solovskaya I., Severin N., Paskalenko S. // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 3, Issue 4 (87). P. 45–50. doi: 10.15587/1729-4061.2017.102999
6. Kartashevskiy I. V., Buranova M. A. Vliyanie mekhanizmov upravleniya QoS na pokazateli kachestva obsluzhivaniya mul’timediynogo trafika seti Internet // T-comm: telekommunikatsii i transport. 2013. Issue 8. P. 54–60.

7. Kempa W. M. A Direct Approach to Transient Queue-Size Distribution in a Finite-Buffer Queue with AQM // Applied Mathematics & Information Sciences. 2013. Vol. 7, Issue 3. P. 909–915. doi: 10.12785/amis/070308
8. Tikhonenko O., Kempa W. M. Queue-Size Distribution in M/G/1-Type System with Bounded Capacity and Packet Dropping // Communications in Computer and Information Science. 2013. P. 177–186. doi: 10.1007/978-3-642-35980-4_20
9. Chydzinski A. Nowe modele kolejkowe dla wezlow sieci pakietowych. Pracownia Komputerowa Jacka Skalmierskiego, Gliwice, 2013. 286 p.
10. Time-Varying Performance Analysis of Multihop Wireless Networks With CBR Traffic / Xu K., Tipper D., Qian Y., Krishnamurthy P., Tipmongkonsilp S. // IEEE Transactions on Vehicular Technology. 2014. Vol. 63, Issue 7. P. 3397–3409. doi: 10.1109/tvt.2013.2297382
11. Gil J. Y., Monni C. Fitting long-tailed distribution to empirical data // Concurrency and Computation: Practice and Experience. 2017. Vol. 29, Issue 34. P. e4223. doi: 10.1002/cpe.4223
12. Cooke R. M., Nieboer D., Misiewicz J. Fat-tailed distributions: Data, diagnostics and dependence. John Wiley & Sons, 2014. 144 p. doi: 10.1002/9781119054207
13. Aliev T. The Synthesis of Service Discipline in Systems with Limits // Communications in Computer and Information Science. 2016. P. 151–156. doi: 10.1007/978-3-319-30843-2_16
14. Verbanov O. V., Lozhkovskiy A. G. Modelirovanie pachechnogo trafika i otsenka koeffitsienta samopodobnosti // Vymiriuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh. 2014. P. 186–188.
15. Ushanev K. V. Imitatsionnye modeli sistemy massovogo obsluzhivaniya tipa Pa/M/1, H_2/M/1 i issledovanie na ih osnove kachestva obsluzhivaniya trafika so slozhnoy strukturoy // Sistemy upravleniya, svyazi i bezopasnosti. 2015. Issue 4. P. 217–251.
16. Implementation experience in multi-domain SDN: Challenges, consolidation and future directions / Katsalis K., Rofoee B., Landi G., Riera J. F., Kousias K., Anastasopoulos M. et. al. // Computer Networks. 2017. Vol. 129. P. 142–158. doi: 10.1016/j.comnet.2017.09.005
17. Are we ready for SDN? Implementation challenges for software-defined networks / Sezer S., Scott-Hayward S., Chouhan P., Fraser B., Lake D., Finnegan J. et. al. // IEEE Communications Magazine. 2013. Vol. 51, Issue 7. P. 36–43. doi: 10.1109/mcom.2013.6553676
18. MECOM: Live migration of virtual machines by adaptively compressing memory pages / Jin H., Deng L., Wu S., Shi X., Chen H., Pan X. // Future Generation Computer Systems. 2014. Vol. 38. P. 23–35. doi: 10.1016/j.future.2013.09.031
19. A Study of Systems with Multiple Operating Levels, Soft Thresholds and Hysteresis / Brandwajn A., Begin T., Castel H., Atmaca T. Inria-Research Centre Grenoble-Rhône-Alpes, 2017. 14 p.