

# Quality Control of Mobile Communication Management Services in a Hybrid Environment

Globa L.<sup>1</sup>, Skulysh M.<sup>1</sup>, Romanov O.<sup>1</sup>, Nesterenko M.<sup>1</sup>

<sup>1</sup> National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Peremoga Ave. 37, 03056 Kyiv, Ukraine  
lgloba@its.kpi.ua, mskulysh@gmail.com, a\_i\_romanov@gmail.com,  
nikolaiy.nesterenko@gmail.com

**Abstract.** There is an integration of telecommunication systems and distributed computing environment, resulting in a single hybrid environment for telecommunication services. The hybrid environment has the ability to control the information flow process at each stage and ensure compliance with high quality standards. Providing the quality of service to end-users of communication networks depends on quality control at all stages of the provision of the service. Today, due to the dynamically changing service structure provided to end users, constantly changing requirements for service quality indicators, with increasing traffic volumes, there is a growing need for well-scalable communication systems that meet the needs of end-users, gaining special significance service management systems. In the article the features of quality control service servicing of flows of the mobile communication network are investigated with the use of partial virtualization of network functions. An architectural solution is proposed for organization of service flows in a hybrid environment, which includes a telecommunication communication network and cloud computing resources that provide maintenance of virtualized functions involved in the organization of service flows. The solution for improvement of the PCRF system as well as a number of procedures that allow ensuring quality control of servicing streams as well as controlling the computing resources of a hybrid system, whose work affects the quality of service of service flows of the system, is proposed.

**Keywords:** NFV, VeCME, VBS, VeEPC, LTE, 5G, TC gibrid-service.

## 1 Introduction

The work of the telecommunications network is inextricably linked with computer systems. According to [1], a hybrid telecommunications service is a service that includes components of cloud and telecommunications services. A mobile network consists of a local area network, a radio access network and a provider core network. The advent of cloud computing has expanded the possibilities for servicing telecommunications systems.

Specifications [2] represents the main architectural solutions in which complex hardware solutions are replaced by different ways of virtualizing network sections. This allows configuring the network computing resources in a flexible way. To do this, it is necessary to create new methods of managing the quality of service that will take into account the features of the process in the telecommunication system and in the computing environment for servicing hybrid services.

The purpose of this work is to improve the quality of service of hybrid telecommunication services. To this end, it is proposed to use methods to control the formation of service request flows and to manage the allocation of resources. Realization of the set goal is achieved by solving the following tasks:

1. Research of developments of the scientific community in the field of monitoring and ensuring the quality of service of hybrid services. Identification of processes regularities and features.
2. Development of a model of servicing hybrid services for a heterogeneous telecommunication environment.
3. Development of methods for ensuring the quality of service in access networks.
4. Development of methods for ensuring the quality of service in the local networks of the mobile operator and on their borders.
5. Development of a model and methods for the operation of the provider core network in a heterogeneous cloud infrastructure
6. Development of a functioning model for provider charging system.

To realize these tasks, it is necessary to take into account such factors as the annual growth of traffic volumes in an exponential progression; the need for differentiated services for a multiservice flow and different quality of service requirements; the need for constant monitoring of quality indicators and timely response to their decline.

Thus, the operator's monitoring system collects and processes a large amount of information about the quality of each service. It also monitors the telecommunications operator subsystems, the number of failures, etc.

For an adaptive response to the decline in quality of service today, such mechanisms are used:

- Monitoring the workload of the network;
- Monitoring of queue service quality in communication nodes;
- Managing of subscriber data flows;
- Managing of queues for differentiated servicing of multi-service flows
- Overload warning mechanisms;
- Methods of engineering the traffic for an equable distribution of resources.

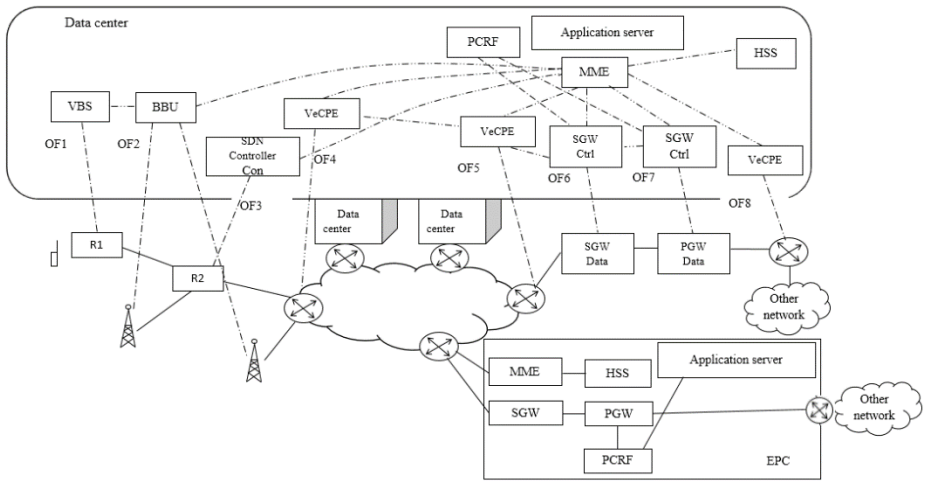
## **2 Organization of a heterogeneous telecommunications network**

According [3], all calculative functions that accompany transfer process are performed in data centers with cloud infrastructure. Virtualization of the base station will reduce the amount of energy consumed by the dynamic allocation of resources and load balancing. In addition to virtual base stations, radio access networks with cloud-based resources organization (Cloud-RAN) is required to create a resource base frequencies processing, which will combine different computing resources of centralized virtual environment. The specification offers virtualization of network functions for the router located on the border of provider local network. A router performs flow classification, routing management and providing firewall protection.

For the organization of virtual base stations and VeCPE it is necessary for data center to be close to base stations and to each output of the local network. So, the of the provider network represents a geographically distributed network of data centers with communication channels delivered primary information of mobile subscribers to each

of them. The network requires conversion at the lowest level, so the signal requires recognition and decoding at higher levels of MAC, RLC, RRC, and PDCP. The specification is also propose provider core virtualization.

Based on this, it can be assumed that most of the network processes are performed in datacenters, and the network is only a means of delivering information messages [4]. In the conditions of program-controlled routers distribution, there is network structure shown in Fig. 1.



**Fig.1.** Provider core network structure using software-controlled routers

Fig. 1 shows how the mobile subscriber communicates with the R1 transponder, which converts the radio signal to optical, and then the signal reaches the R2 transponder managed by the SDN controller, which is also situated in the data center. After attaining the data center, the signal is processed by the virtual base station. Further, according to LTE technology, the flow is sent to the operator's core for further processing. The BBU subsystem is based on the technology of software-configurable networks/virtualization of network operation. This system supports either the work of virtual base stations or hybrid of 2G/3G/4G/Pre5G solutions.

Further direction of data channels is determined by servicing in the core. If the flow is directed to the provider internal network, it is immediately sent to the corresponding virtual base station in the data center for service, and then forwarded to the subscriber through the transponders R2 and R1. If the stream is to be sent outside the operator's local network, it is directed to the boundary virtual router, and then to external networks. This is the example of Next Generation Network

Thus, the data center combines a group of data centers that are connected to a single logical space for servicing virtualized network functions through a secure network. The quality of end users service is influenced by the organization of processes in such a heterogeneous data center based on the cloud computing concept.

According to the ITU-T Y.3500 recommendation, cloud computing is the paradigm network access providing to a scalable and flexible set of shared physical and virtual resources with administration based on on-demand self-service.

The structure of described data center in which the group of functional blocks shown in fig. 1 are servicing is shown in fig. 2. There is a transport network and connected data centers, forming a single virtualized space.

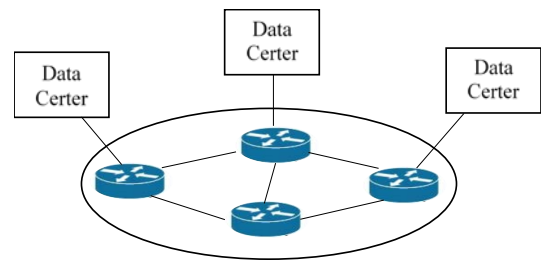


Fig. 2. The structure of the heterogeneous data center

Recommendation ITU-T Y.3511 defines this complex system of data center groups as multi-cloud computing. It is a paradigm for interaction between two or more providers of cloud services. Recommendation ITU-T Y.3520 presented the conceptual architecture of the multi-cloud and multi-platform cloud services management presented in fig. 3 [5]

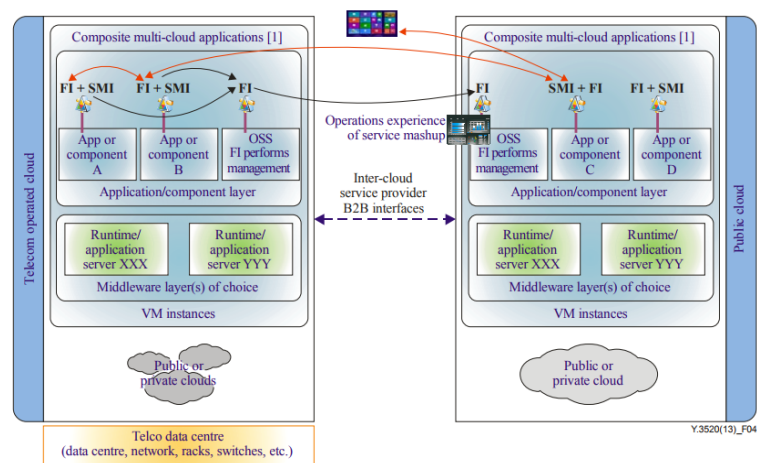


Fig. 3. Architectural vision for multi-cloud, multi-platform cloud management

During the work of provider data center, virtual BS system, the core subsystems and the virtual router are in a single logical space. In fig. 3 we can see that at the middleware level XXX Server is presented in every data center that participates in the inter-cloud computing infrastructure. The corresponding programs that activate the provider functional blocks are performed at the application and component level.

To ensure the work of mobile network using virtualization technology, it is necessary to provide a distributed structure of data centers, organized in a single virtual space. The structure should include deployed logical elements of the mobile service network, process management and flow allocation carried out by the orchestrator (fig. 4).

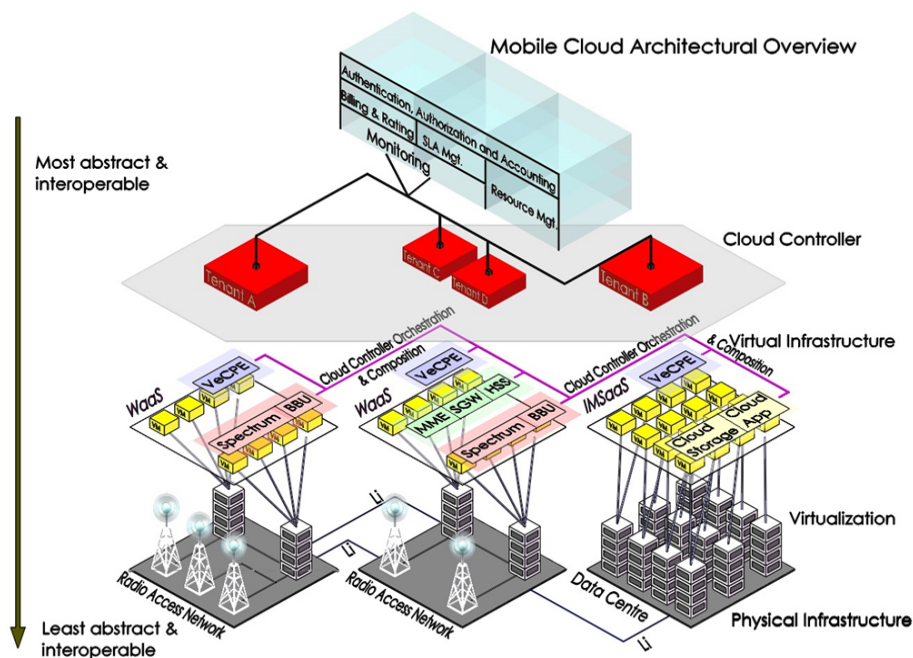


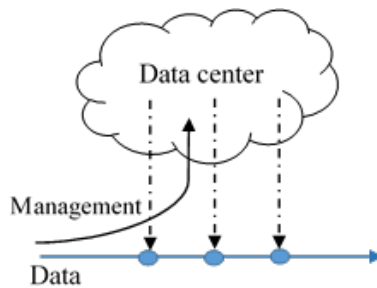
Fig. 4. Organization of service in New Generation Networks

According to the research, the effectiveness of computing processes organization in functional units affects the efficiency of end-users servicing of a mobile operator. The data processing center in this architectural solution is a complex organizational and technical set of computing and telecommunication systems that ensures the smooth operation of the NFV infrastructure. The effectiveness of its operation depends on the choice of physical data centers that will become part of the distributed center structure; the location of network functions in the infrastructure; the organization of flows between virtualized objects and the allocation of resources for their servicing.

### 3 The principle of flow service with the resource virtualization in public telephone network

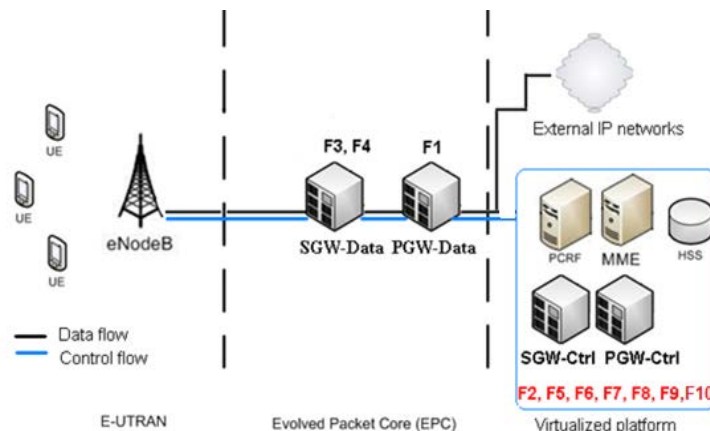
Controllers located in the data center guide all subsystems of mobile communication. The interaction between controllers of subsystems for the purpose of control occurs only in the middle of the data center. The functions of managing the service process, namely: searching for the subscriber, searching for the physical elements involved in the transmission process, and passing the guidance on the corresponding physical elements, are transferred to the cloud.

The subscriber device for connection organization interacts with the base station controller located at the data center. According to the protocols, subsystem controllers interact at the level of the data center, sending the final hardware solutions to the physical equipment to start the data transmission process (Fig. 5).



**Fig. 5.** The principle of flow service with the resource virtualization in public telephone network

There are two principles of virtualization of network resources. The first principle redirects through the cloud resources only control flows. The second principle is to use cloud-based data centers to process both network and information flows. In this paper, the first principle is considered. According to it, virtualization of network functions allows separating the control system of the mobile network nodes from the data transmission system. In fig. 6 you can see an example of separating the functions of the core network. The main functions of the core subsystem were analyzed, and the functions associated with the control and data transfer were selected. Data transfer functions are solved in a virtualized environment deployed on the basis of data centers group [6].



**Fig. 6** Distribution of network core functions between physical and virtual devices

**F1** – packet filtering by users and legitimate interception of traffic;

**F2** – IP pool distribution functions for UE and PCEF Functionality;

**F3** – basic routing and interception of packet traffic;

**F4** – the function of an anchor point (traffic aggregation point) for a handover between the NodeBs within one access network in the base station service area according to a set of rules and instructions;

**F5** – processing of BBERF functionality;

**F6** – Traffic Detection Function;

**F7** – User Data Repository (UDR);

**F8** – Application Function (AF);

F9 – Online Charging System (OCS).

Fig. 7 shows the processes of network subsystems interaction with the separation of control functions and data transmission with virtualization in the provision of data transfer functions. In fact, each arrow on this scheme is a service request in this virtual (or physical) node. The number of requests per time unit is the load intensity on given service node.

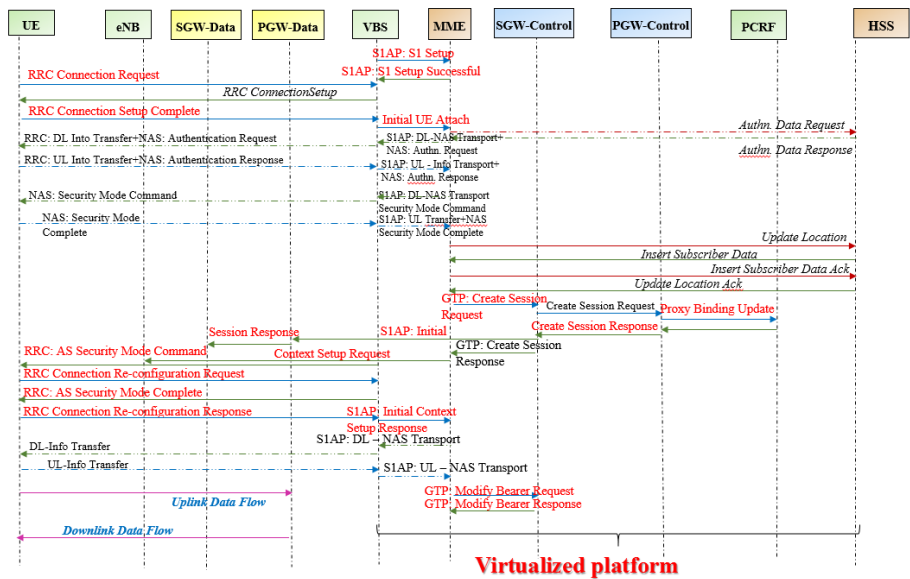
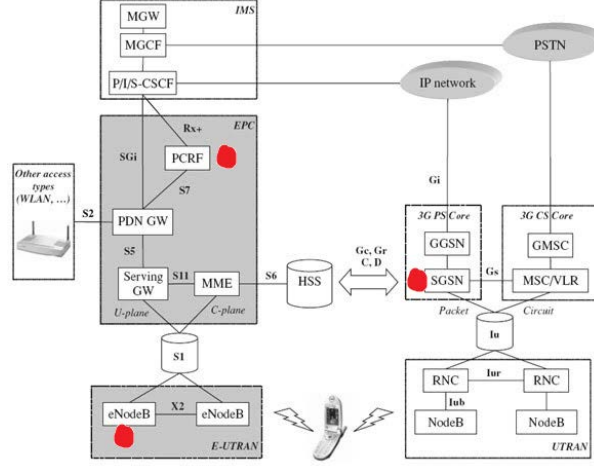


Fig. 7. Procedure of subsystems interaction during subscriber's service

Network structure and user service quality control take place in the nodes. Traditionally, the subsystems of LTE network perform a set of functions, in accordance with standards and specifications. The paper proposes to divide subsystem management functions and functions that are associated with the data transfer process directly to the LTE network. The feature is the expansion of subsystems functionality, compared with the networks of previous generations. More than half of the subsystem functions are connected not with the service process, but with the management of the communication system. Service quality control occurs in the subsystems eNodeB, SGSN, PCRF (fig. 8). Delay control in virtualized network nodes, where service intensities depend on computing resources requires PCRF modification.



**Fig. 8.** General architecture of the standard LTE network

The efficiency of hybrid telecommunication services is estimated by quantitative indicators of service quality:

- $t_d$  – time delay in the maintenance of the hybrid telecommunication management service ( $t_{data} - t_{start}$ ), where  $t_{start}$  is the moment of request by the subscriber for permission to transmit data information flows,  $t_{data}$  is the moment when the subscriber begins to transmit information streams;
- $P$  – the probability of refusal in service.
- 

$$P = \prod_{i=1}^N P_i$$

where  $P_i$  is failure probability in virtualized service node for one of the requests types to the subsystem of the heterogeneous telecommunication environment.

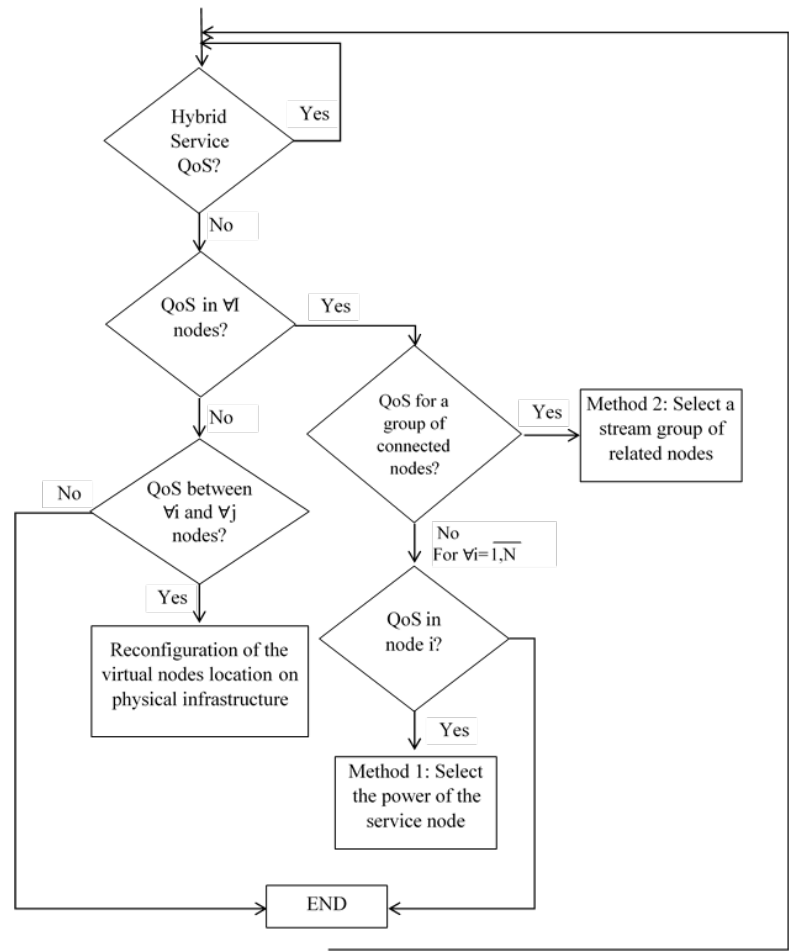
#### 4 Procedure of guaranteeing the adjusted quality of service

The principle of dynamic quality control is as follows: the delay value in maintaining the application for connection (disconnection, recovery) is compared with the service quality policy of the subscriber. If the metric does not match, then the quality metrics in virtual nodes and VLANs are consistently compared with the thresholds of the corresponding policies stored in the PCRF subsystem. This principle analyzes the following quantitative indicators of the effective system operation, such as: the time of service flow request delay in the virtual node and the probability of queries loss in the service node. Service node is a virtual machine that performs the functions the network node managing.

After discovering the reason of service efficiency indicators problem, then appropriate measures are taken. If there is a problem in the time of transmission between service nodes, then it is recommended to reconfigure the system, namely to change the location of virtual nodes in physical nodes of the heterogeneous data center structure. If the problem is identified in one service nodes, then it is recommended to



increase the number of service resources. If there is a decrease in service quality rates in a group of linked interface nodes, for example, which form a single core of the EPC network, then it is recommended to limit the flow of applications sent to service the corresponding core. For this purpose it is recommended to calculate the intensity of the load on the group of nodes. The algorithm of the procedure is shown in fig. 9

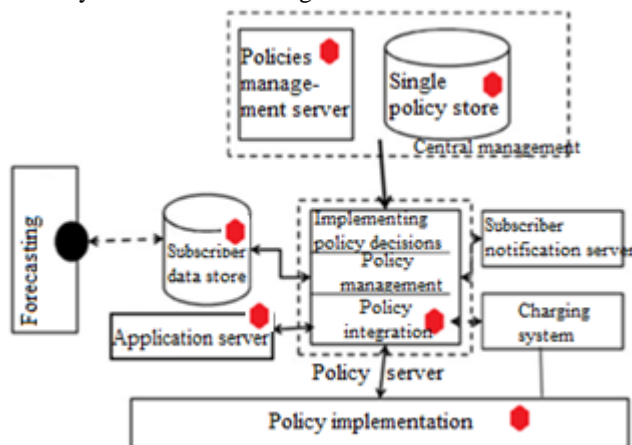


**Fig. 9.** Procedure for guaranteeing the preset quality of service

To implement the principle of dynamic quality control, a modification of the PCRF system subsystems is required. The "Single Policy Storage" subsystem is expanding, and the following policies regarding quality management service flow rates are added:

1. The allowable delay time for an application service flow in a virtual host.
2. Permissible loss of requests in the virtual node
3. Permissible time for serving requests in groups of virtual nodes that provide a given service.
4. Permissible delays in transmission between service nodes
5. The value of the admissible delivery delays of the guiding influence on network nodes.

An expanded subsystem is shown in Fig. 10



**Fig. 10.** PCRf subsystem modification

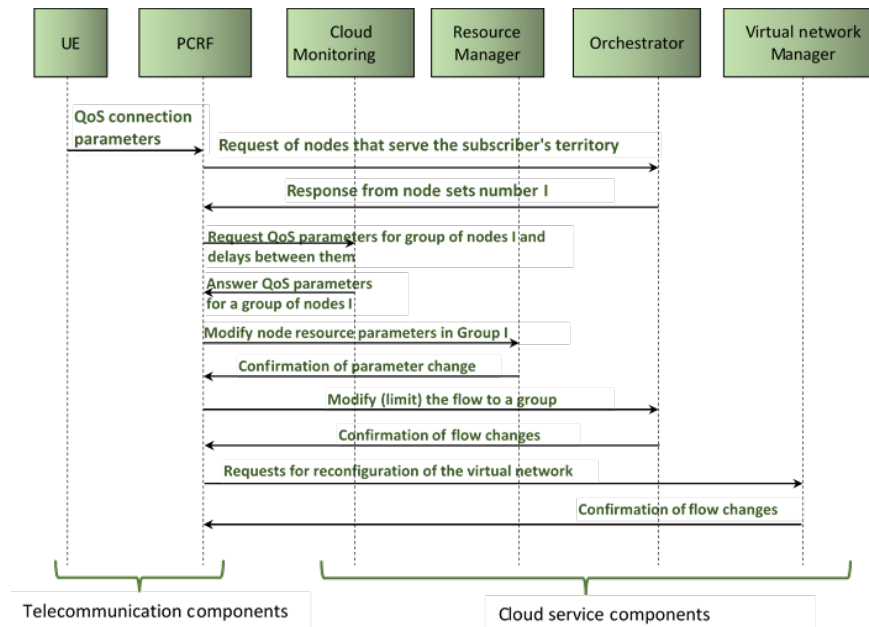
- The "Policy Management" subsystem creates a set of requirements for implementing a set of policies in relation to different flows of management.
- The "Policy Server" subsystem detects a problem of inconsistency of the current quality metrics with the declared subscription service policies.
- In the "Application Server" subsystem, program modules in which calculations are performed according to the proposed methods are implemented. The source data for the methods is the statistics obtained from the monitoring system and policy data that is provided to respective subscribers.
- The "Subscriber Data Store" subsystem is supplemented by information about virtual nodes, or a separate virtual network maintenance statistics database is created. This database collects information about service requests flows; the statistics of the relative dependence of the service intensity on service resources for each type of request.

The principle of dynamic quality control requires new procedures: it is necessary to arrange the interaction of mobile communication management system with virtualized resources management system (Fig. 11).

The quality control of management procedures implementation is evaluated at the level of User Equipment:

The User Equipment records the time delay in execution of service procedures, namely the time from the moment of connection initialization to the moment of data transmission beginning, and transfers to the subsystem of PCRf.

The PCRf receives this information from the subscriber and analyzes the policy server; in the policy implementation sub-system it compares the received data to the correspondence of chosen subscriber policy that is stored in the "Subscriber data store".



**Fig. 11.** Interaction of the mobile communication control system and the virtualized environment management system

If the delay values are not in accordance with the policy, PCRF requests the "Orchestrator" subsystem to identify the group of nodes  $i$  that serve the subscriber.

Orchestrator sends the numbers of nodes serving the subscriber, located in a given area. PCRF sends request to "Cloud Monitoring" for information on the delay and loss parameters in the nodes  $i$ , and information on the delay between nodes services. The Cloud Monitoring collects information regarding the latency and loss performance of hybrid services that are served on the nodes of the virtual network. The data about the service node group is transferred to the PCRF, where the principle of dynamic quality control of the service of hybrid services is realized. According to the management decisions, the PCRF subsystem sends inquiries:

- for reconfiguration of the virtual network, to the "Virtual Network Manager";
- to reconfigure resources to "Resource Manager";
- to change flows of service to "Orchestrator" streams over a virtual network.

When implementing the principle of dynamic quality control, most subsystems of the PCRF system are involved.

On the basis of a large amount monitoring system data, it is necessary to develop service system configurations that will meet the requirements for the quality of the maintenance process. On the basis of configurations and statistics, develop a schedule of resource use and periodically check the sufficiency of resources (Fig. 12).

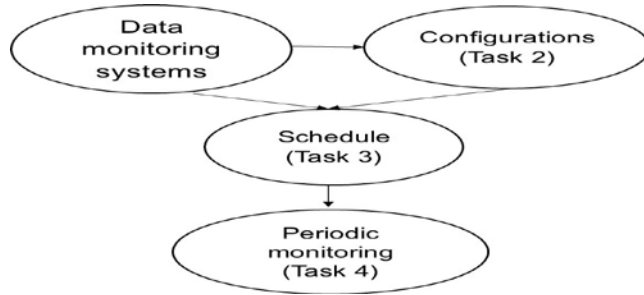


Fig. 12. The principle of forecasting the resource

The main factors influencing the forecast are the unsteadiness of subscriber activity during the day and the possibility of flexible configuration of resources when using the cloud computing environment.

*The method of constructing a schedule for choosing configurations based on long-term statistics* is based on a large number of statistical data collected over a long period of time (for example, during the year). These data are grouped on different days of the week. It analyzes the long-term load statistics created by the access node group if this method is used to plan the configuration of the service node groups or the load on the service node if the method is used to plan node resource allocation configurations.

The method solves the problem of determining the order of application of service resource configurations, namely the switching points. The basic scientific idea of the method is to use iterative procedures for define the configurations in different time intervals and to check the inequality of Chebyshev to meet the requirements regarding the adequacy of service resources throughout the time line.

**Input data:**

- $\lambda_k$  – maximum service intensity that equals to the  $k$ -th configuration;
- $\vec{\lambda}_t$  – vectors of statistical values of the requests quantity,  $t \in [0, T]$ ;
- $\vec{\lambda}_t = (\lambda_t^1, \lambda_t^2, \dots, \lambda_t^{50})$  – this is a set amount values of the requests, which were received at the  $i$ -th moment time, during the last 50 weeks.

**Output data:**

- $(k, t)$ , where  $t \in [0, T]$  – the recommended moments of servicing intensity change,  $k$  – the recommended configuration of the system, on which should switch to at moment time  $t$ .

**Algorithm of the method:**

*Step 1.* Split the time axis into time intervals of duration  $\Delta t$ ;

*Step 2.* For each interval  $\forall i = \overline{1, \frac{24}{\Delta t}}$  to find the average value of  $M\lambda$  and the dispersion  $\sigma^2$ . For the analysis takes a set of data collected over a long servicing time;

*Step 3.* For each interval to find the minimum configuration  $k$ , for which the following inequality is satisfied:

$$M\lambda + 3\sigma^2 \leq \lambda_k \quad (16)$$

create a temporary set  $Kt = \{Kt_i\}$ ,  $Kt_i = (k, i)$ ,  $i = \overline{1, T}$ ;

*Step 4.* Create a constant set  $Kt'$ , if  $Kt' = \emptyset$ , then  $Kt' := Kt$ , *Step 5*; else, compare the elements of the sets  $Kt'$  and  $Kt$  in pairs, if  $\forall Kt'_i$  executed  $Kt'_i = Kt_i$ , then *Step 6*, else  $Kt'_i = \max(Kt'_i, Kt_i)$ , *Step 5*;

*Step 5.* Change the value  $\Delta t$ , go to *Step 2*;

*Step 6.* Based on  $Kt'_i$  to define  $(k, t)$ , where  $t \in [0, T]$  – the recommended moments of servicing intensity change,  $k$  – the recommended configuration of the system, on which should switch to at moment time  $t$ .

#### **Short-term load predicting method**

The short-term load predicting method is improved on the based method ARIMA (Autoregressive integrated moving average) – auto-regression method with sliding mathematical expectation. However, unlike the known method, it is proposed to solve the problem of finding the minimum sliding interval, the use of which will satisfy the requirements, which will minimize the number of floating point operations to perform the prediction, which will ensure the optimal speed of prediction.

The proposed method consists of two stages: the calculation of the prediction interval based on the statistics of the operation of this servicing node and directly periodic forecasting of the load and the control of the sufficiency resources.

#### **Input data:**

- $T_p$  – the time interval for which the prediction is needed;
- $\lambda_i$  – number of requests per 1 ms, ( $i \in 0, \dots, N$ ),  $N = T_{inf} / 1 \text{ ms}$ ,  $\lambda_i \in \Lambda$ ,  $\Lambda$  – the set of requests number statistics values received during the time  $T_{inf}$  (first set, and then adjusted to the 2nd step of the method) before the start of the prediction,  $|\Lambda| = N$ ;
- $T_{pred}$  – the prediction period, the time after which the prediction algorithm is started;
- $M$  – the limit of the requests quantity that can be serviced with a given configuration of the servicing device;
- $P$  – likelihood of a prediction error.

Short-term statistics are collected locally at the servicing device and stored no longer than  $T_{inf} + T_{pred}$ , the interval of sampling is 1 ms.

#### **Output data:**

- $z \in \{0, 1\}$  –  $z = 0$  do not change the configuration;  $z = 1$  change the configuration.

#### **Algorithm of the method:**

Preparatory stage. The tuition of system based on statistical data. The search of the minimum  $T_{inf}$  (interval time for data collection)

$$T_{inf} \rightarrow \min, \quad (17)$$

for which the limitation is accomplished:

$$\bar{\lambda}_{T_{\text{pred}}} + 3\sigma > M, \quad (18)$$

$\bar{\lambda}_{T_{\text{pred}}}$  and  $\sigma$  – is calculated according to the main stage.

The limitation is accomplished for  $p=P*100\%$  of statistical samples obtained at different time intervals.

Solution: check values for the sequence established on the principle of  $T_{\text{inf}}^{k+1} = T_{\text{inf}}^k + \Delta$ ;  $T_{\text{inf}}^0 = T_{\text{pred}}$ .

The main stage of dynamic control.

1. To analyse the statistical data  $\lambda_i$  for the interval  $T_{\text{inf}}$  time preceding the moment of calculation. Construction by the least square method of coefficients  $\hat{a}$  and  $\hat{b}$

$$\lambda = ai + b. \quad (19)$$

2. To calculate  $\bar{\lambda}_{T_{\text{pred}}} = \hat{a}T_{\text{pred}} + \hat{b}$ .

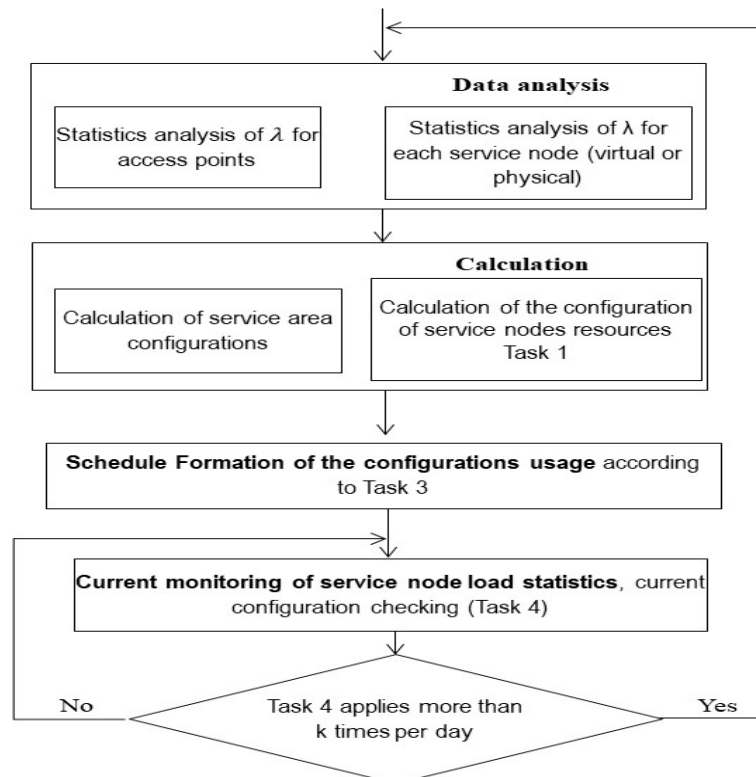
3. If  $\bar{\lambda}_{T_{\text{pred}}} + 3\sigma \leq M$ , then  $z = 0$ , else  $z = 1$ .

Planning of HTE resources is based on the analysis of a large number of statistical data. It allows you to calculate servicing configurations and a plan for switching between them during the operation of the control system.

The main stages in planning of the HTE resources are:

- data analysis;
- calculation of configurations;
- calculation of the plan for the use of configurations that ensure the maintenance of the use of resources NFV within the specified limits;
- dynamic control of the sufficiency resources, according to short-term statistics.

Data analysis involves the construction of data flow models for each access point, load statistics model for each servicing node (both virtual and physical), analysis of the flows map over the network (Fig. 13.).



**Fig. 13.** The algorithm of planning of the HTE resources on the basis of statistical data analysis

The calculation of the configurations consists of the calculation of the maintenance zones and the calculation of service resources, according to the statistics of loads on the nodes, using the results of task 1.

The calculation of service area configurations involves identifying groups of access points that has load statistics, and determining the required configuration of the nodes for servicing the virtualized (or physical) kernel system of the telecom operator.

Then by the method of constructing a schedule on the basis of long-term statistics (task 3), the graph of the use of calculating system configurations is determined.

Further on the method of planning a schedule based on long-term statistics (task 3), the schedule for using the calculating configurations of the system is determined.

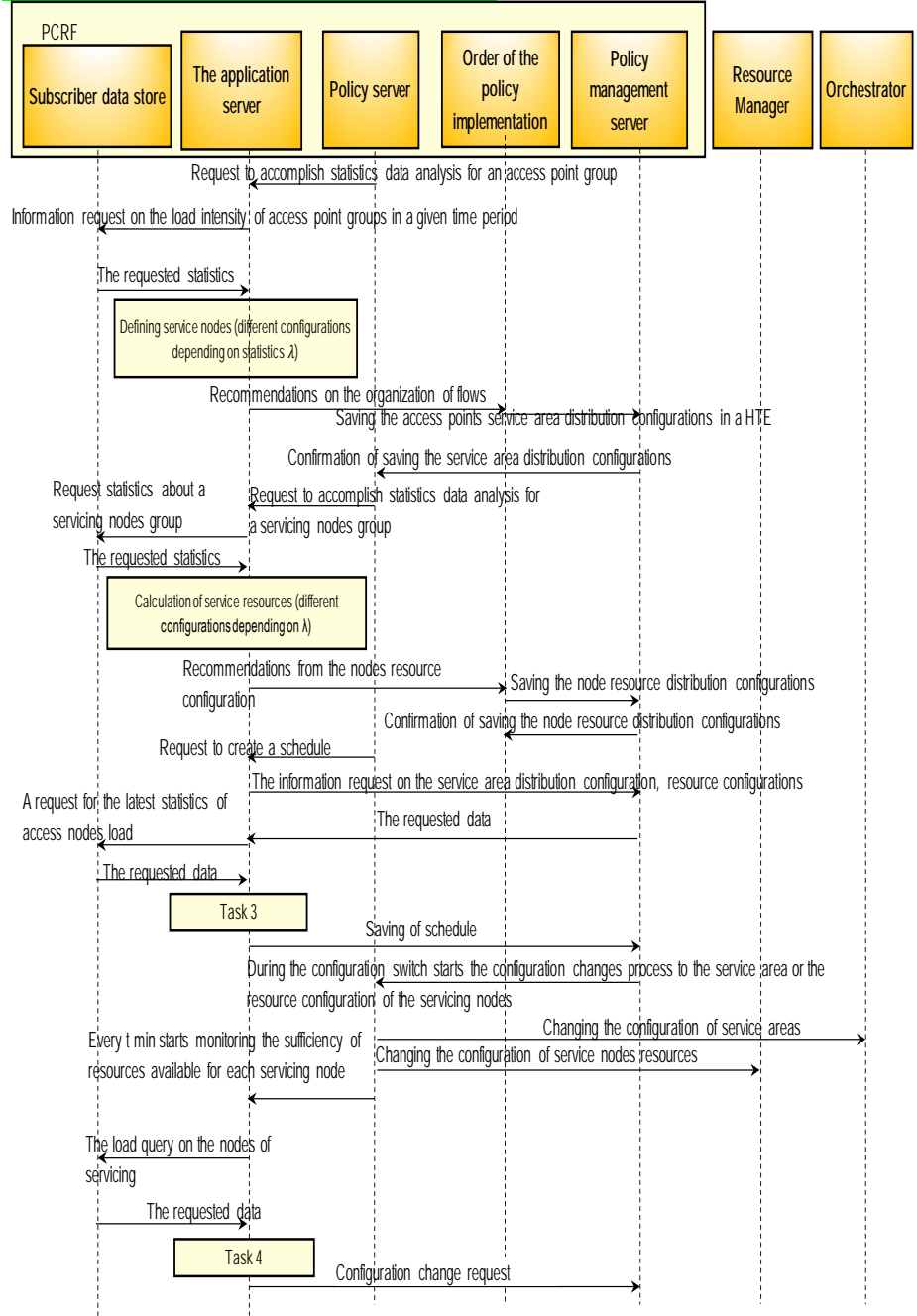
To ensure control of overload at service nodes, the sufficiency of resources for the next period of time is monitored on the basis of short-term statistics (task 4).

In fig. 14 shows the basic procedures for resource planning. The diagram shows which systems are involved for monitoring and planning resources.

Basic operations are performed in the PCRF system when interoperating between its subsystems namely between "Subscriber data store", "The application Server", "Policy Server", "Order of the policy implementation", and a heterogeneous environment management system - "Resource Manager" and "Orchestrator".

Virtual nodes data collection also uses "Virtual entity monitoring" - system that also refers to a heterogeneous environment management system, but is not included in

the description of the procedure. It is assumed that the long-term statistics of virtual nodes are stored in the "Subscriber data store"



**Fig. 14.** Procedure for planning the volume of resources of a heterogeneous telecommunication environment



Based on the proposed principles, models and methods, a methodology for providing high-quality service of hybrid telecommunication services was proposed, which solves the problem of the lack of tools for monitoring the quality of service in heterogeneous telecommunication environment (HTE) with the virtualization of network functions. In HTE, the quality of service users depends not only on the organization of the telecommunications network, but also on the organization of computing and interaction between the telecommunications and computing environment.

The work systematized all parameters of the HTE and allocated groups:

- group parameters which are changing dynamically, the statistics of which are fixed by monitoring systems cloud service system and telecommunication system;
- groups of sets parameter marginal values specified by standards and specifications and stored on the server by service policies in the PCRF system;
- group of rarely changing parameters;
- group sets of parameters that are calculated by known methods based on the group of parameters described above.

In fig. 16 the principle of methodology for providing high-quality service of hybrid telecommunication services is given.

*Group of dynamically changing parameters:*

- $V=\{V\}$  – set of nodes available to service;
- $E=\{e_{ij}\}$  – the matrix of node connectivity at the studied time;
- $A = \{\lambda_{ij}\}$  – the set of the estimated load of the  $i$ -th service node;
- $R = \{R_{ij}\}$  – matrix of available resources
- $N = \{n_{ik}\}$  – set of requests number that simultaneously are serviced at the  $i$ -th node

- $M = \{\mu_i\}$  – set of the service intensities at node.

*Group of sets of parameters limit values:*

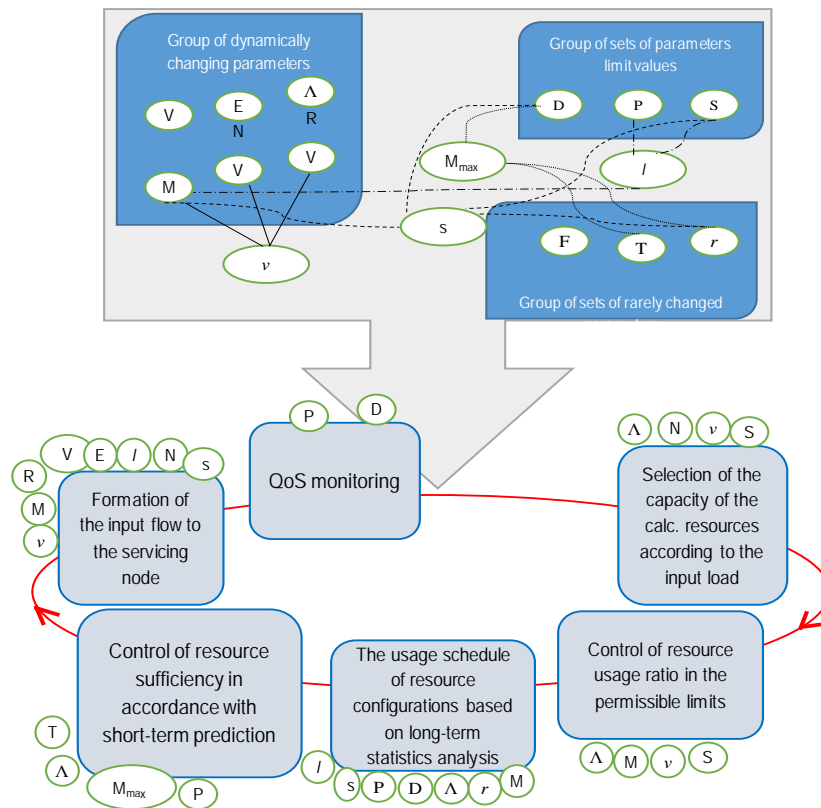
- $D = \{d_{ij}\}$  – set of the delay limit values in the node for various types of services;
- $P = \{P_i\}$  – set of limit values for indicator of packet loss for different types of services,
- $S = \{(S_{min\ i}, S_{max\ i})\}$  – a set of pairs of boundary values of the ratio of optimal use of resources on  $i$ -th node.

*Group of rarely changed parameters:*

- $r = \{k_{ij}\}$  – the possible configurations matrix of node resources, where the  $i$  – node number,  $j$  – number of the configuration
- $T$  – set of the time auxiliary parameters used to build the statistical samples and the formation of the schedule.

*A group of sets of calculated parameters that are found by known methods:*

- $M_{max}$  – set of number of requests that can be served in a node,
- $s$  – set of permissible queues lengths for different types of traffic
- $l$  – a set of limit values for the number of requests in the queue for which the not applicable method of early system overload
- $v = \{v_{ij}\}$  – the set of the volume of the  $j$ -th type resource that is used to service a single request in the  $i$ -th node.



**Fig. 16.** Principle of methodology for providing high quality service of hybrid telecommunication services

A number of simulation models were presented that implemented the models and methods proposed in the dissertation, on the basis of which the estimation of the quality of service of hybrid telecommunication services was calculated in accordance with the proposed methodology. Service quality assessment was performed on a set of statistical data that was received from the company of the communication operator. The evaluation process proceeded as follows, used modeling the set of kernel nodes of the communication operator in the GPSS system.

*Input data of the simulation model:*

$dt = 0,1$  ms – interval for time sampling;

$T_{simul} = 864000$  intervals – total simulation time (discrete);

$N_{\Sigma}^i$  – the number of requests that came in the system for time  $i$ -th simulation,  $i = 1, 100$ ;

$M = \begin{cases} 28, & \text{when using additional resources} \\ 14, & \text{else} \end{cases}$ ;

$V_j = \{V_{R1j}, V_{R2j}\}$  – available resources of the node  $j$ .

For each simulation, the statistics of servicing quality indicators for each  $j$ -th node, as well as the system as a whole, were recorded.

$t_j^i$  – average delay time at node  $j$  ( $j = \overline{1, M}$ ), during the  $i$ -th simulation  $i = \overline{1, 100}$ ;  
 $z_j^i$  – number of lost queries in node  $j$  ( $j = \overline{1, M}$ ), during the  $i$ -th simulation  $i = \overline{1, 100}$ ;  
 $t_{all}^i$  – the average delay time in the system during the  $i$ -th simulation  $i = \overline{1, 100}$ ;

$z_{all}^i$  – the number of lost queries in the system during the  $i$ -th simulation  $i = \overline{1, 100}$ .

Also, the monitoring system keep statistic data on the number of resources in each small interval of time:

$R_{1j}^i = \{R_{1j}^1, \dots, R_{1j}^k, \dots, R_{1j}^{T_{\text{MOD}}}\}$  – a set data monitoring of resource R1 in node  $j$  ( $j = \overline{1, M}$ ), during the  $i$ -th simulation  $i = \overline{1, 100}$ ;

$R_{2j}^i = \{R_{2j}^1, \dots, R_{2j}^k, \dots, R_{2j}^{T_{\text{MOD}}}\}$  – monitoring of resource R2 in node  $j$  ( $j = \overline{1, M}$ ), during the  $i$ -th simulation  $i = \overline{1, 100}$ .

To assess the performance of hybrid services, the calculation of the likelihood of violating the requirements of standards and specifications regarding the service time and the probability of timely service of the service was performed, the corresponding formulas for calculating probabilities were given in Table 1. The simulation results are summarized in Table 2.

In the work of the communication operator, an important indicator of the functioning of the system as a whole is the utilization rate of resources. The practice of the telecommunication company has shown that the resource utilization rate should vary from 30% to 80%. Since, if the resource utilization rate exceeds 80%, start unforeseen crashes, if the resource utilization rate is less than 30%, then not used effectively hardware and arises surplus of their maintenance costs are recorded. Therefore, in the simulation process, the probability that the system resources are used less than a given threshold value  $a$ , as well as the probability that the system resources are used more than a given threshold value  $b$

$$\begin{aligned} R_{2j}^i \supset R_{2j\_a}^i &= \{R_{2j}^k | R_{2j}^k < a * V_{R2j}; k = \overline{1, T_{\text{simul}}}\} & |R_{2j\_a}^i| &= A^i & (20) \\ R_{2j}^i \supset R_{2j\_b}^i &= \{R_{2j}^k | R_{2j}^k > b * V_{R2j}; k = \overline{1, T_{\text{simul}}}\} & |R_{2j\_b}^i| &= B^i & (21) \end{aligned}$$

Table 1. – Quality scores and appropriate quality scores

Quality score	Threshold value	Score	Values of evaluation
$t_d$ - delay time	$P_{tj}=0.8 \text{ ms}$	$p_{1j}$ $\forall j = \overline{1, M}$	$p_{1j} = 1 - (\sum_{i=1}^{100} k_{ij}) / 100$ $k_{ij} = \begin{cases} 1 & t_j^i > P_{tj} \\ 0 & \text{else} \end{cases}$
	$P_{tall}=8 \text{ ms}$	$p_{1all}$	$p_{1all} = 1 - (\sum_{i=1}^{100} K_{iall}) / 100$ $K_{iall} = \begin{cases} 1 & t_{all}^i > P_{tall} \\ 0 & \text{else} \end{cases}$
$P$ – probability of successful service	$P_{zi}=0.98$	$p_{2j}$ $\forall j = \overline{1, M}$	$p_{2j} = 1 - (\sum_{i=1}^{100} K_{iz}) / 100$ $K_{iz} = \begin{cases} 1 & \frac{N_{\Sigma}^i - z_j^i}{N_{\Sigma}^i} > P_{zi} \\ 0 & \text{else} \end{cases}$

	$P_{zall}=0,98$	$p_{2all}$	$p_{2all} = 1 - (\sum_{i=1}^{100} K_{i2all}) / 100$ $= \begin{cases} 1 & \frac{N_{\Sigma}^i - z_{all}^i}{N_{\Sigma}^i} > P_{zall} \\ 0 & else \end{cases}$
$\alpha$ – resource utilization rate HTE	a=0,3	$P_{3R_{2j}}$ $\forall j = 1, M$	$P_{3R_{2j}} = (\sum_{i=1}^{100} \frac{A_i}{T_{simul}}) / 100$
	b=0,8	$P_{4R_{2j}}$ $\forall j = 1, M$	$P_{4R_{2j}} = (\sum_{i=1}^{100} \frac{B_i}{simul}) / 100$

Table 2. – Simulation results

	Standard service	Management based on a methodology	Standard service	Management based on a methodology
	Average delay in servicing the service at the node ( $\bar{t}_j = \sum_{i=1}^{100} t_j^i / 100$ )		Assessment of timely service in the node ( $p_1 = (\sum_{j=1}^M p_{1j}) / M$ )	
Max <sub>j</sub>	7	9	0,8	0,805
Min <sub>j</sub>	1	1		
Average	3,8	4,4		
	Average service servicing delay over the system ( $\bar{t}_{all} = \sum_i t_{all}^i / 100$ )		Assessment of timely service ( $p_2 = p_{1all}$ )	
Max	80	90	0,82	0,84
Min	20	23		
Average	55	62		
	Probability of successful servicing in the node ( $\bar{z}_j = \sum_i \frac{N_{\Sigma}^i - z_j^i}{N_{\Sigma}^i} / 100$ )		Estimation of the probability of successful service in the node ( $p_3 = (\sum_{j=1}^M p_{2j}) / M$ )	
Max <sub>j</sub>	1	1	0,95	0,99
Min <sub>j</sub>	0,96	0,99		
Average	0,98	0,999		
	Probability of successful service in the system ( $\bar{z}_{all} = \sum_i z_{all}^i / 100$ )		Estimation of the probability of successful service in the system ( $p_4 = p_{2all}$ )	
Max <sub>j</sub>	1	1	0,94	0,99
Min <sub>j</sub>	0,95	0,99		
Average	0,975	0,999		
	Resource utilization rate HTE $\alpha$		Probability of using the resource R1 and R2 with a coefficient of utilization of computing resources less than a	

			given threshold value ( $\overline{P_{3R1j}}$ i $\overline{P_{3R2j}}$ )	
$(\overline{Max_j R_{1j}}) / V_{R2j}$	0.95	0.9	0.4 i 0.35	0.15 i 0.25
$(\overline{Min_j R_{1j}}) / V_{R2j}$	0.15	0.25	Probability of using the resource R1 i R2 with the coefficient of computing resources more than a given threshold value ( $\overline{P_{4R1j}}$ , $\overline{P_{4R2j}}$ )	
$(\overline{Max_j R_{2j}}) / V_{R2j}$	1	0.85	0.2 i 0.15	0.05 i 0.1
$(\overline{Min_j R_{2j}}) / V_{R2j}$	0.1	0.2	Estimation of the coefficient of utilization of computing resources used within the norm $p_5 = 1 - \frac{\sum_{g=1}^2 w_g p_{3Rgj}}{2} - \frac{\sum_{g=1}^2 w_g p_{4Rgj}}{2}$	
Average	0.4	0.2	0.43	0.75

Table 2 shows that such indicators of quality of service as the average delay in servicing the service at the node and in the system as a whole, the probability of successful servicing of the service in the node and the system as a whole were kept within the limits of acceptable values and acquired a slight improvement. However, the utilization rate of resources for servicing according to the proposed models and methods is kept within the specified limits with an average probability of 32%.

### 5 Conclusions

An approach to managing a heterogeneous telecommunication environment for increasing the efficiency of the service process of hybrid telecommunication services in new generation systems is proposed. A unified solution for telecommunication systems, where the maintenance of hybrid telecommunication services is carried out with the use of software is proposed. This approach allows to avoid reducing the quality of service during dash of overload and to maintain quality of service indicators at a given level, subject to compliance the resource utilization rate within the specified limits.

The modification of PCRF subsystems and new procedures for organizing the interaction of the mobile telecommunication network subsystems and the virtualized environmental management subsystems is proposed. It provides a process for monitoring the quality of service of hybrid telecommunication streams in the telecommunication environment, which allow providing the quality of service control and planning the amount of service resources for the efficient operation of heterogeneous telecommunication environment.

### References

1. ITU-T Recommendation M.3371 of October 2016.
2. ETSI GS NFV 001 v.1.1.1 (10/2013).

3. ETSI GS NFV 001 v.1.1.1 (10/2013).
4. Skulysh M. Romonov O. The structure of a mobile provider network with network functions virtualization // TCSET 2018 : 14-th International conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, 20–24 February 2018 : conference proceedings. — Lviv–Slavske, 2018. — P. 1032 - 1034.
5. J. ITU-T Y.3520 Telecommunication standardization sector of ITU (06/2013) / Series Y: Global information infrastructure, internet protocol aspects and next-generation networks.
6. Skulysh M., Klimovych O. Approach to virtualization of Evolved Packet Core Network Functions // Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 2015 13th International Conference The. – IEEE, 2015. – C. 193-195.
7. Larisa G., Mariia S., Andriy R. Control strategy of the input stream on the online charging system in peak load moments // Microwave & Telecommunication Technology (CriMiCo), 2014 24th International Crimean Conference. – IEEE, 2014. – C. 312-313.
8. Mariia Skulysh. The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters // Radio Electronics & Info Communications (UkrMiCo), 2017 International Conference.
9. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking Pages: 1 - 5, DOI: 10.1109/BlackSeaCom.2016.7901546.
10. O.I. Romanov, M.M. Nesterenko, L.A. Veres. IMS: Model and calculation method of telecommunication network's capacity // Proceedings of the 2017 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo) 11-15 Sept. 2017 Year, Odessa, Ukraine. - IEEE Conference Publications (IEEE Xplore Digital Library, DOI:10.1109/UkrMiCo.2017.8095412), 2017. – P. 1-4
11. Romanov, O.I., Oryschuk, M.V., Hordashnyk, Y.S. Computing of influence of stimulated Raman scattering in DWDM telecommunication systems. 2016 IEEE International Scientific Conference "Radio Electronics and Info Communications", UkrMiCo 2016 - Conference Proceedings
12. Skulysh, M.A., Romanov, O.I., Globa, L.S., Husyeva, I.I. Managing the Process of Servicing Hybrid Telecommunications Services. Quality Control and Interaction Procedure of Service Subsystems. *Advances in Intelligent Systems and Computing*. Volume 889, 2019, Pages 244-256.
13. V. V. Kurdecha, N. A. Zingeva, "Optimal reconfigurable base stations (R-BS) architecture and requirements to R-BS," 2011 21st International Crimean Conference "Microwave & Telecommunication Technology", Sevastopol, 2011, pp. 465-466.
14. Ilchenko M., Uryvsky L., Moshynska A. Developing of Telecommunication Strategies based on the Scenarios of the Information Community. *Cybernetics and Systems Analysis* November 2017, Volume 53, Issue 6, pp. 905–913
15. Moshynska A., Osypchuk S., Pieshkin A., Shmihel B. The effect of the features of signalcode constructions forming on indicators of functionality and reliability of communication systems based on the 802.11 N/AC standards. *Journal «Sciences of Europe»* (ISSN 3162-2364), Praha, Czech Republic. № 26 (2018), Vol.2. – p.p. 38...47 <http://european-science.org/wp-content/uploads/2018/05/VOL-2-No-26-2018.pdf>
16. Burak Gorkemli, Sinan Tatlicioglu, A. Murat Tekalp, Erhan Lokman. Dynamic Control Plane for SDN at Scale. *IEEE Journal on Selected Areas in Communications* PP(99):1-1
17. Adel Nadjaran Toosi, Redowan Mahmud, Qinghua Chi, Rajkumar Buyya. Management and Orchestration of Network Slices in 5G, Fog, Edge and Clouds. *Preprint*

18. Victor Lira, Eduardo Tavares, Meuse Oliveira, etc Virtual network mapping considering energy consumption and availability DOI: 10.1007/s00607-018-0620-y