

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Кваліфікаційна наукова  
праця на правах рукопису

**МУХІ-АЛДІН ХАССАН МОХАМЕД**

УДК 621.391

**ДИСЕРТАЦІЯ**

**МОДЕЛІ ТА МЕТОДИ ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ОБСЛУГОВУВАННЯ У  
МЕРЕЖАХ З ПІДТРИМКОЮ ФУНКЦІЙ ВІРТУАЛІЗАЦІЇ**

Спеціальність: 05.12.02 – телекомунікаційні системи та мережі  
05 «Технічні науки»

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело



Мухі-Алді́н Ха́ссан Моха́мед

Науковий керівник: ХЛАПОНІН Юрій Іванович, доктор технічних наук,  
професор

Ідентичність всіх примірників дисертації засвідчую:

Учений секретар спеціалізованої вченої ради

/Р.С. Одарченко/

Київ – 2020

## АНОТАЦІЯ

*Мухі-Алді́н Хассан Мохамед*. Моделі та методи забезпечення якості обслуговування у мережах з підтримкою функцій віртуалізації. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук (доктора філософії) за спеціальністю 05.12.02 «Телекомунікаційні системи та мережі» (172 – Телекомунікації та радіотехніка). – Національний авіаційний університет, Київ, 2020.

Дисертаційна робота присвячена вирішенню актуальної науково-прикладної задачі, яка пов'язана з забезпеченням належного рівня якості обслуговування у мережах з підтримкою функцій віртуалізації (network virtualfunction, NFV) за рахунок впровадження математичних моделей та методів аналізу та верифікації відповідності складових елементів архітектури віртуалізованих мереж.

У результаті проведеного в роботі аналізу встановлено, що розвиток та повноцінне функціонування технології віртуалізації мережевих функцій ускладнено рядом чинників. По-перше, не існує стандартизованих механізмів та правил побудови архітектури NFV: документація міжнародних консорціумів та інститутів стандартизації у галузі телекомунікацій носить здебільшого рекомендаційний узагальнений характер і не містить чітких вимог щодо впровадження та подальшої підтримки функціонування віртуалізованих мереж. По-друге, у більшості випадків інтерфейси мережевого обладнання з підтримкою технології NFV, що надається різними розробниками, наприклад, Cisco, HP, Juniper, Mikrotik, залишаються закритими та мають комерційне спрямування. По-третє, реалізація різних механізмів віртуалізації мережевої інфраструктури та використання різних систем управління та моніторингу призводить до виникнення функціональних розбіжностей та значного зниження якості обслуговування кінцевих користувачів.

Також встановлено, що формалізація вимог щодо функціонування мереж з підтримкою віртуалізації, своєчасний аналіз ефективності та коректності

розподілу мережевих ресурсів у процесі надання послуг за рахунок впровадження математичних моделей та методів дозволяє виявляти та усувати ряд розбіжностей та можливих помилок на ранніх етапах розробки мережевих рішень з підтримкою технології NFV. Розробка ж нових та вдосконалення існуючих моделей та методів формалізації та аналізу ефективності функціонування віртуалізованих мережах у процесі надання послуг дозволить усунути ряд критичних помилок і, як наслідок, покращити рівень якості обслуговування у телекомунікаційних мережах у цілому.

В основі методу формування комплексних сервісів покладено аналіз зміни значень показників QoS та їх вплив на сумарне значення функції ефективності результуючого комплексного сервісу. Математично обґрунтований вибір складових атомарних сервісів, що входять до складу комплексної та можливість та можливість їх динамічної заміни. Своєчасна заміна атомарних послуг у разі деградації QoS результуючого комплексного сервісу, що дозволяє підтримувати необхідний рівень якості протягом його надання.

При динамічному формуванні комплексних сервісів можуть бути зроблені модифікації або заміни атомарних сервісів, які дозволяють мережевій інфраструктурі адаптуватися до вимог кінцевого користувача і управляти функцією повернення в разі виявлення помилки. Іншими словами, склад комплексного сервісу може формуватися кожен раз при зверненні до нього користувача або компонент системи MANO. Метою даного підходу є досягнення високої гнучкості формування комплексних сервісів і поліпшення надійності композиції.

Удосконалений у даній роботі метод динамічного розподілу пропускної здатності каналів зв'язку базується на принципах розподілення залишкової пропускної спроможності з урахуванням коефіцієнту завантаження каналів та пріоритету сервісу, що надається. Встановлено, що у порівнянні з динамічним алгоритмом CentralQueuing описаний метод балансування навантаження дозволяє підвищити ефективність використання каналів передачі на 7% та

зменшити середню затримку загалом на 16%, а застосування розробленого методу реплікації сервісів дозволяє підвищити доступність сервісу в 14 разів у випадку перевантажень.

Для моделювання зміни мережевих характеристик в процесі надання послуг запропоновано застосування теорії мережевого обчислення, на основі якої складено ряд визначень, що дозволяють найбільш повно моделювати елементи NFVI. Запропоновано методи аналізу поведінки NFVI у двох часових моделях: безперервному часі: криві навантаження і обслуговування та дискретному часі: функція прибуття і відправлення. Розроблено алгоритм резервування ресурсів під час надання послуг, що базується на використанні елементів теорії мережевих обчислень. В основу методу покладено алгоритми аналізу та оцінки граничних показників затримки і швидкості обробки, що характерні для кожного вузла.

Розроблено метод динамічної реплікації сервісів в мережах з підтримкою функцій віртуалізації мережевих ресурсів, який за рахунок використання теорії мережевого обчислення дозволяє підвищити доступність і ефективність надання сервісів за рахунок своєчасного формування, видалення копій різних типів сервісів і їх раціонального розподілу між фрагментами мережі.

Стратегія динамічної реплікації або адаптивна стратегія реплікації дозволяє створювати репліки послуг, розподіляти і управляти ними на основі гнучких поведінкових шаблонів, що відповідають кожній копії. В цьому випадку рішення про формування і розміщення реплік засновані на зміні робочого навантаження серверів, їх доступності та продуктивності, кількості і локалізації запитів користувачів. У разі динамічної замість постійної підтримки копій на віддаленому обчислювальному вузлі, реплікація послуги виконується тільки тоді і в тому фрагменті мережі, де цього вимагають користувачі.

Підмножини даних, що належать популярній послугі, можуть бути збережені як на найближчий сервер, так і на локальний вузол в якості репліки. Таким чином, при використанні динамічної стратегії, репліки створюються і видаляються відповідно до зміни структури доступу.

У мережах, де мережеві ресурси є вузьким місцем (низька пропускна здатність каналів зв'язку), для зберігання реплік вибираються локальні фрагменти мережі, де дані репліки мають найбільшу популярність. Найчастіше в процесі динамічної реплікації формується рішення про необхідність створення репліки послуги і вибір її місця розташування, який ґрунтується на розрахунку відношення реальної кількості реплік, що зберігаються в даному фрагменті мережі, і кількості реплік, запитуваних у даний момент, а також аналізі мережевих характеристик сервера, на якому дані репліки повинні бути розміщені.

Одна з реалізацій стратегії динамічної реплікації запропонована в механізмі FIFO. В цьому випадку, планувальник Fair, з метою формування рішення щодо вибору місця розташування реплік послуги, які дозволять скоротити витрати мережевих ресурсів, використовує ряд певних критеріїв. В даному випадку копія популярного блоку даних або послуги зберігається на локальному вузлі або сервері.

Удосконалено метод формування та управління складом комплексних сервісів в мережах з підтримкою функцій віртуалізації шляхом урахування поточного стану ресурсів мережі під час формування і управління складом сервісів, що дозволяє забезпечити необхідний рівень якості обслуговування у відповідності до погодженого договору про якість обслуговування. Розподілений комплексний сервіс може бути представлений у вигляді об'єднання безлічі сервісів  $P_i$ . При цьому, QoS параметри атомарних сервісів  $P_i$  можуть мати значення нижче необхідних, проте, за рахунок їх одночасного використання, сумарні показники продуктивності розподіленого сервісу зростають.

Метод формування розподіленого сервісу дозволяє сформувати сервіс з заданими значеннями параметрів якості у випадку, коли відсутній сервіс з необхідними показниками якості в мережі з підтримкою технології віртуалізації.

Удосконалено метод балансування навантаження в мережі з підтримкою віртуалізації мережевих ресурсів за рахунок ранжування сервісів по якості їх надання та необхідній пропускній здатності та послідовного перерозподілу наявних вільних мережевих ресурсів, що дозволяє на основі аналізу поточного навантаження як на віртуальну, так і на фізичну складову мультисервісної мережі забезпечити заданий рівень якості обслуговування з меншими накладними витратами.

Оптимального розподілу пропускної здатності можна досягнути, відповідно до запропонованого методу, шляхом розподілу потоків трафіку всіма допустимими каналами зв'язку пропорційно доступним пропускним здібностям цих каналів в поточний момент часу або у момент часу надходження заявки.

Передбачається, що для кожного потоку доступні  $n$  каналів зв'язку. Вибір каналу зв'язку ґрунтується на значенні його пропускної здатності. Вибір політики розподілу пропускних здатностей каналів зв'язку залежить від початкових QoS вимог до потоку даних (пріоритетність запиту).

Вдосконалення методу розподілу пропускної здатності полягає в первинному виборі шляху максимальної пропускної здатності для пріоритетного потоку і шляхів, що доповнюють, для розподілу непріоритетних потоків. В принципі функціонування запропонованого алгоритму покладено алгоритми Едмондс-Карпа і максимізації пропускної здатності з деякими припущеннями.

Метод формування розподіленого сервісу дозволяє сформуванню сервісу з заданими значеннями параметрів якості у випадку, коли відсутній сервіс з необхідними показниками якості.

**Ключові слова:** віртуалізація мережевих функцій, система управління та оркестровки, якість обслуговування, комплексний сервіс, балансування навантаження.

## ABSTRACT

*Muhi-Aldin Hassan Mohamed.* Models and methods for ensuring quality of service in networks with function virtualization – Qualifying research work as a manuscript.

Dissertation for the degree of Candidate of Technical Sciences (Doctor of Philosophy) in the specialty 05.12.02 – Telecommunication Systems and Networks. – (172 – Telecommunications and radio engineering). – National Aviation University, Kyiv, 2019.

The dissertation is devoted to the solution of a relevant scientific and applied problem, which is related to ensuring the proper level of service quality in virtualized networks by introducing mathematical models and methods for analyzing and verifying the consistency of the architecture components of virtualized networks.

As a result of the analysis carried out in the work, it was established that the development and full functioning of virtualization technology of network functions is complicated by a few factors. First, there are no standardized mechanisms and rules for constructing the NFV architecture: the documentation of international consortia and telecommunication standardization institutes is mostly of a general recommendation nature and does not contain clear requirements for the implementation and further support of the functioning of virtualized networks. Secondly, in most cases, NFV network equipment interfaces provided by different developers, such as Cisco, HP, Juniper, Mikrotik, remain closed and commercially oriented. Thirdly, the implementation of various mechanisms for virtualization of the network infrastructure and the use of different management and monitoring systems leads to functional differences and a significant reduction in the quality of service of end users.

It has also been established that the formalization of requirements for the functioning of virtualized networks, timely analysis of the efficiency and correctness of network resources allocation in the process of providing services through the introduction of mathematical models and methods allows timely identifying and eliminating a number of differences and possible errors at the early stages of

development of network solutions with the support for NFV technology. The development of new and improved existing models and methods of formalization and analysis of the effectiveness of virtualized networks in the process of providing services will eliminate a number of critical errors and, consequently, improve the quality of service in telecommunication networks in general.

The basis of the method of formation of complex services is the analysis of changes in the values of QoS indicators and their impact on the total value of the function of the efficiency of the resulting comprehensive service. Mathematically based choice of components of atomic services, which are part of the complex and the possibility and possibility of their dynamic replacement. Timely replacement of atomic services in case of degradation of QoS resulting comprehensive service, which allows maintaining the required level of quality during its delivery.

At dynamic formation of complex services modifications or replacements of atomic services can be made, which allow the network infrastructure to adapt to the requirements of the end user and control the return function in case of an error. In other words, the composition of a comprehensive service may be formed each time a user or component of the MANO system accesses it. The purpose of this approach is to achieve the high flexibility of forming complex services and to improve the reliability of the composition.

The method of dynamic allocation of bandwidth of the communication channels proposed in the work is based on the principles of distribution of residual bandwidth taking into account the load factor of channels and the priority of the provided service. It has been determined that the proposed load balancing method can increase the efficiency of the use of transmission channels by 7% and reduce the average delay by 16% in comparison with the dynamic Central Queuing algorithm, and the application of the developed service replication method can increase the service availability by 14 times in case of overloads.

To simulate the change of network characteristics in the process of providing services, the application of the theory of network calculus has been proposed, on the basis of which a number of definitions are made that allow the most complete



modeling of the NFVI elements. The methods of analysis of NFVI behavior in two time models are proposed: continuous time: arrival and service curves and discrete time: arrival and departure function. The resource allocation algorithm is developed during the provision of services, which is based on the use of elements of the theory of network calculus. The method is based on algorithms for analysis and evaluation of the boundary indicators of delay and processing rate, which are characteristic for each node.

The method of dynamic replication of services in networks with support of functions of virtualization of network resources is developed, which through the use of network computing theory allows to increase the availability and efficiency of rendering services by timely formation, deleting of copies of different types of services and their rational distribution between fragments of the network.

Dynamic Replication Strategy or Adaptive Replication Strategy lets you create, distribute, and manage service replicas based on flexible behavioral templates that match each copy []. In this case, the decisions on the formation and placement of replicas are based on changes in the server workload, their availability and performance, the number and location of user requests. In the case of dynamic instead of constant support of copies on the remote computing node, service replication is performed only then and in that part of the network where users require it.

The subsets of data belonging to a popular service can be stored either on the nearest server or on a local node as a replica. Thus, when using a dynamic strategy, replicas are created and deleted according to changes in the access structure [108, 109].

In networks where network resources are a bottleneck (low bandwidth of communication channels), local fragments of the network where these replicas are most popular are selected for storing replicas. Most often, in the process of dynamic replication, the decision is made about the need to create a replica of the service and the choice of its location, which is based on the ratio of the actual number of replicas stored in this fragment of the network, and the number of replicas currently

requested, as well as analysis of network characteristics of the server, on which the replica data should be placed.

One of the implementations of the dynamic replication strategy is proposed in the FIFO mechanism. In this case, Fair Scheduler uses a number of criteria to determine the location of replica services that will reduce the cost of network resources. In this case, a copy of a popular data or service block is stored on a local node or server.

The method of forming and managing the composition of complex services in networks with support for virtualization functions has been improved by taking into account the current state of network resources during the formation and management of the composition of services, which provides the necessary level of quality of service in accordance with the agreed service quality agreement.

Improved network load balancing method with support for virtualization of network resources due to the ranking of services by the quality of their provision and the necessary bandwidth and consistent redistribution of available free network resources, which allows, based on the analysis of current load on both virtual and physical components of the multiservice network set the level of quality of service with less overhead.

The optimal bandwidth allocation can be achieved, according to the proposed method, by distributing traffic flows by all admissible communication channels in proportion to the available bandwidth of these channels at the current time or time of receipt of the application.

It is assumed that  $n$  channels are available for each stream. The choice of a communication channel is based on the value of its bandwidth. The choice of bandwidth allocation policy depends on the initial QoS requirements for the data flow (request priority).

The proposed throughput allocation method is to initially select the maximum throughput path for priority flow and complementary paths for the distribution of non-priority flows. The principles of operation of the proposed algorithm are based

on the Edmonds-Carp algorithm and throughput maximization with some assumptions.

Distributed integrated service can be represented as an aggregation of multiple Pi services. At the same time, the QoS parameters of the atomic Pi services may have values below the required ones, however, due to their simultaneous use, the total performance of the distributed service increases. The method of forming a distributed service allows you to create a service with set values of quality parameters in the case when there is no service with the required quality indicators.

**Key words:** network functions virtualization, management and orchestration system, quality of service, complex service, load balancing.

*Список публікацій здобувача:*

1. Abdulghafoor Raed Yahya. Performance analysis of load balancing mechanisms in Software-Defined Networking / O.B.Tkachova, Abdulghafoor Raed Yahya, Hassan Mohamed Muhi-Aldeen // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2016.– №187. – P.197-203.

2. Мухи-Алдин Х.М. Комплексный метод повышения доступности сервисов в программно-конфигурируемых сетях, основанный на динамической репликации [Электронный ресурс] / Х.М. Мухи-Алдин, Е.Б. Ткачова, М.Т. Салах // Проблемы телекоммуникаций. – 2016. – № 2 (19). – С. 81 - 93. – Режим доступа к журн.: [http://pt.journal.kh.ua/2016/2/1/162\\_tkachova\\_rep.pdf](http://pt.journal.kh.ua/2016/2/1/162_tkachova_rep.pdf).

3. Ткачева Е.Б. Применение теории сетевого исчисления при моделировании программно-конфигурируемых сетей / Е.Б. Ткачева, Е.В. Дуравкин, Х.М. Мухи-Алдин // Системи управління, навігації та зв'язку. – 2017.– №6(46). – С. 136-141.

4. Ткачева Е.Б. Адаптивный алгоритм перераспределения сетевых ресурсов в сетях с поддержкой технологии NFV / Е.Б. Ткачева, Хассан Мохамед Мухи-Алдин // Радиотехника: Всеукр. межвед. науч.-техн. сб. – 2017. – № 191 – С. 135-141.

5. Hassan Mohamed Muhi-Aldeen. A method of modeling a behavior of NFV infrastructure based on network calculus theory / Hassan Mohamed Muhi-Aldeen // *Scholars Journal of Engineering and Technology (SJET) ISSN 2321-435X (Online) Sch. J. Eng. Tech.* – 2017. – Vol 4(11).– pp. 517-524.

6. Хассан Мохамед Мухи-Алдин. Методологія побудови кореляційних алгоритмів класифікація мережевого трафіку / Хассан Мохамед Мухи-Алдин // *Міжвузівський збірник "Наукові нотатки".*– 2018, – Вып.64, – С. 128-135.

7. Хассан Мохамед Мухи-Алдин. Комплексный метод динамической репликации данных в программно-конфигурируемых сетях / Хассан Мохамед Мухи-Алдин, Е.Б. Ткачева // *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*» – 2018. – Том 29 (68).– 2018. – Вып. 5 – С. 18-26.

8. Hassan Mohamed Muhi-Aldeen. A method of network resource allocation in the networks supporting NFV technology / Hassan Mohamed Muhi-Aldeen, О.В.Тkachova // *Науковий журнал "Комп'ютерно-інтегровані технології: освіта, наука, виробництво".* – 2018, – Вып. 33 – С.5-11.

9. Одарченко Р.С. Модель оцінки підвищення енергоефективності підсистем базових станцій оператора стільникового зв'язку. Р.С. Одарченко Мухі-Алді́н Хассан Мохамед, О.О. Полігенько, Ю. І Хлапонін. *Вісник інженерної академії*, 2019. № 3. С. 72-74.

10. Hassan Mohamed Muhi-Aldeen A network load balancing algorithm for overlay-based NFV solutions / Hassan Mohamed Muhi-Aldeen, Abdulghafoor Raed Yahya, О.В. Tkachova // *Second International IEEE Conference «Problems of Infocommunications. Science and Technology» PICS&T–2016, 4–7 October 2016: proc. of the conf.* – Kharkiv, Ukraine, 2016. – Pp.1-4.

11. Хлапонін Ю. І, Мухі-Алді́н Хассан Мохамед. Метод динамічної реплікації даних в мережах з підтримкою функцій віртуалізації. Сьома міжнародна науково-технічна конференція "Проблеми інформатизації", 12 грудня 2016 р., тези доп. – С. 94, м. Київ.

12. Hassan Mohamed Muhi-Aldeen. An evaluation of load distribution between servers' platforms inside a data center / Hassan Mohamed Muhi-Aldeen, Abdulghafoor Raed Yahya, O.B. Tkachova // 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 21-25 February, 2017 – P. 357-359.

13. Мухи-Алдин Хассан Мохамед. Модель оценки показателей качества обслуживания в облачных решениях // Информатика, математика, автоматика. Матеріали науково-технічної конференції, 17-21 квітня 2017 р. – Суми: Сумський державний університет, 2017. – С. 104.

14. Tkachova O. An Method of Service Composition with Optimal Resources Allocation in Software-Defined Networking / O.Tkachova, Ie. Duravkin, Muhi-Aldeen Hassan Mohamed // Forth International IEEE Conference «Problems of Infocommunications. Science and Technology» PICS&T–2017, 10–13 October 2017 proc. of the conf. – Kharkiv, Ukraine, 2017. – Pp.1-4.

15. Ткачева Е.Б., Мухи-Алдин Х.М. Анализ сценариев оркестровки сервисов в облачных технологиях / Е.Б. Ткачева, Мухи-Алдин Х.М. // Інформаційна безпека та комп'ютерні технології: Збірник тез доповідей II Міжнародної науково-практичної конференції, 20-22 квітня 2017 року, м. Кропивницький: ЦНТУ, 2017. – С. 173-175.

## ЗМІСТ

ВСТУП.....	17
<b>РОЗДІЛ 1</b>	
<b>АНАЛІЗ МЕТОДІВ УПРАВЛІННЯ СЕРВІСАМИ В МУЛЬТИСЕРВІСНИХ МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ ..</b>	<b>25</b>
1.1. Аналіз архітектури мультисервісних мереж з підтримкою віртуалізації мережевих функцій.....	32
1.2. Аналіз механізмів оркестрації послуг в мультисервісних мережах з підтримкою технології віртуалізації мережевих функцій.....	36
1.3. Аналіз методів забезпечення якості обслуговування в мультисервісних мережах з підтримкою віртуалізації мережевих функцій.....	42
1.4. Постановка наукового завдання і формулювання особистих завдань дослідження.....	48
1.5. Висновки до першого розділу.....	50
<b>РОЗДІЛ 2</b>	
<b>МОДЕЛІ ПРОЦЕСІВ НАДАННЯ СЕРВІСІВ У МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ.....</b>	<b>52</b>
2.1. Застосування елементів теорії мережевого обчислення, що застосовуються при моделюванні процесів надання сервісів.....	59
2.2. Моделювання процесу надання сервісів у мультисервісних мережах з підтримкою функцій віртуалізації за допомогою елементів теорії мережевого обчислення.....	65
2.3. Аналіз затримки «від краю до краю» в процесі надання сервісів у мультисервісних мережах з підтримкою функцій віртуалізації.....	70
2.4. Висновки до другого розділу.....	73
<b>РОЗДІЛ 3</b>	
<b>МОДЕЛІ І МЕТОДИ АНАЛІЗУ І УПРАВЛІННЯ ЯКІСТЮ СЕРВІСІВ У МЕРЕЖАХ З ВІРТУАЛІЗАЦІЄЮ ФУНКЦІЙ.....</b>	<b>74</b>

3.1. Оцінка якості надання комплексних сервісів в мережах з підтримкою віртуалізації функцій.....	74
3.2. Метод формування комплексних сервісів з урахуванням показників якості обслуговування.....	79
3.3. Метод динамічного розподілу пропускної здатності каналів зв'язку в мережах, що програмно конфігуруються .....	85
3.4. Оцінка ефективності запропонованого методу динамічного розподілу пропускної здатності каналів зв'язку.....	91
3.5. Висновки до третього розділу.....	96
<b>РОЗДІЛ 4</b>	
<b>РОЗРОБКА МЕТОДУ РЕПЛІКАЦІЇ СЕРВІСІВ.....</b>	<b>98</b>
4.1. Аналіз методів реплікації сервісів в мультисервісних мережах.....	98
4.2. Розробка стратегії динамічної реплікації сервісів.....	108
4.3. Визначення ключових показників якості в процесі формування репліки комплексного / атомарного сервісу.....	113
4.3.1. Визначення моментів часу формування реплік.....	115
4.3.2. Визначення кількості реплік і їх розташування в мережі.....	117
4.3.3. Визначення моменту прийняття рішення про видалення репліки сервісу.....	121
4.4. Висновки до четвертого розділу.....	125
Висновки до роботи.....	127
СПИСОК ЛІТЕРАТУРИ.....	129
ДОДАТОК А.....	144

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

BGP	Border Gateway Protocol
CPU	central processing unit
EPL	Ethernet Private Line
ETSI	European Telecommunications Standards Institute
ETSI	European Telecommunications Standards Institute
FIFO	First In First Out
FG	Forwarding Graph
IETF	Internet Engineering Task Force
ITU-T	International Telecommunication Union-Telecommunication
MANO	Management And Network Orchestration
MPLS	Multiprotocol Label switching
NFV	Network Function Virtualization
NFVI	Network functions virtualization infrastructure
OPERA	OPEn Replication Architecture
QoE	Quality of Experience
QoS	Quality of Service
RAM	Random Access Memory
RSVP	Resource ReSerVation Protocol
SDN	Software Defined Networking
MANO	Management And Orchestration
SLA	Service Level Agreement
SOA	Service Oriented Architecture
TCP	Transmission Control Protocol
ToS	Type of Service
UDDI	Universal Description Discovery & Integration
VIM	Virtualized Infrastructure Manager
VNF	Virtual function network
VTEP	Virtual Tunnel End Point
TC	Точки з'єднання



## ВСТУП

Вдосконалення методів забезпечення та підтримки гарантованого рівня якості обслуговування користувачів на сьогоднішній день виступає одним із основних напрямків розвитку телекомунікацій. Слід зазначити, що ускладнення протоколів управління і передачі, а, отже, і ускладнення мережевого обладнання, що є стримуючим фактором для подальшого розвитку традиційних телекомунікаційних систем, відбувається через стрімке зростання обсягів переданих даних і підвищення вимог до функціональних можливостей телекомунікаційних систем.

Для підвищення ефективності процесів надання послуг необхідно будувати та впроваджувати мережі з підтримкою технології віртуалізації мережевих функцій (Network Function Virtualization, NFV). Завдяки тому, що технологія NFV дозволяє програмно реалізувати широкий набір функцій і послуг, які на сьогоднішній день надаються лише мережевим обладнанням реалізованим апаратно (мережеві екрани, прикордонні маршрутизатори, комутатори, сервери доступу та інші) у вигляді відкритого програмного забезпечення. Таким чином, з використанням технології NFV будь-який вид послуги може бути оперативно впроваджений і наданий кінцевому користувачеві в потрібний час з заданим рівнем якості.

Однак розвиток і повноцінне функціонування технології віртуалізації мережевих функцій на сьогоднішній день ускладнений рядом факторів. По-перше, не існує стандартизованих механізмів і правил побудови архітектури NFV: документація ETSI і IETF носить рекомендаційний узагальнений характер і не містить чітких вимог до впровадження і подальшої підтримки NFV. По-друге, в більшості випадків інтерфейси мережевого обладнання з підтримкою технології NFV, що надається різними розробниками, наприклад, Cisco, HP, Juniper, Mikrotik, залишаються закритими і мають комерційний напрямок.

Формалізація вимог щодо функціонування віртуалізованих мереж, аналіз ефективності і оцінка коректності розподілу ресурсів мережі в процесі надання послуг за рахунок впровадження математичних моделей і методів дозволять

своєчасно виявляти і усувати ряд розбіжностей і можливих помилок на ранніх етапах розробки мережеских рішень з підтримкою функцій віртуалізації. Однак методи, що застосовуються на сьогоднішній день при розробці мережеских рішень, не дозволяють забезпечити належну якість послуг в мережах з підтримкою технології NFV: методи формалізації вимог специфікації, що набули широкого поширення під час розробки традиційних телекомунікаційних систем і мереж (SDL, UML та інші) не дозволяють повноцінно формалізувати причинно-наслідкові зв'язки між віртуалізованими мережескими компонентами; а недостатній розвиток методів аналізу і верифікації рішень NFV може призвести до виникнення ряду критичних помилок, що впливають на подальшу якість обслуговування кінцевих користувачів.

Таким чином, розробка нових і вдосконалення існуючих моделей і методів формалізації та аналізу ефективності функціонування віртуалізованих мереж в процесі надання послуг дозволить усунути ряд можливих помилок і, як наслідок, поліпшити рівень якості обслуговування.

На розвиток технології віртуалізації мережеских функцій, а саме завдання дослідження і вдосконалення методів розробки та впровадження механізмів формування та надання послуг з підтримкою гарантованої якості обслуговування, спрямована увага таких великих європейських телекомунікаційних провайдерів, як Alcatel-Lucent, AT & T, BT, Deutsche Telekom, Orange, Telecom Italia, а також міжнародних інститутів і консорціумів - ETSI, IRTF, IETF SFC WG, DMTF OVF. Методам підвищення якості обслуговування в мережах з підтримкою технології NFV, зокрема, розвитку методів підвищення ефективності розподілу мережеских ресурсів і коректності функціонування систем управління ними присвячені роботи дослідників AT & T (Б. Хана, В. Копалакрішна), Deutsche Telekom (В. Мішель, К. Мішель), Telecom Italia (Е. Демар, А. Піннола), Vodafone (С. Сабатер, А. Нено) і ін. Моделям і методам аналізу, розробки і впровадження NFV рішень також присвячений ряд робіт російських і українських учених - Захарова В.А., Смоленського Р.Л., Кучерявого А.Е., Чемеріцького Е.В. та ін.

Результат аналізу існуючих досліджень і практичних реалізацій мережевих рішень з підтримкою NFV дозволяє зробити висновок, що перспективним напрямком усунення існуючих недоліків, що виникають в процесі їх функціонування, є впровадження математичних моделей і методів аналізу та верифікації як заключних стадій кожного з етапів процесу розробки. Так, чітка формалізація вимог специфікації по впровадженню віртуальних мережевих функцій, математичні методи аналізу розподілу ресурсів та оцінки відповідності кінцевої реалізації NFV вимогам специфікації в процесі розробки дозволять підвищити доступність послуг і якість їх надання в кінцевій реалізації NFV рішень.

Виходячи з цього, науково-прикладна задача, яка полягає в розробці методів і моделей формування та надання комплексних послуг в мережах з підтримкою функцій віртуалізації, що дозволяють раціонально використовувати мережеві ресурси відповідно з поточним станом мережі і QoS вимог користувачів, а, отже, і тема дисертаційної роботи «Модель і методи забезпечення якості обслуговування в мережах з підтримкою функцій віртуалізації», яка спрямована на вирішення зазначеного завдання, є актуальною.

### **Мета роботи**

Метою роботи є забезпечення необхідного рівня якості обслуговування в мультисервісних мережах з підтримкою функцій віртуалізації.

Для досягнення поставленої мети в роботі запропоновані і вирішені наступні завдання:

1. провести аналіз архітектури, принципів побудови мультисервісних мереж з підтримкою функцій віртуалізації та методів управління ресурсами для забезпечення необхідного рівня якості обслуговування;
2. розробити метод динамічної реплікації сервісів в мережах з підтримкою функцій віртуалізації мережевих ресурсів;
3. удосконалити метод формування та управління складом комплексних сервісів в мережах з підтримкою функцій віртуалізації;

4. удосконалити метод балансування навантаження в мережі з підтримкою віртуалізації мережевих ресурсів;
5. удосконалити метод динамічного розподілу пропускнуої здатності каналів передачі даних під час надання комплексних сервісів;
6. провести дослідження працездатності та ефективності запропонованих в роботі рішень.

**Об'єкт дослідження:** процеси надання послуг в мультисервісних мережах з підтримкою функцій віртуалізації.

**Предмет дослідження:** моделі та методи забезпечення необхідного рівня якості обслуговування в мультисервісних мережах з підтримкою функцій віртуалізації.

**Методи дослідження.**

Під час вирішення поставлених завдань, зокрема, під час оцінки ефективності взаємодії елементів мережевої інфраструктури NFV в процесі надання комплексних сервісів були використані положення теорії управління багаторівневими системами; під час аналізу та складу комплексного сервісу використані положення теорії множин; при вирішенні завдання балансування навантаження в мережах з підтримкою функцій віртуалізації - теорія масового обслуговування та теорія графів; під час оцінки коректності та ефективності розподілу мережевих ресурсів - апарат E-мереж та методи структурного аналізу; під час проведення та оцінки результатів експериментального дослідження - методи імітаційного моделювання і математичної статистики; під час розробки методу оцінки показників якості надання послуг «від краю до краю» в інфраструктурі NFV - методи аналізу граничних значень показників якості, засновані на принципах теорії мережевого обчислення.

**Наукова новизна отриманих результатів.** В ході вирішення поставлених завдань отримані наступні нові наукові результати:

1. Розроблено метод динамічної реплікації сервісів в мережах з підтримкою функцій віртуалізації мережевих ресурсів, який за рахунок використання теорії мережевого обчислення дозволяє підвищити доступність і

ефективність надання сервісів за рахунок своєчасного формування, видалення копій різних типів сервісів і їх раціонального розподілу між фрагментами мережі.

2. Удосконалено метод формування та управління складом комплексних сервісів в мережах з підтримкою функцій віртуалізації шляхом урахування поточного стану ресурсів мережі під час формування і управління складом сервісів, що дозволяє забезпечити необхідний рівень якості обслуговування у відповідності до погодженого договору про якість обслуговування;

3. Удосконалено метод балансування навантаження в мережі з підтримкою віртуалізації мережевих ресурсів за рахунок ранжування сервісів по якості їх надання та необхідній пропускній здатності та послідовного перерозподілу наявних вільних мережевих ресурсів, що дозволяє на основі аналізу поточного навантаження як на віртуальну, так і на фізичну складову мультисервісної мережі забезпечити заданий рівень якості обслуговування з меншими накладними витратами.

4. Набув подальшого розвитку метод динамічного розподілу пропускної здатності каналів передачі даних у NFV мережах, що базується на принципах розподілення залишкової пропускної спроможності. Застосування даного методу дозволяє уникнути повного завантаження та перевантаження каналу передачі.

### **Обґрунтованість і достовірність результатів дослідження.**

Обґрунтованість і достовірність отриманих в роботі нових наукових результатів забезпечувалася і підтверджувалася коректним використанням ключових положень відомих і апробованих математичних методах: логіках вищого порядку, теорії управління багаторівневими системами, теорії графів, теорії масового обслуговування та мережевого обчислення, а також поширеними і широко вживаними методами верифікації.

### **Наукове і практичне значення отриманих результатів.**

Запропоновані в дисертаційному дослідженні математичні моделі і методи формалізації вимог специфікації, аналізу ефективності розподілу мережевих ресурсів, перевірки відповідності готового мережевого рішення з підтримкою технології NFV вимогам специфікації і оцінки якості надання послуг дозволяють забезпечити належний рівень якості обслуговування у віртуалізованих мережах.

Результати дисертаційної роботи можуть бути рекомендовані при проектуванні і вдосконаленні структури мереж з підтримкою технологій віртуалізації, в тому числі технології NFV. Запропоновані моделі і методи можуть бути використані як науково-методична база для подальших досліджень функціонування систем управління і оркестровки, а також при розробці та впровадженні механізмів віртуалізації в інфраструктурі NFV.

### **Практичне значення одержаних результатів досліджень.**

Практичне значення одержаних результатів досліджень полягає в тому, що запропоновані математичні моделі та методи можуть бути використані під час розробки, впровадження та підтримки мультисервісних мереж, що базуються на хмарному підході з використанням технології віртуалізації мережевих функцій. Зокрема, запропоновані методи і моделі аналізу і оцінки якості надання послуг були використані під час розробки процесів пошуку і бронювання системою MANO каналів зв'язку з достатньою пропускнуою здатністю, процесу резервування ресурсів за допомогою протоколу RSVP і формування агрегованого потоку даних уздовж віртуальних каналів зв'язку. Розроблені моделі та методи можуть бути використані як науково-методична база для подальших досліджень функціонування і надання сервісів у різних типах розподілених мереж, що функціонують на основі концепції NFV: під час розробки, підтримки, проектування і впровадження різних конвергентних мультисервісних мереж, заснованих на хмарному підході.

Результати дисертаційної роботи впроваджено у Державному підприємстві «Науковий центр точного машинобудування» Державного космічного

агентства України (акт впровадження № 1 від 4.04.2018 р.), ТОВ "Світ-ІТ", (акт впровадження № 1203-1 від 12.03.2019 р.), у навчальному процесі в Київському національному університеті будівництва і архітектури (Довідка про впровадження від 08.11.2019 р.).

### **Особистий внесок здобувача.**

Всі дослідження в рамках дисертаційної роботи автор виконав самостійно або в співавторстві. Автор самостійно виконав всі теоретичні і експериментальні дослідження, які складають основу дисертаційної роботи. Всі основні результати дослідження, представлені до захисту, отримані автором самостійно.

У статтях, виконаних у співавторстві, особисто автору належать наступні результати. В роботі [1] автором проведено дослідження методів формалізації взаємодії комплексних Web-сервісів, які використовуються в роботі сучасних інформаційних систем. В роботі [2] автором запропоновано та досліджено роботу методу реплікації найпростіших сервісів; в роботі [3] запропонований метод аналізу ключових показників функціонування мережевих ресурсів в мережах з підтримкою віртуалізації. В роботі [4] автором проведено аналіз розвитку концепції NFV і сформований ряд основних завдань дослідження. В роботі [5] отримав подальший розвиток метод формалізації взаємодії атомарних сервісів в мережах з підтримкою функцій віртуалізації, зокрема було запропоновано використовувати елементи теорії мережевого обчислення. В роботі [6] автором проведено дослідження методів побудови кореляційних алгоритмів класифікації мережевого трафіку в розподілених мережах; в роботах [7] і [8] запропонована методика оцінки масштабованості мережевих ресурсів, а також наведено оцінку зміни коефіцієнта масштабованості трьох основних структур рівня управління SDN.

### **Апробація результатів дисертації.**

Основні наукові результати і положення дисертації були представлені, доповідались та всебічно обговорені на дев'яти міжнародних науково-технічних конференціях і форумах: Conference BEAR, University of Birmingham,

24 June 2013 (Birmingham, UK), IX Міжнародна конференція “Електроніка та прикладна фізика”, жовтень 2013 р. (Київ, Україна), XII International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET’2014, Slavske, Ukraine), Радиоэлектроника и молодежь в XXI веке: 18-й Международный молодежный форум, 14–16 апреля 2014 г. (Харьков, Украина), Perspective technologies and methods in MEMS design, 22-24 June 2014 (MEMSTECH’2014, Polyana-Svalyava (Zakarpattia), Ukraine), The First International Scientific-Practical Conference Problems of Infocommunications Science and Technology, 14-17 October 2014 (Kharkiv, Ukraine), IX Міжнародна конференція “Електроніка та прикладна фізика”, 22-25 жовтня 2014 (Київ, Україна), 5th Annual Conference BEAR PGR, 15 December, 2014 (Birmingham, UK), XIIIth International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM’2015, Polyana-Svalyava (Zakarpattia), Ukraine), Сьома міжнародна науково-технічна конференція "Проблеми інформатизації", 12 грудня 2016 р.

**Публікації.** Основні результати дисертаційної роботи опубліковані в дев’яти наукових статтях. З них вісім наукових статей опубліковано в наукових фахових виданнях України [1-4, 6-9] та одна стаття у зарубіжному фаховому виданні [5]. Матеріали дисертаційних досліджень опубліковано у шести збірниках вітчизняних та міжнародних науково-технічних конференцій, в тому числі й публікації, проіндексовані у наукометричних базах IEEEExplore та SCOPUS.

**Структура та обсяг дисертації.** Дисертація складається зі вступу та чотирьох розділів. Загальний обсяг роботи становить 147 сторінок, в тому числі 126 сторінок основного тексту, 28 рисунків та 16 таблиць на 35 сторінках. Список використаних джерел містить 140 найменувань, викладених на 14 сторінках.



## РОЗДІЛ 1

### АНАЛІЗ МЕТОДІВ УПРАВЛІННЯ СЕРВІСАМИ В МУЛЬТИСЕРВІСНИХ МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ

За останні роки можна спостерігати як стрімко підвищується обсяг різноманітного трафіку, що шириться в інформаційно-комунікаційних системах, а також, розширюється спектр затребуваних сервісів, додатків тощо. За даними статистики корпорації Cisco Systems глобальний IP-трафік зріс більш ніж в чотири рази за останні 5 років [24, 78]. Через значне зростання кількості та різноманітності переданих даних, вагоме розширення спектра сервісів виникло суттєве збільшення навантаження на мережу та ускладнились завдання управління мережевою інфраструктурою. На теперішній час число протоколів, які підтримують процес передачі даних, стрімко збільшується (на даний момент використовується понад 600 стандартизованих протоколів), їх зростання спричинює ускладнення мережевих пристроїв, що в свою чергу погіршує повноцінне впровадження нових сервісів, а, відповідно, й швидкість адаптації.

Вище вказані обставини мали значний вплив на розвиток мережевої інфраструктури: принципи віртуалізації мережевих сервісів, відділення рівня додатків від рівня передачі даних, побудова центрів обробки даних і хмарних обчислень приходять на зміну традиційним технологіям зберігання та надання даних.

Обладнання користувача (CPE), яке за допомогою комутаторів чи мультиплексорів з'єднується з маршрутизаторами широкосмугового віддаленого доступу, виступає кінцевими компонентами мережі. На граничному діапазоні мережі використовуються різні програмно-апаратні засоби захисту переданих даних: firewall and DPI, NAT, etc. Мережа надання контенту і ряд граничних маршрутизаторів, що забезпечують обмін інформацією і передачу даних.

Здебільшого використовується безліч middleboxes - програмно-апаратних засобів для підтримки та надання сервісів у мережах зв'язку, які здатні реалізувати певні послуги. Оскільки програмне забезпечення та апаратна платформа досить пов'язані між собою, впровадження будь-якої нової послуги на існуючих програмно-апаратних компонентах, що утворюють мережу, вкрай складне. Без залучення виробника устаткування внести певні зміни у вже існуючу конфігурацію даного обладнання є дуже клопітною задачею. [88, 90, 129].

Ситуація, що склалася, призвела до зміни концепції надання послуг в цілому - основна увага сучасних науково-дослідних кіл і розробників у сфері телекомунікацій на сьогоднішній день звернена не на вдосконалення традиційних мультисервісних мереж, а на перехід до нової архітектури побудови інфраструктури мультисервісних мереж [73, 118]. Основною ідеєю в даному випадку є відділення рівня надання даних від рівня управління, а також віртуалізація ресурсів і функцій мережі. Даний підхід дозволить реалізовувати велику кількість послуг без прив'язки до фізичного обладнання.

Ідея відділення рівня управління від рівня передачі даних реалізована в концепції Software-Defined Networking [73, 118]. Відповідно до концепції SDN вся логіка і функції управління переносяться на окремий централізований пристрій - контролер [27, 75, 76, 118]. Саме на контролері реалізуються функції управління і моніторингу, необхідні для повноцінної роботи мережі. На відміну від SDN, в традиційних мережах ці функції реалізовані в одному пристрої, на базі загального (єдиного) набору системної логіки, тому їх відділення неможливе [27, 65, 90].

Основною ідеєю технології віртуалізації мережевих функцій (Network Function Virtualization, NFV) є логічний поділ апаратного і програмного забезпечення, надаючи при цьому різноманітні моделі сервісів (рис. 1.1) відповідно до типових вимог в рамках запитів найбільш широких груп користувачів [24-27, 29 ]:

- модель надання інфраструктури (IaaS: Infrastructure as a Service);

- модель надання апаратних ресурсів (Hardware as a Service);
- модель надання платформи (Platform as a Service);
- модель надання програмного забезпечення (Software as a Service).

В рамках моделі HaaS користувач забезпечується апаратними ресурсами (сервери, мережеве обладнання, накопичувачі та ін.), на основі якого розгортає свою власну інфраструктуру. Подібним чином в схемі IaaS апаратні ресурси надаються відповідно до масштабування системи. Користувач при цьому не здатний управляти інфраструктурою сервісу, але має контроль над операційною системою, протоколами зберігання і захисту даних, а також йому надається вибір мережевих компонентів. Таким чином, споживач організовує захист платформ і додатків, а провайдер хмарного сервісу - захист інфраструктури мережевого сервісу. З іншого боку, PaaS будується шляхом додавання до моделі IaaS операційної системи і інтерфейсу програмування API (Application Programming Interface). Споживач при цьому не може управляти базовою інфраструктурою хмари, але отримує контроль над додатками і параметрами конфігурації середовища хостингу. У моделі SaaS користувач отримує доступ до програмного забезпечення через мережевий доступ. В даній схемі питання оновлення та отримання ліцензій на ПЗ регулюються провайдером, а виплати здійснюються за фактичне використання ПЗ.

Ефективність використання моделі хмарного сервісу залежить від галузі її застосування. Відповідно до цього підходу мережеві сервіси можуть бути розділені на три основні групи:

- приватний мережевий сервіс;
- мережевий сервіс загального доступу;
- гібридний мережевий сервіс.

Приватний мережевий сервіс має на увазі високу собівартість на впровадження і підтримку системи, але при цьому дозволяє вибрати її конфігурацію і в індивідуальному режимі налаштувати систему безпеки. Даний вид сервісу підходить для великих корпорацій і ІТ-компаній. У мережевих сервісах загального доступу використовується стандартна конфігурація, тому

впровадження відбувається найбільш просто і собівартість сервісу виявляється мінімальною. Дані сервіси використовуються малими компаніями та індивідуальними користувачами. Комбінація даних видів сервісів, відома як гібридний мережевий сервіс, яка в свою чергу використовується організаціями з сезонними періодами активності.

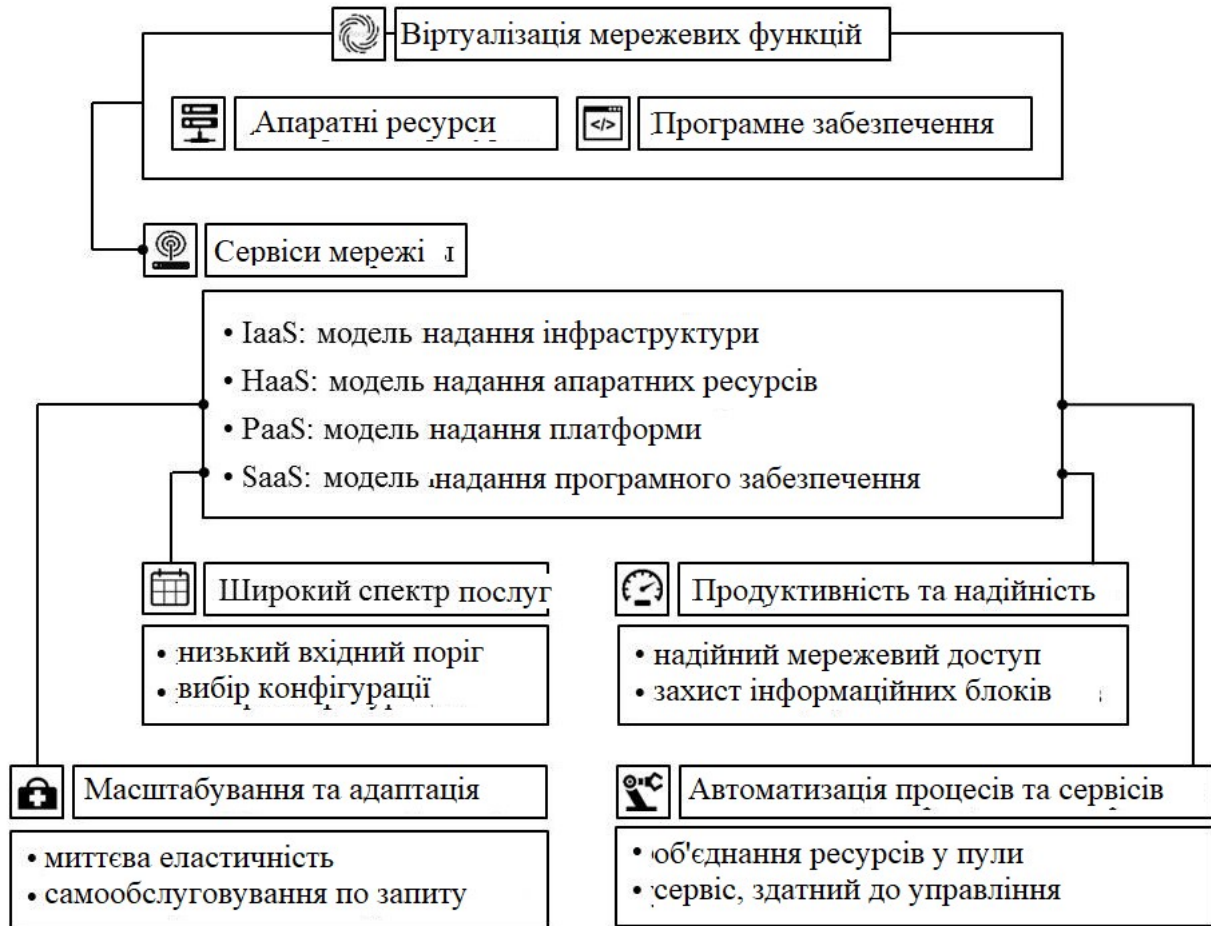


Рис. 1.1. Базова схема віртуалізації мережевих функцій

Аналогічно можна сформулювати рекомендації для класифікації мережевих сервісів типу IaaS, NaaS, PaaS і SaaS

Табл. 1.1.

Класифікація мережевих сервісів і області їх використання

	Модель надання послуг в рамках мережевого сервісу			
	IaaS	NaaS	PaaS	SaaS
<b>Переваги</b>	Ефективність ресурсоспоживання		Висока ефективність впровадження нових версій ПЗ	Низька собівартість
	Масштабування			Автооновлення ПЗ

<b>Недоліки</b>	Складність впровадження	Відсутність контролю над апаратними ресурсами	Відсутність контролю над апаратними ресурсами і ПЗ
	Висока собівартість		
<b>Області застосування</b>	Великі корпорації	ІТ-компанії	Малі компанії
			Окремі користувачі

NFV дає змогу операторам розгортати мережеві рішення (DPI, NAT, Firewall іт.д.) як програмні додатки, а не як окреме мережеве обладнання. Робота додатків NFV і сама реалізація віртуальних функцій можлива на високопродуктивних мережевих платформах і серверах, які з'єднані один з одним за допомогою пристроїв комутації і маршрутизації [25, 27].

На сьогоднішній день технологія NFV підтримується і розробляється низкою великих виробників обладнання та постачальників послуг, таких, як Cisco, AT & T, BT, Deutsche Telekom, Orange, Telecom Italia, Telefonica, Verizon і ін. Підтримкою та розробкою вимог і рекомендацій щодо функціонування та впровадження технології NFV займається Європейський інститут телекомунікаційних стандартів (European Telecommunications Standards Institute, ETSI) [23-27].

Основними перевагами, що очікуються від впровадження технології NFV, є:

- доступність великого спектру послуг;
- забезпечення високого рівня надійності в процесі надання;
- високий рівень масштабованості і адаптації мережевої інфраструктури;
- автоматизація процесів управління і надання сервісів.

У короткостроковій перспективі впровадження технології NFV в сучасні мультисервісні мережі, відповідно до досліджень Infonetics Global Service Provider Survey [74], дозволить підвищити ефективність таких складових, як масштабованість послуг, ефективність надання сервісів, швидка адаптація мережевої інфраструктури, підвищення швидкості надання, економія матеріальних і енергетичних витрат. Діаграма, що відображає відсоткове співвідношення виграшу від впровадження NFV, наведена на малюнку

1.2.

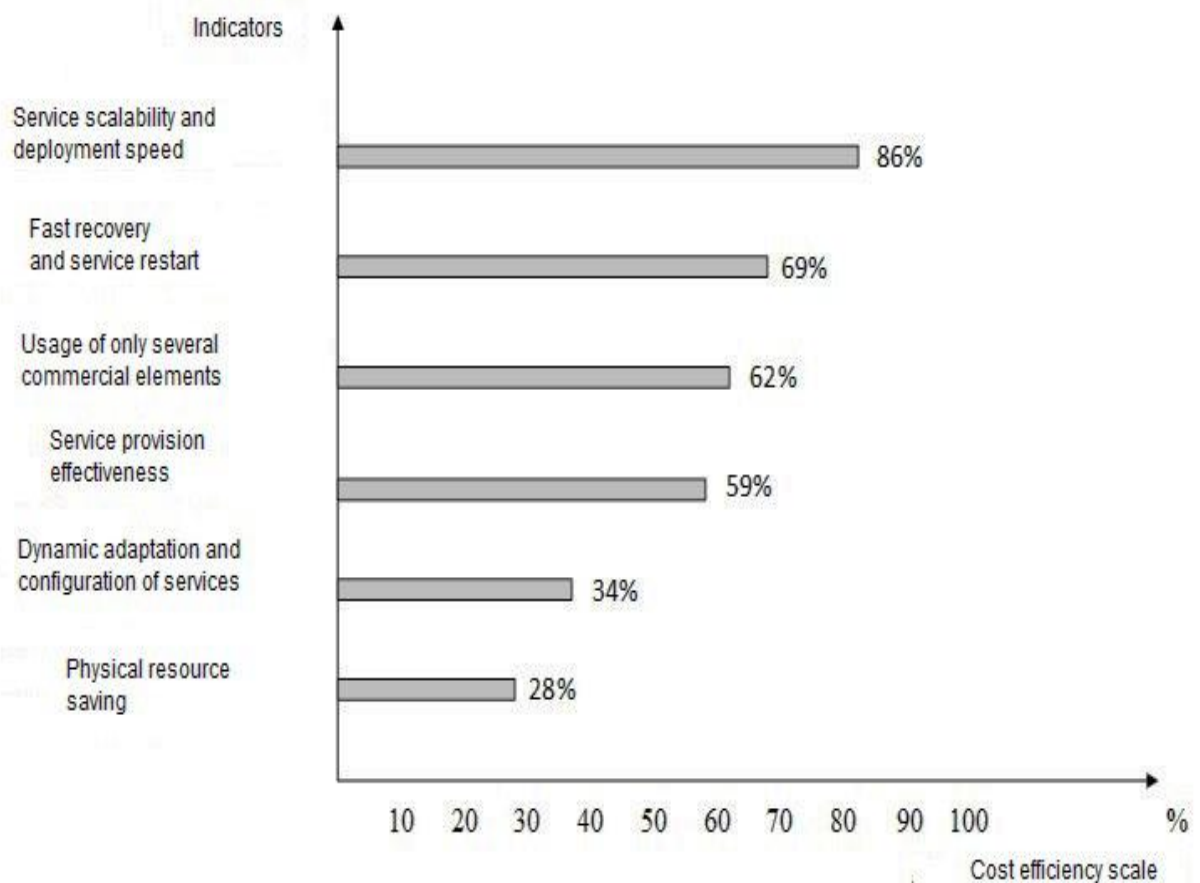
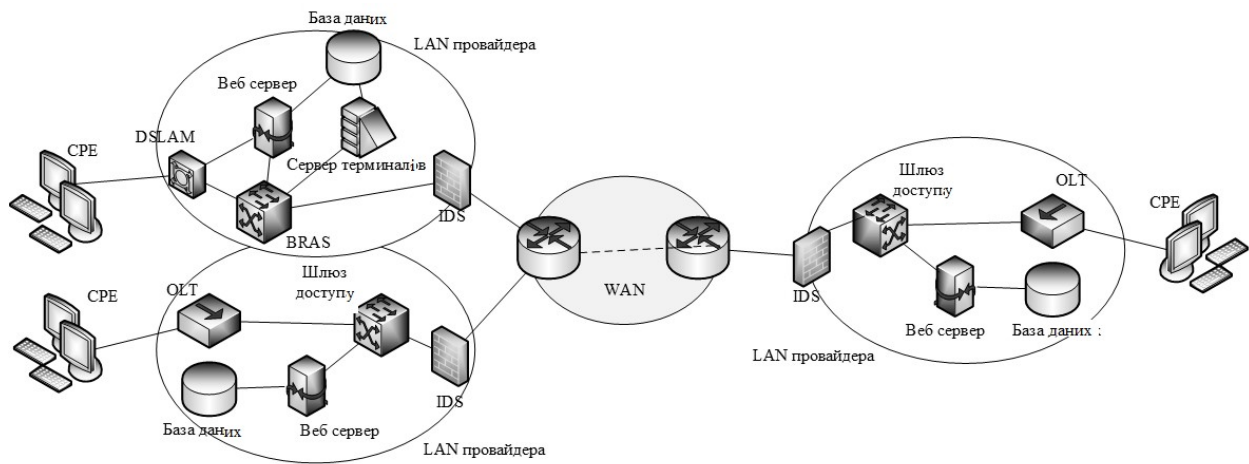
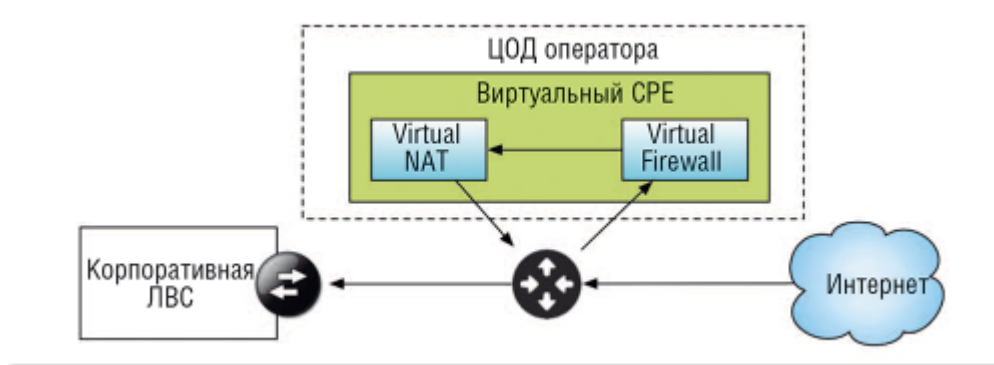


Рис. 1.2. Відсоток підвищення ефективності послуг мережевого сервісу під час впровадження технології NFV

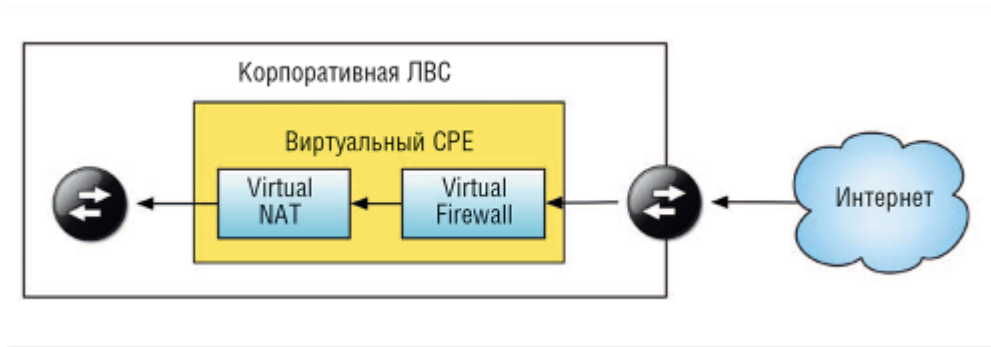
Технологія NFV має на увазі формування на базі універсальних серверів віртуального обчислювального середовища, яке і забезпечує весь необхідний набір послуг [116]. Відмінні риси побудови традиційної мережевої інфраструктури від мультисервісної мережі з підтримкою технології NFV, наведені на рисунку 1.3.



а)



б)



в)

Рис.1.3. Структурна модель традиційної мультисервісної мережі (а) і мережі із підтримкою технології NFV (б,в)

Для досягнення найкращих результатів в процесі проектування мереж на базі NFV рішень необхідно дотримуватися наступних базових принципів [29, 37, 39]:

1. Принцип відмовостійкості. У разі виникнення відмови NFV рішення повинно забезпечувати автоматичне відновлення працездатності за рахунок механізмів міграції VM, переходів на резервні маршрути і реплікації сервісів.

2. Принцип забезпечення гарантованої якості обслуговування при передачі різного типу даних. NFV рішення повинно забезпечувати процес надання сервісів з різними класами обслуговування шляхом забезпечення гарантованих показників QoS.

3. Принцип підтримки широкого спектра сервісів. При проектуванні NFV рішення мається на увазі використання певного набору сервісів;

4. Принцип ефективного розподілу і управління ресурсами NFV інфраструктури. У процесі формування і надання сервісів повинні бути використані різні механізми, які забезпечують швидку і гнучку адаптацію мережевої інфраструктури під вимоги додатків і послуг.

5. Принцип масштабованості. Можливість розширення мережевої інфраструктури без істотної деградації її ефективності.

6. Принцип забезпечення обліку витрачених мережевих ресурсів. Система управління та оркестровки (Management and Orchestration, MANO) і система підтримки бізнес послуг (Business Support System, BSS) NFV рішення повинні забезпечувати можливість збору статистичної інформації про кількість збережених, оброблених і переданих даних, а також інформації про збої і неполадки, що виникли в процесі надання сервісів.

### **1.1. Аналіз архітектури мультисервісних мереж з підтримкою віртуалізації мережевих функцій**

Відповідно до ETSI [24-27], архітектура NFV складається з трьох ключових елементів: мережевої функції віртуалізації інфраструктури (NFVI), VNFs і NFVMANO [23-27]. Структурна схема архітектури NFV приведена на Рис. 1.4.



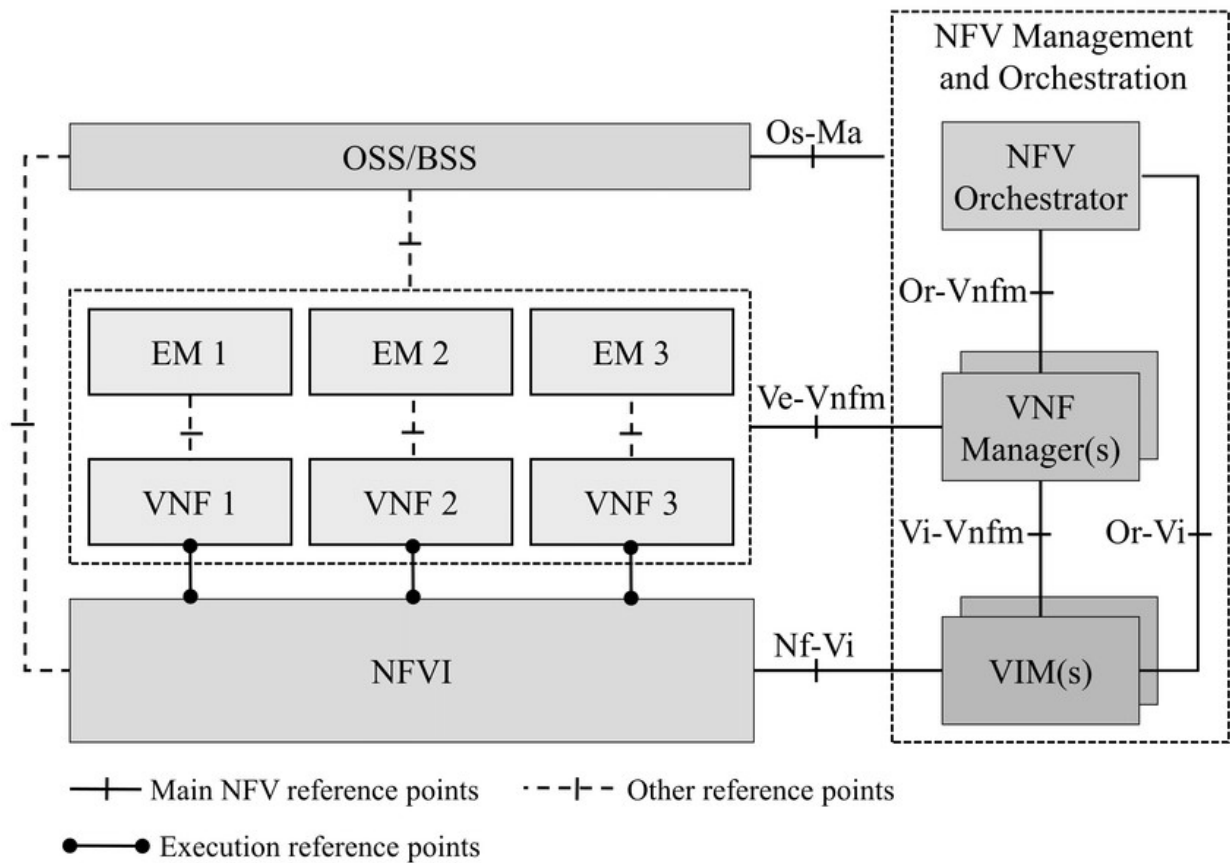


Рис.1.4. Архітектура віртуалізації мережевих функцій

NFV інфраструктура (NFVI) [23] є єдиною платформою обробки, зберігання та передачі даних, яка реалізується за допомогою взаємодії фізичних і віртуальних мережевих ресурсів. Фізичні ресурси NFVI включають в себе серверне обладнання, обладнання комутації та маршрутизації, системи зберігання даних і канали зв'язку. Злагоджена взаємодія цих компонентів забезпечує коректну передачу даних від кінцевого користувача до обчислювальних елементів, і навпаки.

Рівень віртуалізації є безліччю абстракцій: віртуальні машини з різними операційними системами (ОС) і додатками та центри зберігання даних. Саме вони забезпечують формування та надання послуг кінцевим користувачам.

За коректну взаємодію між фізичними і віртуальними компонентами мережевої інфраструктури відповідає гіпервізор [36, 109]. Він забезпечує розподіл (виділення, звільнення і ізоляцію) мережевих ресурсів між віртуальними машинами в процесі емуляції і надання сервісів. Віртуальна

мережа, яка об'єднує обчислювальні ресурси, формується за допомогою таких протоколів, як VXLAN, NVGRE, BGP L2, BGP L3 [41, 16, 119, 113]. Точками входу в віртуальну мережу інфраструктури NFV виступають гіпервізори.

Модуль мережевих функцій і сервісів представляє собою відокремлений компонент інфраструктури NFV. Він включає в себе безліч віртуальних машин і виконує набір функцій і послуг, характерний для емуляції певного мережевого обладнання. Широкого поширення набули емулятори обладнання домашньої мережі, такі як Residential Gateway (RGW) [5, 7], емулятори DHCP-серверів, міжмережеві екрани, маршрутизатори і т.д. Залежно від кінцевого призначення складу модуля VNF може змінюватися - запуск і зупинка віртуальних машин коригується системою управління і оркестровки MANO.

Згідно специфікацій і рекомендацій ETSI [25] основним призначенням системи MANO є управління процесами надання послуг кінцевим користувачам: система MANO координує всі дії обладнання інфраструктури NFV, включаючи пошук і розподіл мережевих ресурсів, необхідних для формування та надання послуг, а також підтримку і моніторинг за станом послуг протягом усього життєвого циклу. Основними завданнями, які розв'язуються компонентами системи MANO є [26, 98, 113]:

- Resource Management / Управління ресурсами мережевої інфраструктури;
- Function and Services Management / Управління функціями і сервісами, що надаються;
- Fault Management / Управління відмовами;
- Configuration Management / Управління конфігурацією;
- Accounting Management / Облік;
- Performance Management / Управління продуктивністю;
- Security Management / Управління безпекою.

Компоненти, що вирішують завдання управління ресурсами мережевої інфраструктури, такі як оркестратор, NFM, VNF & S, несуть відповідальність за коректний розподіл мережевих ресурсів, формування оптимального маршруту передачі і вибору місця розташування сервісу.

Компоненти, що вирішують завдання управління функціями і сервісами, такі як дескриптори віртуальних функцій і сервісів, модулі управління мережевою інфраструктурою, VIM, NVFO, виконують опис і формування сервісів відповідно до SLA, а також створення і вибір правил та політики управління компонентами мережі в процесі надання сервісів [73, 103].

Компоненти, що вирішують завдання управління відмовами, такі як контролер і менеджер мережевого управління, VIM, несуть відповідальність за виявлення і виправлення неполадок мережі. Основними функціями даних компонентів є моніторинг за станом фізичного мережевого обладнання, його тестування і діагностика, і, як результат, формування сигнальних повідомлень про виникнення аварійних ситуацій.

Компоненти, що вирішують завдання конфігурації і реконфігурації віртуальних і фізичних мережевих елементів, такі як оркестратор, VNF & S, здійснюють збір статистичних даних, формування карти мережевої інфраструктури, моніторинг і контроль за станом віртуальних і фізичних елементів мережі, включаючи канали зв'язку, і будь-якої їх модифікації.

Компоненти, що вирішують завдання розподілу мережевих ресурсів, такі як MNF, MIF, VNF & S, виконують функції розподілу та обліку мережевих ресурсів і функцій.

Компоненти, що вирішують завдання оцінки та управління продуктивністю, такі як оркестратор, VLD, VNF менеджер пакетів і менеджер сервісів, виконують збір статистичних даних про функціонування мережі та мережевих елементів в реальному масштабі часу, в разі необхідності забезпечують перерозподіл потоків даних і міграцію мережевого обладнання.

Компоненти, що вирішують завдання управління безпекою, такі як пристрої аутентифікації і аудиту, VIM, NS, забезпечують контроль надання сервісів, ведення журналів доступу, захист від зовнішніх і внутрішніх порушників.

## 1.2. Аналіз механізмів оркестрації послуг в мультисервісних мережах з підтримкою технології віртуалізації мережевих функцій

Процес надання послуг за допомогою технології NFV має ряд істотних відмінностей, обумовлених, в першу чергу, наявністю і взаємодією фізичних і віртуальних компонент. У [24, 71, 72] визначено, що процес надання послуги «від краю до краю» спочатку моделюється за допомогою графа перенаправлення мережевих функцій (VNF Forwarding Graph). Основною метою побудови графа перенаправлення мережевих функцій є визначення і формування зв'язків між упорядкованим безліччю модулів NRF & S з подальшим наданням послуг. Результатом побудови графа перенаправлення є формування послідовного ланцюжка взаємодії модулів VNF із зазначенням всіх фізичних і віртуальних каналів (ВК) зв'язку, а також точок з'єднання (ТС), задіяних в процесі надання послуги. МС є інтерфейс модуля реалізації мережевих функцій та послуг, вони містять ряд правил і політик необхідних для обробки вхідного і вихідного потоків даних. Вибір певного правила або політики залежить від рішення рівня управління та оркестровки. Структурна схема графа перенаправлення приведена на рисунку 1.5.

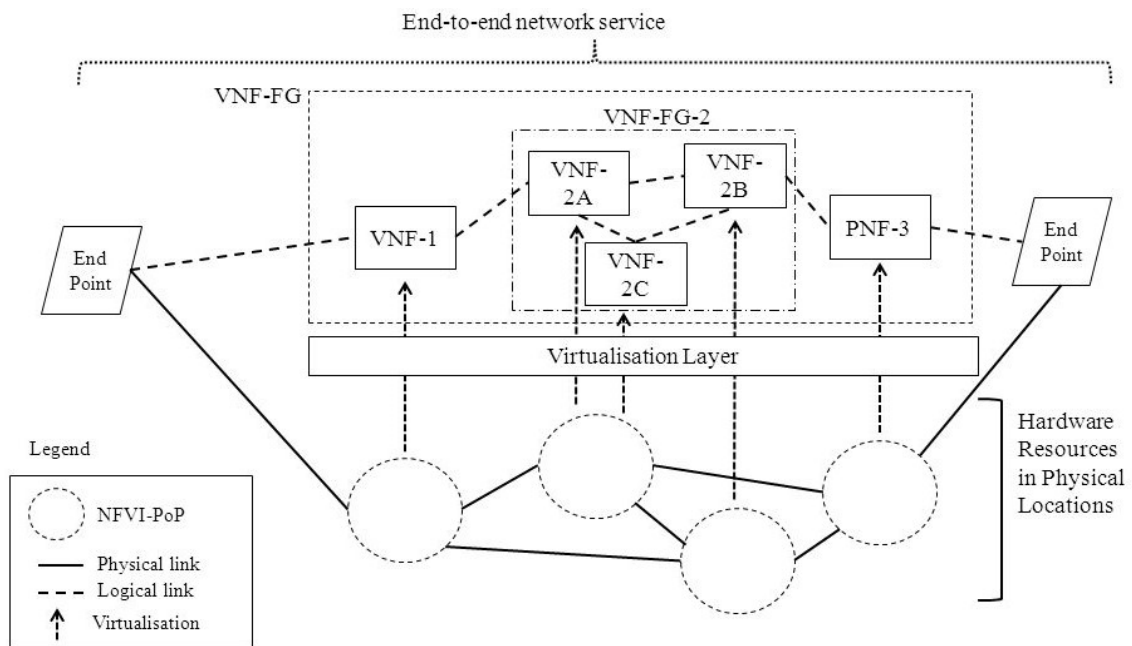


Рис. 1.5. Структурна схема графа перенаправлення FG в рамках надання послуги кінцевому користувачу

Аналіз специфікацій [23-27], рекомендацій [39, 40, 71, 73] і готових архітектурних рішень [2, 4, 5, 83] показав, що процес формування графа перенаправлення мережевих функцій є одним з ключових аспектів функціонування технології NFV: від ефективності його побудови залежать показники QoS і QoE кінцевих сервісів. Слід зазначити, що в першу чергу в процесі формування графа перенаправлення послуги повинні бути враховані наступні аспекти:

- здатність обробки і підтримки необхідного набору функцій фізичним обладнанням мережі. Оцінюються фізичні характеристики мережевого обладнання (обсяги RAM, CPU, SSD), а також його програмне забезпечення (реалізація механізмів побудови віртуальних тунелів, підтримка протоколів передачі даних і подання кінцевого сервісу);

- коректна взаємодія фізичного обладнання з віртуальними елементами. Оцінюється ефективність функціонування гіпервізора і формування кінцевих точок формування віртуального каналу - точок VTEP [50] віртуальної мережі;

- ефективне формування керуючих команд, яке залежить від обраного типу системи управління і оркестровки. Оцінюється логіка формування керуючого рішення, частота оновлення керуючої інформації і коректність керуючого рішення.

На сьогоднішній день формування графів перенаправлення здійснюється на основі шаблонів поведінки, які заздалегідь відомі, погоджені із замовником послуг і визначені в SLA (наприклад, надання послуг телефонії або захист периметра мережі за допомогою VNF).

У загальному випадку процес формування графа перенаправлення мережевих функцій включає в себе три основних етапи [71, 72, 103]:

1. Обробка запиту користувача системою управління. Даний процес включає в себе збір інформації про сервіс, формування великої кількості атомарних сервісів або комплексного сервісу, який відповідає необхідним показникам якості обслуговування.

2. Вироблення вимог до інфраструктури управління, пошук та конфігурація модулів VNF, які здатні забезпечити необхідний рівень якості.

3. Формування шляху передачі сервісу. Даний процес включає в себе формування фізичного маршруту передачі даних, включаючи вибір протоколів передачі та інкапсуляції даних на фізичному рівні, і формування віртуальної мережі, включаючи вибір механізмів віртуалізації каналів.

Формування комплексних сервісів в мультисервісній мережі з підтримкою технології NFV може бути виконано за допомогою статичних або динамічних методів [86].

Формування комплексних сервісів в мультисервісній мережі з підтримкою технології NFV може бути виконано за допомогою статичних або динамічних методів [86].

У разі застосування методу статичного формування сервісів, дії, сервіси, керуючі потоки (MANO-NFV) і потоки передачі даних (NFV-PNF) визначаються в період проектування на основі SLA сервіс провайдером і замовником і мають статичне значення протягом усього життєвого циклу сервісу [26]. Це означає, що після аналізу вимог комплексний сервіс з фіксованим складом сервісів є сформованим і внести будь-які зміни в процесі його представлення неможливо. Однак, на відміну від архітектури SOA, може бути сформований ряд обхідних маршрутів, що збільшує показники надійності сервісу [64, 66, 80].

У загальному вигляді цей процес задається у вигляді моделі [80, 82, 101], що містить наступні компоненти:

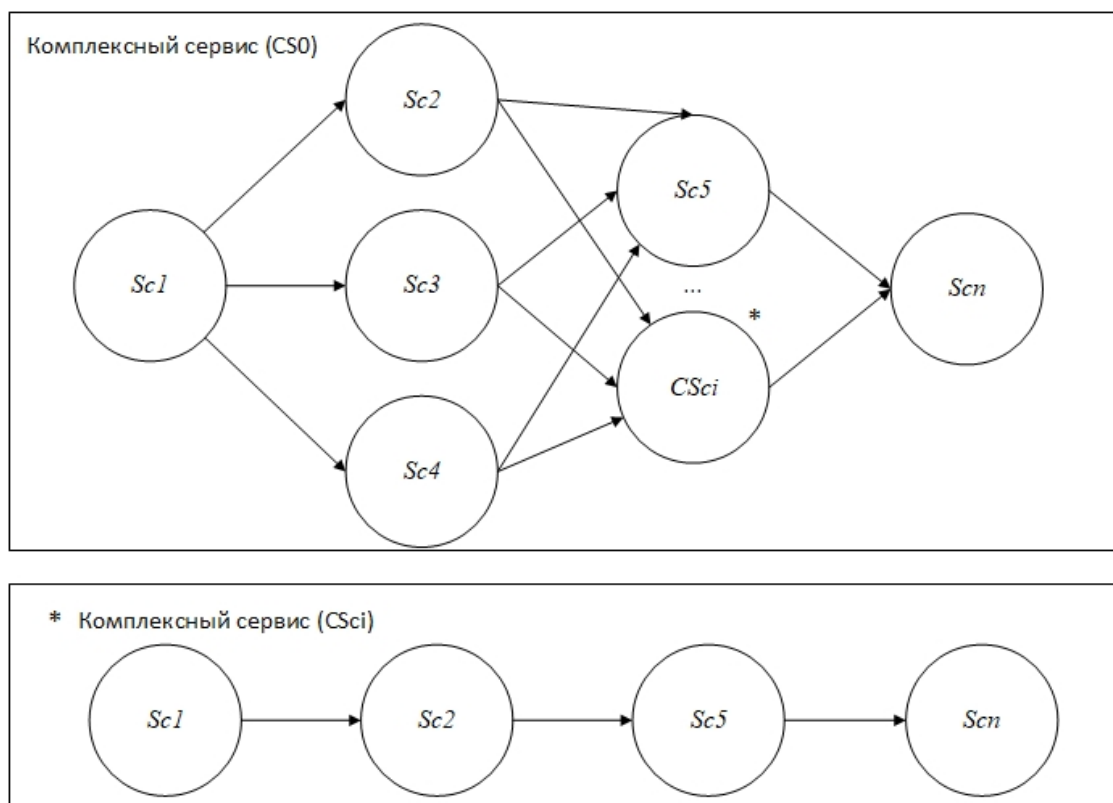
У загальному вигляді цей процес задається як модель [80, 82, 101], що містить наступні компоненти:

- Набір атомарних сервісів, які можуть бути реалізовані NFV інфраструктурою сервіс провайдера. Значення набору атомарних послуг міститься в реєстрі функцій і сервісів NRF & S;
- Набір функцій, які можуть бути реалізовані на фізичному обладнанні NFV інфраструктури і залежать від характеристик інфраструктури: і

віртуальної мережевої інфраструктури. Значення набору функцій міститься в модулі VIM, MFV, NS;

- Набір класів обслуговування, відповідних SLA. Набір класів обслуговування є шаблоном для формування кінцевої послуги, яка відповідає вимогам QoS. Значення набору класу обслуговування міститься в OSS / BSS.

На підставі значень набору класу обслуговування, а також характерних для NFVI значень, і формується безліч кандидатів комплексного сервісу, таких що відповідають заданим параметрам QoS. На підставі комбінації можливих кандидатів формується вектор вибору, який містить всі можливі комбінації сервісів, які можуть бути надані в рамках даної NFVI і відповідають вимогам QoS. Процес формування комплексного сервісу наведено на рисунку 1.6.



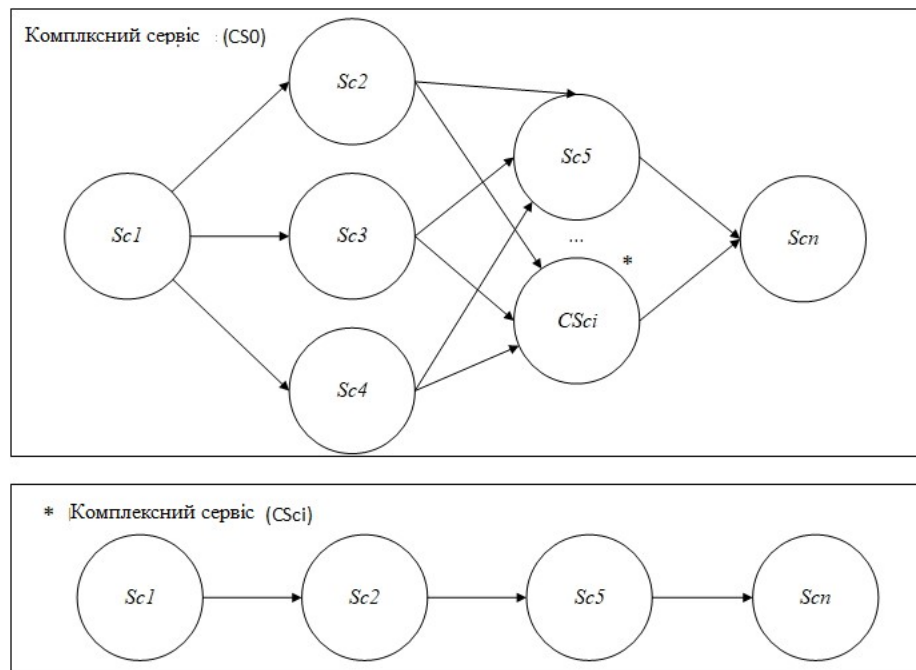


Рис. 1.6. Процес формування комплексного сервісу CS

При динамічному формуванні комплексних сервісів можуть бути зроблені модифікації або заміни атомарних сервісів, які дозволяють мережевій інфраструктурі адаптуватися до вимог кінцевого користувача і управляти функцією повернення в разі виявлення помилки [89, 108]. Іншими словами, склад комплексного сервісу може формуватися кожен раз при зверненні до нього користувача або компонент системи MANO. Приклад динамічного формування складу комплексного сервісу описаний в [88]. Метою даного підходу є досягнення високої гнучкості формування комплексних сервісів і поліпшення надійності композиції.

Зазначені методи формування сервісів можуть бути використані разом для забезпечення максимальної гнучкості, так як вони можуть впливати на різні рівні формування комплексних сервісів. Також необхідно визначити конкретні завдання і процеси, щоб виконання методів під час виконання сервісу було таким же, як і в період проектування.

До динамічних методів формування сервісів відноситься метод, відомий як Re-Binding [87, 103, 108]. Даний метод дозволяє замінити один сервіс іншим функціонально еквівалентним сервісом. Мається на увазі, що в період



проектування повинні визначатися лише деякі альтернативні атомарні сервіси, які покривають загальну функціональність комплексного сервісу.

На рисунку 1.7 представлений комплексний сервіс CS, що складається з набору атомарних. Для кожного атомарного сервісу, що входить до складу комплексного CS0, існує набір функціонально еквівалентних сервісів  $S_{ki}$ , що належать тим самим вектором розмітки QoS. У разі, якщо один з атомарних сервісів потребуватиме заміни, в зв'язку з погіршенням його показників QoS, то запуситься процес re-binding, який вибере новий сервіс з безлічі {}.

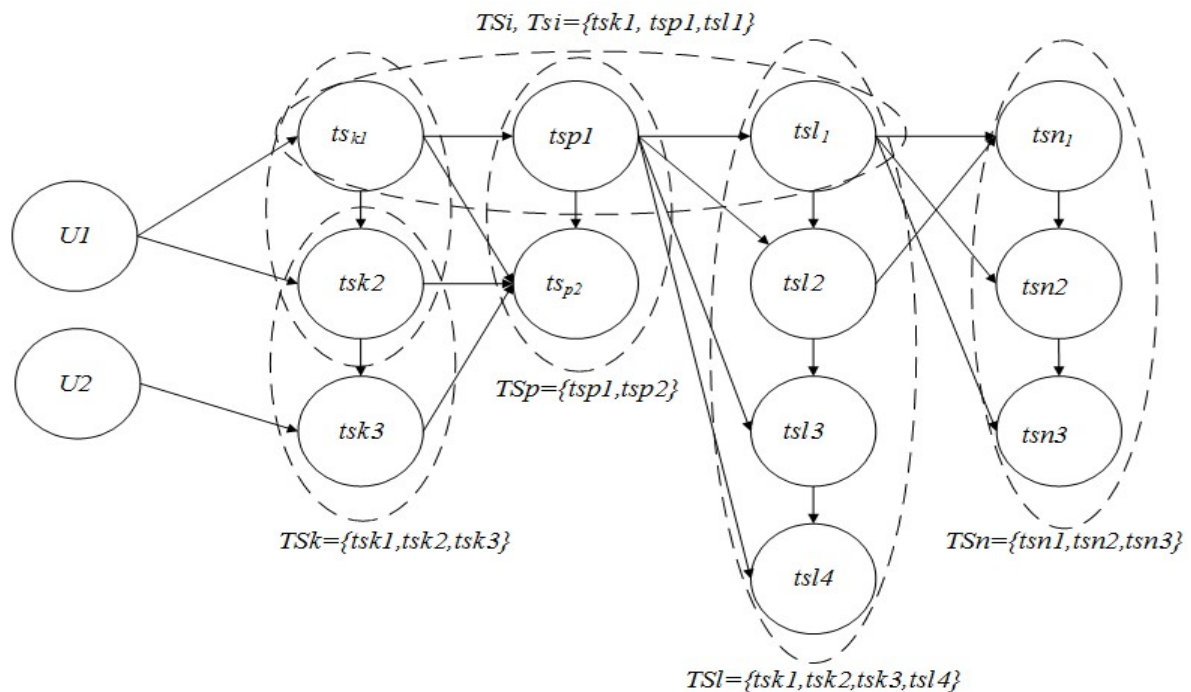


Рис.1.7. Процес динамічного формування комплексної послуги.

### Метод Re-Binding

**Variability Points.** В даному методі в комплексному сервісі виділяють певні точки внесення змін, в яких вказуються різні можливі варіанти виконання певної дії. Залежно від ситуації, коли комплексний сервіс затребуваний, і умови для кожної частини сервісу визначені, одна з альтернативних реалізацій складеного сервісу буде обрана і запущена. Відмінності між альтернативними реалізаціями частин сервісу можуть визначатися різницею в характеристиках сервісу або їх продуктивності під час виконання.

**Re-Planning.** Даний метод використовується у випадку, коли взаємодія сервісів в композиції невдала і відсутня можливість замінити один атомарний

сервіс іншим. В такому випадку необхідно модифікувати композицію сервісів таким чином, щоб відновити втрачену функціональність.

### 1.3. Методи забезпечення якості обслуговування в мультисервісних мережах з підтримкою віртуалізації мережевих функцій

Як було відзначено, функціонування мультисервісних мереж з підтримкою віртуалізації в процесі надання послуги значно відрізняється від процесів надання послуг в традиційних телекомунікаційних мережах [103, 112]. У першу чергу дана ситуація пов'язана з тим, що атомарні сервіси, що формуються в рамках модулів MANO є слабо пов'язаними між собою абстракціями і не залежать від програмної і апаратної платформи (Рис. 1.8).

Незважаючи на те, що послуги формуються і надаються кінцевим користувачам на верхніх рівнях моделі OSI (IP телефонія, функції файрволла з перевіркою контенту), їх якість обслуговування багато в чому залежить від стану транспортного середовища [88, 89]. У тих випадках, якщо показники фізичного обладнання або каналів зв'язку недостатні, наприклад, пропускна здатність транспортної мережі має низькі значення або сервер не оснащений необхідною кількістю RAM, то, незважаючи на високі показники QoS послуги у віртуальній мережі, час її надання буде істотно збільшено.

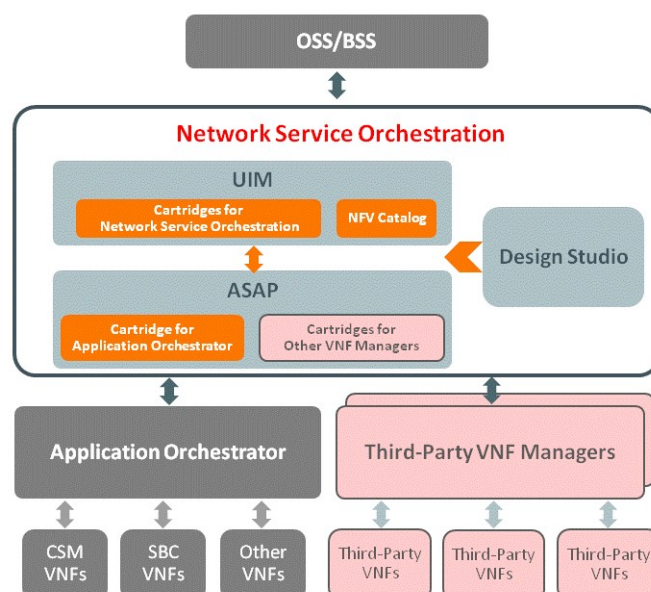


Рис.1.8. Архітектура оркестрації сервісів

Компанії, що займаються розробкою і випуском стандартів і специфікацій архітектури та принципів функціонування NFV, основну увагу приділяють організації слабозв'язаної, платформено-незалежної взаємодії споживача і постачальника сервісу. Такий підхід дозволяє забезпечити легкість і гнучкість взаємодії споживача сервісу та провайдера незалежно від технології реалізації сервісу, апаратної і програмної платформи [108, 114]. Недоліком такого підходу є неможливість гарантувати характеристики продуктивності, надійності, доступності сервісу.

У стандарті, розробленому інститутом ETSI [26] показники якості функціонування NFV і надання послуг розділені на дві основні групи: функціональні і нефункціональні. Відповідно до [27, 73, 82] категорії QoS кожного сервісу мультисервісної мережі з підтримкою NFV архітектури включають: бізнес вимоги, вимоги за рівнем обслуговування користувача, вимоги щодо сумісності, вимоги до бізнес процесу, вимоги до керованості, вимоги до безпеки.

Зазначені категорії поділяються на бізнес групу (Business quality group, BQG) і системну групу (System quality group, SQG) в залежності від того, до якого аспекту системи вони відносяться.

Показники QoS, що відносяться до бізнес групи, характеризують екземпляр сервісу з точки зору економічних факторів (ціна, функціональність, репутація провайдера і власника і т.п.). Дана група показників якості використовується при комерційному виборі екземпляра сервісу.

Системні показники якості в свою чергу поділяються так само на дві групи: статичні і динамічні.

Статичні показники якості вносяться в опис примірника сервісу в момент його впровадження і не змінюються протягом усього часу його існування. До таких показників відносяться показники бізнес-логіки, керованості, сумісності, безпеки. Управління даними показниками можливо лише на рівні статичних моделей, що використовуються в процесі розробки сервісів, телекомунікаційних протоколів і систем мережевого управління.

Динамічні показники якості обслуговування, що відображають стан примірника сервісу в процесі його функціонування, заслуговують на більшу увагу. До групи динамічних показників віднесені: час відгуку, продуктивність, готовність, доступність, надійність, кількість відмов.

Під час відгуку розглядається час з моменту відправки запиту до моменту отримання відповіді. Час відгуку в процесі надання послуг NFV в загальному залежить від чотирьох компонент: затримки обробки на стороні клієнта (обладнання CPE), фізичної мережі передачі даних (включаючи канали зв'язку і затримку обробки на серверах), затримки обробки гіпервізова або пристроїв віртуалізації і затримки віртуальної інфраструктури (включаючи віртуальні канали зв'язку і формування послуги на віртуальній машині).

Для забезпечення і підтримки заданого рівня якості обслуговування необхідне виконання основних функцій MANO [8, 57, 78], а саме:

Функція управління топологією відповідає за виявлення і підтримку мережевих зв'язків на підставі даних, отриманих від OpenFlow комутаторів.

Функція управління ресурсами відповідає за збір актуальної інформації про зайнятість та розподіл ресурсів в мережі. При розрахунку маршрутів передачі даних або ж управління чергою необхідне отримання даних.

Функція управління чергою забезпечує підтримку QoS, засновану на інформації про пріоритетність черг.

За збір визначення потоку, отриманого від постачальника послуг через інтерфейс контролера обслуговування відповідає функція управління потоком, яка, також, здатна забезпечити ефективне управління потоками за допомогою агрегування визначення потоку.

Функція розрахунку маршруту відповідає за визначення маршрутів (наприклад, найкоротший шлях і маршрути QoS) для різних типів потоків. Для забезпечення відповідності вимог до продуктивності та завдань різних потоків певна кількість алгоритмів маршрутизації може виконуватися паралельно.

Насамперед, щоб вирішити поставлене завдання, слід визначити структуру показників якості сервісів, які можуть бути використані для оцінки

рівня QoS [43, 80], що надаються інфокомунікаційними мережами, побудованими на базі NFV.

Відобразити в повному обсязі залежність показників якості і рівень QoS для різних видів послуг в інфокомунікаційних мережах, заснованих на базі NFV не дають змоги наступні показники якості мережі: затримка; варіація затримки (джиттер); кількість пакетів з помилками; кількість втрачених пакетів, описаних в рекомендації Y.1540 [99].

Представлені показники якості можна розділити на дві групи: статичні і динамічні. Статистичні показники якості – це стаціонарні значення, які підтримуються протягом всього життєвого циклу. До даної групи показників можна віднести надійність, доступність, керованість, готовність, безпека тощо. Динамічні показники якості обслуговування відображають стан параметрів мережі в процесі надання сервісу. До таких показників якості відносяться: час відгуку, затримка і її варіації, продуктивність, пропускна здатність, кількість відмов.

Готовність відображає частину часу, протягом якого екземпляр сервісу знаходиться в робочому стані:

$$K = 1 - \frac{T(t_i)}{\sum_{i=1}^N t_i}, \quad (1.1)$$

де  $T(t_i)$  – час, протягом якого сервісу не здатен опрацювати запит від користувача.

Доступність є ймовірність того, що екземпляр сервісу доступний і здатний обробити запит користувачів. Ймовірність доступності сервісу може бути представлена у наступній формулі:

$$P_{\text{дост\_серв}}(A) = \frac{P(N_{\text{запит}})}{P(N_{\text{відповідь}})}. \quad (1.2)$$

Згідно з [33] опис завдання надійності можна представити в такий спосіб (рис. 1.9):

$$R(t_i) = \sum_{i=1}^N (1 - P_{\text{дост\_серв}}(A)) + K. \quad (1.3)$$

де  $N$  – кількість запитів користувача.

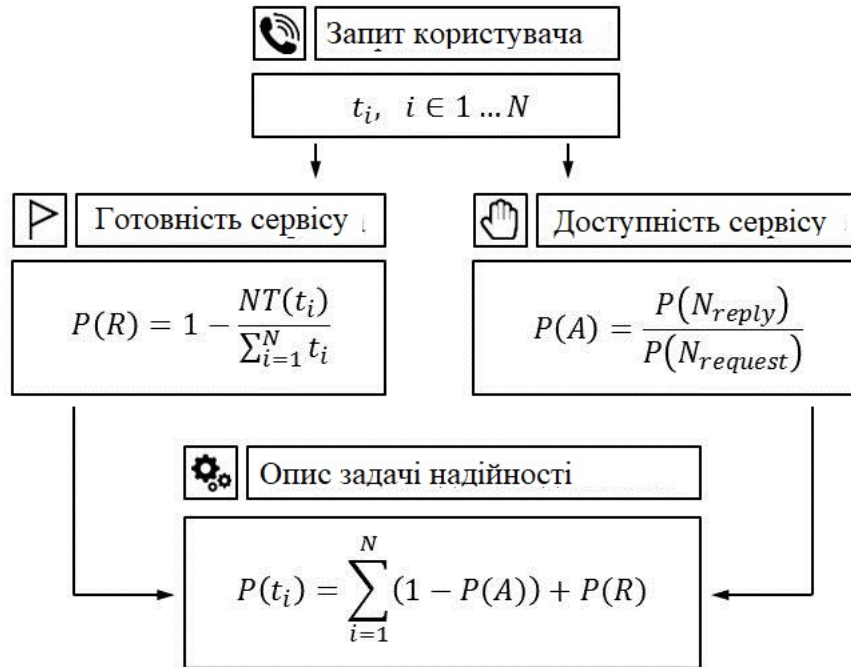


Рис. 1.9. Схема розрахунку надійності мережевого сервісу.

При оцінці QoS для мережевої інфраструктури з можливістю виникнення як паралельних, так і послідовних процесів застосовуються такі формули:

$$T(t_{i,j}) = \min(T(t_1), \dots, T(t_2)) \quad (1.4)$$

$$C(t_{i,j}) = \sum_{i \leq k \leq j} C(t_k)$$

$$R(t_{i,j}) = \sum_{m=0}^l f \sum_{k=i}^b (k - s) * (2k - 1) R(t_1) * \dots * ((1 - k_j) + (2k_j - 1) R(t_j)). \quad (1.5)$$

де  $T(t_{i,j})$ —мінімальний час відгуку при виконанні завдання  $t_{i,j}$ ;  $C(t_{i,j})$  — мінімальна пропускна здатність при виконанні повного завдання  $t_{i,j}$ ;  $R(t_{i,j})$ — надійність мережі.

Вказані рівняння дають змогу оцінити залежність показників QoS NFV, від фізичної складової мережі.

Для того, щоб в повній мірі проаналізувати та дати оцінку показникам якості, необхідно брати до уваги значення, які вносяться програмним компонентом (контролер) та протоколами управління ресурсів.

На теперішній час механізми забезпечення заданої якості сервісів, що надаються, можуть виконувати вимоги додатків лише за рахунок управління інфраструктурою мережі.

До провідних механізмів забезпечення заданого рівня QoS відносяться механізми управління чергами, пріоритезація трафіку, QoS маршрутизація і методи балансування навантаження, резервування ресурсів.

Поділ переданого трафіку на кілька класів, що надає кожному класу певний пріоритет, являється головною задумкою, яка лежить в основі механізму пріоритезації. Ресурси кожного комутатора і контролера розподіляються таким чином, щоб трафік з великим пріоритетом отримував більшу кількість його ресурсів [81, 126].

На практиці бажаний перерозподіл ресурсів проводиться за рахунок планування черговості обробки пакетів з різних потоків. Отже, пріоритезація трафіку вимагає від комутаторів підтримки здатності перерозподілу ресурсів в залежності від пріоритетності трафіку.

QoS маршрутизація являє собою потужний механізм забезпечення підтримки необхідного рівня якості. Звичні протоколи маршрутизації (RIP, OSPF, EIGRP) не дають змоги враховувати вимоги якості окремих потоків. Це призводить до можливості трафіку з високим пріоритетом бути направленим альтернативним маршрутом з поганою якістю.

QoS маршрутизація здійснює маршрутизацію кожного потоку відповідно до вимог QoS і незалежно від трафіку інших додатків. Для того, щоб

реалізувати QoS маршрутизацію необхідно, щоб комутатори мали здатність запам'ятовувати правила маршрутизації для кожного з потоків, що проходять через нього, і співставляти з ними пакети, що надходять.

Наявність даної вимоги призводить до ускладнення комутаційних пристроїв. Цей далекосяжний напрямок може мати складнощі при подальшому розвитку через нестачу стандартних рішень.

#### **1.4. Постановка наукового завдання і формулювання особистих завдань дослідження**

Розширення спектру, обсягів і методів надання послуг кінцевим користувачам призводить до значної зміни парадигми побудови мультисервісних мереж. Однією з найбільш перспективних технологій в цій галузі є технологія NFV. Маркетингові дослідження, проведені компанією HP [35] і Infonetics [74] показали, що при її впровадженні рентабельність надання послуг підвищиться в середньому на 35-40%. Однак впровадження даної технології супроводжується істотною зміною концепції управління мережею і принципів надання послуг кінцевим користувачам. Архітектура мережі в даному випадку поділяється на три компоненти: фізичне середовище передачі, віртуальне середовище формування послуг та середовище управління і оркестровки. Дані компоненти є пов'язаними і зміни однієї з них ведуть до значної модифікації процесу надання послуг і, як правило, до деградації послуги.

Наприклад, обмін даними між віртуальною мережевою інфраструктурою та фізичними компонентами блоку NFVI забезпечується на каналному і мережевому рівнях за допомогою спеціалізованих інтерфейсів. Взаємодія на каналному рівні здійснюється за допомогою формування віртуального мостового з'єднання на основі протоколу BGP, BGP Ethernet VPN. Взаємодія на мережевому рівні також забезпечується за допомогою формування тунелю з використанням протоколу BGP, BGP IP VPN. У цьому випадку формується



безліч окремих віртуальних мереж, кожна з яких має свій власний незалежний діапазон IP-адрес. Діапазони IP адрес, які використовуються віртуальними мережами можуть перекривати одна одну, тобто служби L3 VPN повинні підтримувати ізоляцію простору адрес. Як і в попередньому випадку даний функціонал накладається на площину управління. Однак не у всіх реалізаціях системи MANO забезпечується підтримка механізму ізоляції IP-адрес. Відсутність підтримки цих галузей в одному з фрагментів мережі веде до некоректного надання сервісів. Застосування уніфікованого підходу до побудови NFVI, формалізація вимог специфікації, а також пошук можливих протиріч у вимогах специфікації дозволить усунути причини появи подібних помилок.

Конвергенція традиційних мереж і мереж з підтримкою технології NFV покликана забезпечити розширення спектру і підвищення ефективності надання сервісів. Для цього повинен бути реалізований механізм, який дозволяє побудувати карту традиційної мережевої інфраструктури, здійснювати коректну обробку вхідних потоків даних, (визначати клас обслуговування, інкапсулювати / декапсулювати пакети і направляти до місця призначення). Найчастіше дані властивості накладаються на демаркаційні пристрої. При цьому, подібні пристрої повинні забезпечувати як підтримку віртуальних портів - для зв'язку з NFV інфраструктурою, так і фізичних. Аналіз показників якості послуг і верифікація процесів надання послуг за допомогою технології NFV, на ранніх стадіях її розробки, дозволяє своєчасно виявити «вузькі місця» мережі.

Таким чином, актуальною є розробка нових методів комплексного аналізу та оцінки коректності функціонування мереж з підтримкою технології NFV.

Відповідно до визначеної наукової задачі необхідно вирішити такі особисті задачі дослідження:

1. Провести аналіз архітектури, принципів побудови мультисервісних мереж з підтримкою функцій віртуалізації та методів управління ресурсами для забезпечення необхідного рівня якості обслуговування;

2. Розробити метод динамічної реплікації сервісів в мережах з підтримкою функцій віртуалізації мережевих ресурсів;
3. Удосконалити метод формування та управління складом комплексних сервісів в мережах з підтримкою функцій віртуалізації;
4. Удосконалити метод балансування навантаження в мережі з підтримкою віртуалізації мережевих ресурсів;
5. Удосконалити метод динамічного розподілу пропускнуої здатності каналів передачі даних під час надання комплексних сервісів;
6. Провести дослідження працездатності та ефективності запропонованих в роботі рішень.

### **1.5. Висновки до першого розділу**

1. Проведений аналіз архітектури технології NFV і принципів формування та надання послуг кінцевим користувачам показав, що на сьогоднішній день не існує єдиних уніфікованих підходів ні до інфраструктури NFV, ні до методів управління процесами в ній. Документація ETSI і IETF носить рекомендаційний характер і не містить точних алгоритмів формування графів виведення послуг, що дозволяє розробникам впроваджувати в обладнання віртуалізації і комутації різні набори функцій.

2. Програмне і апаратне забезпечення NFV залишається пропрієтарним. В такому випадку, ефективна взаємодія зон NFVI, є досить складною: можливе виникнення помилок конфігурації та відсутність підтримки необхідних функцій. Дана ситуація дозволяє зробити висновок про необхідність розробки методів аналізу коректності функціонування і перевірки відповідності функціональності NFVI вимогам специфікації на ранніх етапах розробки NFV.

3. Розглянувши методи формування комплексних сервісів можна зробити висновок про подальші можливості в управлінні рівнем QoS за рахунок зміни складу комплексних сервісів. Зважаючи на це, існує потреба в розробці методу

аналізу характеристик мережевої інфраструктури в процесі надання сервісів для забезпечення необхідного рівня QoS.

4. Рівень управління послугами в мультисервісній мережі, побудованій на основі технології NFV, має власну структуру факторів, що впливають на показники QoS. У зв'язку з чим, необхідна розробка методів оцінки даних показників якості, як для атомарних сервісів, так і для комплексних.

5. Аналіз процесу формування графа перенаправлення послуг показав, що надання послуг з заданим рівнем якості можливе лише в разі узгодженої взаємодії компонентів мережевої інфраструктури і коректності рішень системи управління, а вимоги до даних процесів закладені на етапі формування SLA. Розробка методів перевірки відповідності всім вимогам специфікації і SLA дозволить підвищити ефективність процесу надання послуг за рахунок підвищення якості формування графа перенаправлення послуги.

## РОЗДІЛ 2

### МОДЕЛІ ПРОЦЕСІВ НАДАННЯ СЕРВІСІВ У МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ

Вирішення завдання аналізу і превентивна оцінка ефективності функціонування мереж з підтримкою функцій віртуалізації часто критична. Складність оцінки доступних віртуальних мережевих ресурсів в момент формування та імплементації сервісу, частота релокації сервісів, відсутність достатньої кількості інформації про технічні характеристики і конфігурації обладнання на стороні клієнта і демаркаційних зонах тягнуть за собою непередбачувані зміни в якості надаваних сервісів. Ситуація, що склалася, призвела до виникнення методик оцінки ефективності надання сервісів у мережах з підтримкою віртуалізації, яка базується на аналізі варіювання граничних значень показників якості, а також динаміки їх зміни в процесі надання сервісу «від краю до краю» в залежності від умов середовища передачі [9, 10, 52]. Подібні методики дозволяють вивчити природу зміни показників якості, а також виявити ряд чинників, які мають найбільший вплив на зміну (погіршення) показників якості.

Грунтуючись на рекомендаціях IETF [25] з метою точного визначення причин зміни (погіршення) показників якості, що надаються інфраструктурою NFV сервісів, необхідно визначити межі та оцінити характеристики мережевої інфраструктури двох зон обслуговування: зони замовника послуг NFV і зони провайдера сервісу VNF, а також допоміжних систем і інтерфейсів, що впливають на їх взаємодію.

На рисунку 2.1. показана діаграма надання сервісів кінцевим користувачам з зазначенням зон обслуговування.

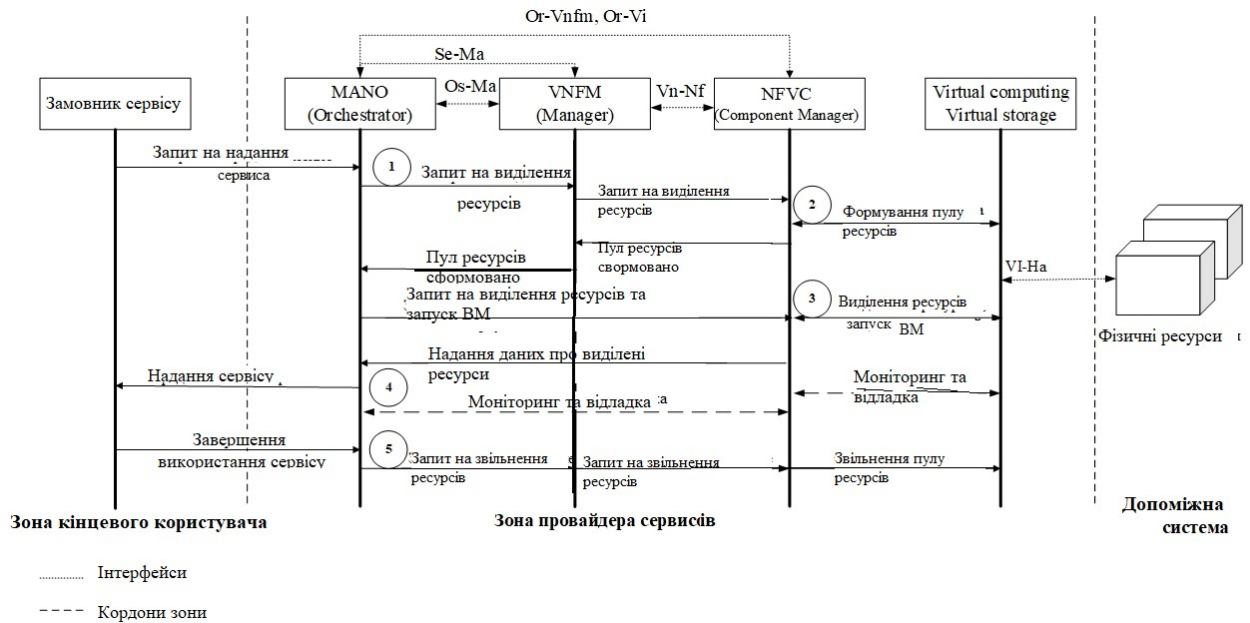


Рис. 2.1. Діаграма надання сервісів кінцевим користувачам

Якість послуг, що надаються за допомогою технології NFV, у першу чергу, залежить від блоків, що реалізують віртуальні мережеві функції (Network Function Component Virtual, NFVC) і системи управління віртуальними мережевими функціями (Virtual Network Function Management, VNFN). Характеристики показників якості є комплексними і включають в себе наступні компоненти:

- блок віртуалізації: віртуальні машини, сховища і обладнання (virtual computing, virtual storage);
- фізичні компоненти: мережеве обладнання і сховища даних;
- блок управління і оркестровки: елементи моніторингу та управління сервісами, віртуальними і фізичними ресурсами, що надається на стороні провайдера (MANO);
- блок підтримки бізнес-процесів і програмного забезпечення: елементи блоку забезпечують взаємодію елементів телекомунікаційної та програмної складової (NFVC, VNFN);
- база даних, що містить екземпляри сервісів і поточні характеристики мережі.

В процесі оркестрації і надання послуг основна увага приділяється такому показнику якості обслуговування, як затримка. Прийнято класифікувати типи затримок за інтерфейсами взаємодії. Виділяють такі типи затримок:

- затримка інтерфейсу Nf-Vi і Op-VM або мережеві затримки, які характеризують мережеву середу користувача і можуть бути розраховані як різниця між часом початку надання сервісу обчислювальним пристроєм (наприклад, фрагмент потокового відео відправлений) і часом отримання сервісу одержувачем (фрагмент потокового відео доставлений):  $D_{NF} = F(T_{NFVi}) - F(T_{Vfo})$
- затримка інтерфейсу Nf-Vi і Op-Vi або затримки обчислювальних ресурсів NFV, характеризуються тимчасовим інтервалом між зупинкою функціонування сервісу -  $T_{io}$  і його поновленням або перезапуском -  $T_{ii}$ :  
 $D_{VM} = F(T_{ii}) - F(T_{io})$ .

При оцінці сумарної якості надання сервісів запропоновано враховувати наступні характеристики:

- 1) продуктивність: ефективність функціонування віртуальних обчислювальних ресурсів / вузлів.
- 2) доступність: швидкість формування віртуалізованої мережевої інфраструктури і відновлення після збоїв.
- 3) надійність: відмовостійкість функціонування IaaS компонент.

Виділені вище показники можуть бути виражені кількісними і якісними параметрами QoS, такими, як швидкість обробки і час відгуку, що впливають на загальну швидкість надання сервісів, значення затримки і джиттера конфігурації примірника сервісу та мережевої інфраструктури [75, 76].

Основні параметри, що впливають на показники якості NFV в процесі надання сервісів наведені в таблиці 2.1 і на рисунках 2.2-2.4. Перший стовпець таблиці вказує на характеристику, другий стовпець містить етап життєвого циклу, в третьому стовпці описані прояви фактору.

Таблиця 2.1

*Надання сервісів за допомогою NFV. Фактори, що впливають на показники якості*

<b>Характеристика</b>	<b>Етап надання</b>	<b>Показники</b>
Доступність надання сервісу	Початкова оркестровка сервісу (аналіз статистичних та розподіл ресурсів, преконфігурація мережевих компонент)	Затримки обробки запитів, Міжінтерфейсні затримки, Затримки розподілу ресурсів
	Формування пула ресурсів	Час відгуку гіпервізора Час обробки запитів Час планування і розподілу ресурсів Тривалість простою обчислювальних ресурсів.
	Виділення ресурсів, розгортання / запуск віртуальної мережевої інфраструктури	Затримка в процесі виділення ресурсів Затримки обробки запитів Час відгуку компонент мережевої інфраструктури
	Функціонування віртуальної мережі / надання сервісів	Затримка обробки запитів, Тривалість простою обчислювальних ресурсів. Тривалість передачі даних
	Фінальна оркестровка (звільнення ресурсів)	Тривалість обробки запиту Тривалість зміни конфігурації ресурсів
Ефективність надання сервісу	Початкова оркестровка сервісу (пошук, розподіл і пре-конфігурація мережевих ресурсів)	Тривалість оркестровки сервісів у відповідності з SLA
	Формування пула ресурсів (синхронізація ресурсів)	Тривалість синхронізації обчислювальних вузлів і функцій управління

	Виділення ресурсів, розгортання / запуск віртуальної мережевої інфраструктури	Тривалість «розгортання» сервісів
	Функціонування віртуальної мережі/надання сервісів	Тривалість обробки запитів
	Фінальна оркестровка (звільнення ресурсів)	Тривалість релокації ресурсів
Надійність надання сервісу	Початкова оркестровка сервісу (аналіз статистичних та розподіл ресурсів, преконфігурація мережевих компонент)	Надійність виділених ресурсів (відношення успішної / неуспішної доставки сервісів, кількість втрачених даних)
	Формування пула ресурсів	Попередній розрахунок виконання/надання сервісу
	Виділення ресурсів, розгортання / запуск віртуальної мережевої інфраструктури	Резервування ресурсів
	Функціонування віртуальної мережі/надання сервісів	Тривалість відновлення при збої фізичних і віртуальних мережевих компонент (VM)
	Фінальна оркестровка (звільнення ресурсів)	Тривалість відновлення при збої чи виході з ладу обчислювальних компонент (VM) в процесі звільнення ресурсів



Рис. 2.2. Фактори, що впливають на показники якості надання сервісів за допомогою технології NFV



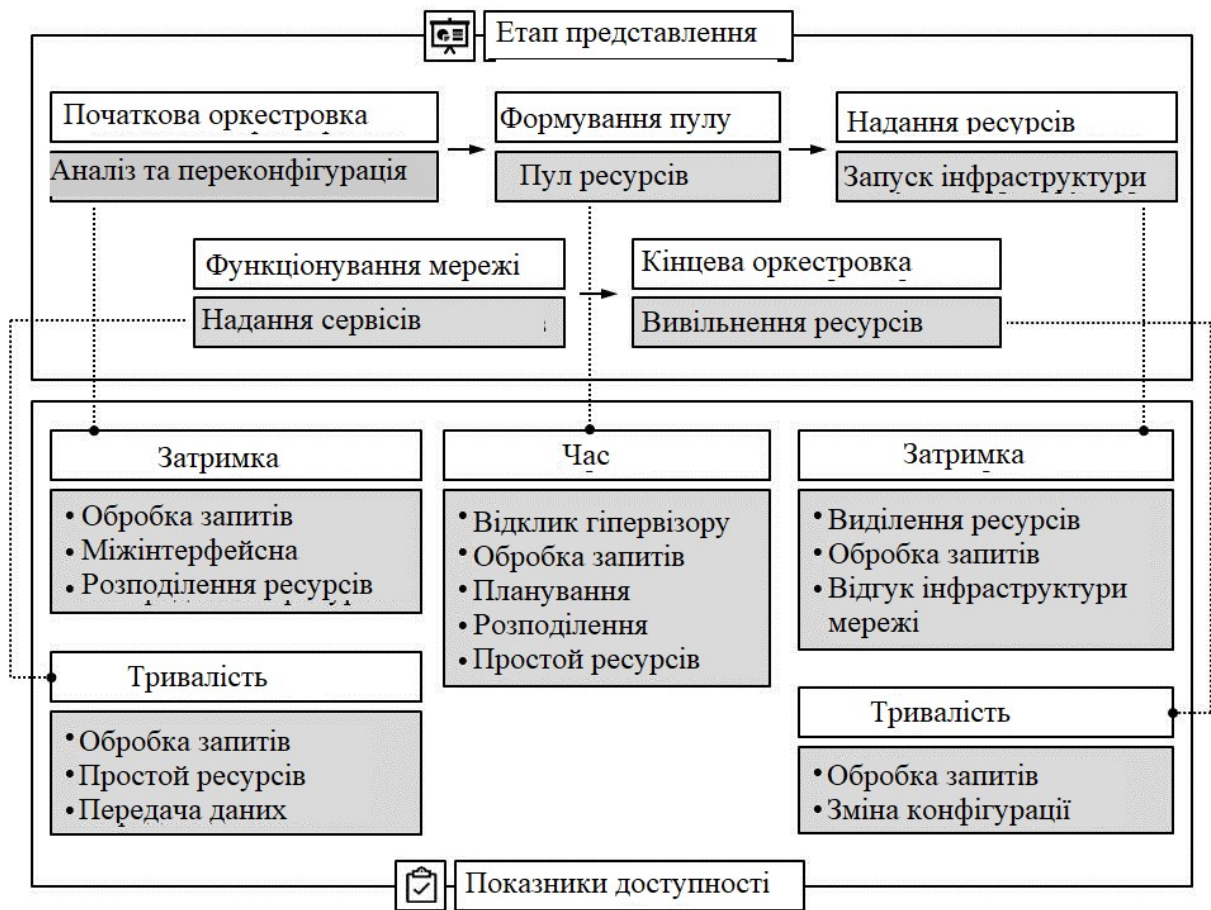


Рис. 2.3. Показники якості надання сервісів за допомогою технології NFV на рівні доступності

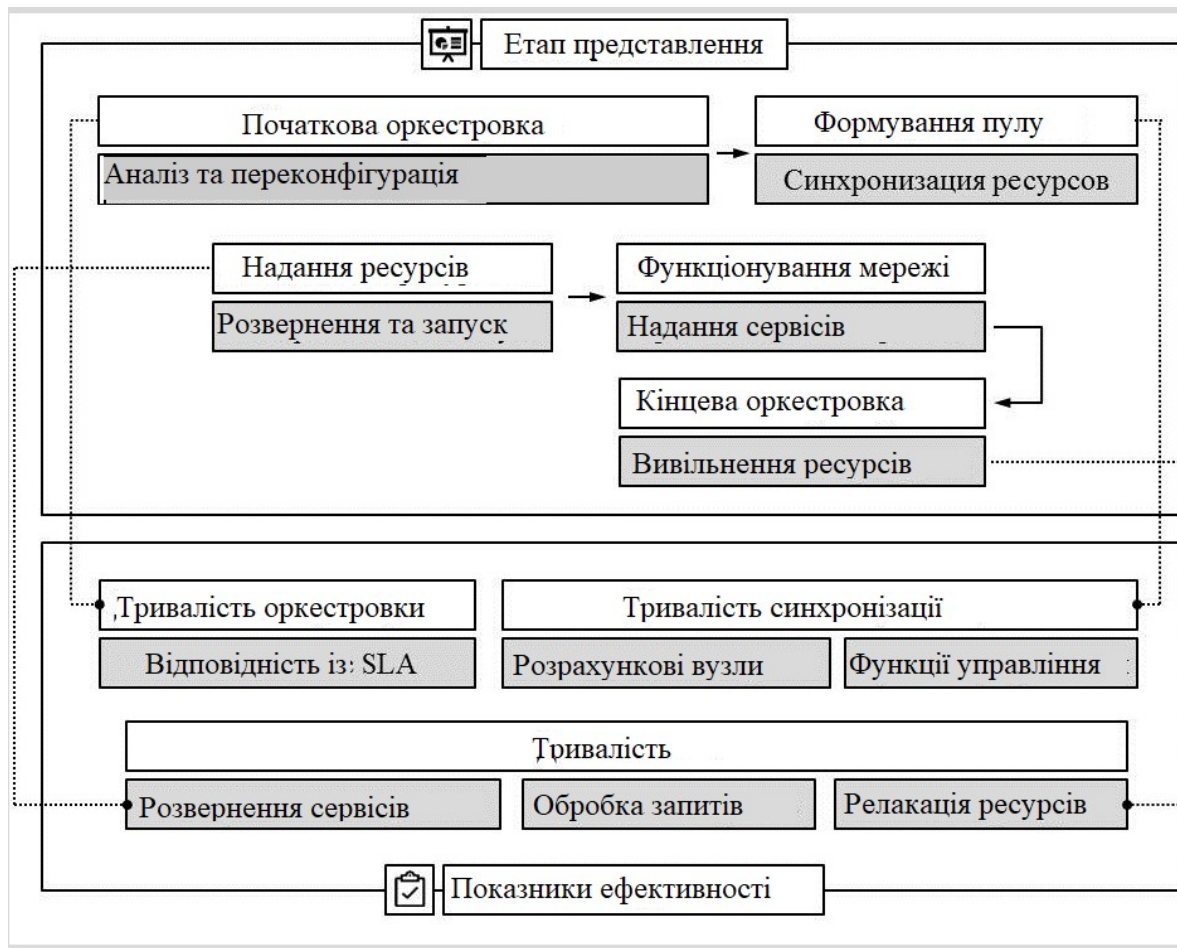


Рис. 2.4. Показники якості надання сервісів за допомогою технології NFV на рівні надійності.

Аналіз життєвого циклу сервісів в мережах з підтримкою віртуалізації мережевих функцій дав змогу визначити залежність QoS сервісів від нижчевказаних характеристик:

- тривалість формування і пре-конфігурації одного або декількох обчислювальних вузлів  $t_{conf}$  (час оркестровки); час інтеграції / запуску нового екземпляра сервісу або віртуального ресурсу  $t_{setup}$ ;
- функціонування ресурсу/надання сервісу  $t_{serv}$ ;
- час звільнення ресурсів і повторне розподілення мережевих ресурсів  $t_{reloc}$ .

Першочерговий вплив на  $t_{conf}$ ,  $t_{setup}$ ,  $t_{serv}$  мають затримки обробки даних і їх варіації. У зв'язку з динамічною оркестрацією сервісів і можливими релокаціями атомарних сервісів, що входять до складу комплексної послуги, в

якості аналізу показників якості сервісів запропоновано використовувати теорію мережевого обчислення.

## 2.1. Застосування елементів теорії мережевого обчислення, що застосовуються при моделюванні процесів надання сервісів

Перевагою теорії мережевого обчислення є моделювання поведінки мережевих елементів в рамках як безперервної, так і дискретної моделі часу. Її інструментарій дозволяє формалізувати і досліджувати залежності параметрів QoS як для композиції мережевих елементів, так і для кожного елемента окремо, а також визначати граничні значення показників якості, що ведуть до деградації сервісу. Визначення граничних показників і своєчасне визначення тенденції наближення до граничних показників дозволить елементам MANO, що управляють, прийняти міри по своєчасній релокації сервісів.

При моделюванні мережевої інфраструктури за допомогою теорії мережевого обчислення, визначення основних складових моделі і їх функціональних особливостей є першочерговим завданням.

Елементи реалізації NFV рішення при моделюванні відповідно до основами мережевої обчислення можуть бути представлені таким чином:

1. Вузли-обробники (сервери додатків, VM, віртуалізовані маршрутизатори, комутатори, файєрволи). Залежно від деталізації вузли-розробники можуть бути представлені одиничними елементами або їх комбінацією (однією VM або цілим кластером). Вузли-обробники характеризується кривою обслуговування (service curve), кривою відставання (backlogged / busy period) і затримки (delay) [112].

2. Крива навантаження  $I(\square t)$  – це функція, що зростає, яка характеризується відхиленням поточного обсягу  $V_{serv}(t_i)$  даних, що надходять в момент часу  $t_i$  від даних, що надійшли в попередній момент  $V_{serv}(t_i - t_0)$ .

Крива навантаження дозволяє оцінити зміну обсягу оброблюваних даних в певний момент часу, а, отже, і зміну часу обробки даних ( $t_{serv}$ ):

$$f(I(\square t)) = \frac{dI(t_i)}{dt} = \lim_{b \rightarrow y} \frac{(V_{serv}(t_i) + b) - V_{serv}(t_i - t_0)}{b}, \quad (2.1)$$

де  $b$  - інтенсивність збільшення / зменшення обсягу переданих даних,  $V_{serv}(t_i)$  - обсяг даних в бітах, які прийшли на вузол-обробник в момент часу  $t_i$ ,  $V_{serv}(t_i - t_0)$  - обсяг даних в бітах, які прийшли на вузол-обробник в момент часу  $t_i - t_0$ .

Зміни інтенсивності потоку даних, що надходять на кожен вузол-обробник, можуть бути оцінені за допомогою визначення граничних значень (глобальних мінімумів і максимумів) кривої навантаження. Верхнє граничне значення кривої вказує на максимальний обсяг потоку даних, і, в тому числі, на максимальну завантаженість каналу зв'язку.

Максимальні значення кривої навантаження можуть бути оцінені за допомогою використання наступних значень:

- зміна кривої навантаження протягом життєвого циклу сервісу

$$\int_{t=0}^j F_{V_{serv}}(t_i) \oslash F_{V_{serv}}(t_{serv});$$

- зміна кривої навантаження протягом періоду оркестровки

$$\text{сервісу} \int_{t=0}^j F_{V_{serv}}(t_i) \oslash F_{V_{conf}}(t_{conf});$$

- зміна кривої навантаження протягом періоду надання сервісу

$$\int_{t=0}^j F_{V_{serv}}(t_i) \oslash F_{V_{setup}}(t_{setup}).$$

Для аналізу процесу обробки пакетів, що надходять на обчислювальний пристрій, сформовано визначення кривої обслуговування, яка дозволяє описати швидкість обробки даних вузлом-обробником.

Крива обслуговування  $S(\square t)$  – це безперервна функція, яка визначає відхилення швидкості обробки даних на поточний момент часу від швидкості обробки даних вузлом-обробником в попередній часовий інтервал  $t_0$   $S(\square t): S(\square t) = R_{serv}(t_i) - R_{serv}(t_i - t_0)$ , де  $R$  - швидкість обробки даних.

У загальному вигляді крива обслуговування може бути представлена як:

$$f(S(\square t)) = \frac{dS(t_i)}{dt} = \lim_{b \rightarrow y} \frac{(R_{serv}(t_i) + b) - R_{serv}(t_i - t_0)}{b}. \quad (2.2)$$

Максимальне значення кривої обслуговування протягом часу надання сервісу ( $t_{setup}$ ) може бути оцінене наступним чином:

$$\int_{t=0}^j F_{R_{serv}}(t_i) \otimes F_{R_{setup}}(t_{setup});$$

При моделюванні вузлів-обробників і їх поведінки ключову роль відіграє властивість накопичення. Значення накопичення характеризується відношенням загального обсягу даних оброблених вузлом і тих, що прийшли на вузол за певний проміжок часу. В якості проміжку часу розглядається різниця між моментом ініціалізації сервісу та поточного моменту надання сервісу  $C(\square t)$ :

$$C(\square t) = \frac{S(\square t)}{I(\square t)}, \square t = t_i - t_0. \quad (2.3)$$

Запізнення при наявності інформації про обсяг даних, що зберігаються в буфері вузла і обсязі отриманих даних на момент часу може бути розраховане за такою формулою:

$$B(\square t) = I(\square t) \otimes S(\square t). \quad (2.4)$$

Максимальне значення запізнювання запропоновано розраховувати наступним чином:

$$\max B(\square t) = \sup_{0 < t \leq t_{end}} \{I(\square t) - S(\square t)\}. \quad (2.5)$$

де  $t_{end}$  - кінцеве значення інтервалу спостереження.

За допомогою запізнювання запропоновано оцінювати затримки обробки.

1. Вузли-споживачі. Кінцеве обладнання споживачів сервісу, що генерують запити на надання послуг. Вузли-споживачі характеризуються кривою навантаження. При моделюванні вузлів-обробників і потоків даних, що надають сервіс кінцевому користувачеві, більшість функцій можна охарактеризувати властивістю накопичення: функція є такою, що повертає на проміжку часу надання сервісу і може виражати загальну кількість інформації, обробленої і переданої вузлами з моменту ініціалізації сервісу (моменту відліку) до поточного моменту часу.

Безліч накопичувальних функцій характеризується безліччю спадаючих і безперервних функцій, графіки є невід'ємними і проходять через початок координат.

Аналіз кривої навантаження, обслуговування та затримки дозволяє оцінити зміни параметрів якості обслуговування. Для моделювання мережевих характеристик і з метою отримання достовірних результатів в теорії мережевого обчислення введені такі визначення.

Кожен потік даних характеризується власною кривою, яка залежить від типу переданих даних і їх інтенсивності.

Однією з вимог до архітектури NFV є успадкування всіх функціональних можливостей мультисервісних мереж. Таким чином, при застосуванні NFV технологій сервіси кінцевого користувача повинні передаватися з гарантованим рівнем якості, не нижче, ніж у традиційних мультисервісних мережах. Для організації та надання якісних сервісів використовуються такі мережеві політики, як DiffServ, IntServ, Traffic Shaping & Policing [57], механізми планування та балансування. В цьому випадку, для оптимального розподілу навантаження компонентів системи MANO необхідна інформація про кожного обчислювального вузла, динаміка зміни його параметрів в залежності від навантаження, що надходить, і часу обслуговування.

В теорії мережевих обчислень, для аналізу процесу обробки пакетів, що надходять на обчислювальний пристрій, сформовано визначення кривої обслуговування, яка дозволяє відобразити швидкість і обсяги оброблених

вузлом-обробником даних. Якщо крива прибуття дозволяє визначати динаміку зміни обсягів вхідних даних, то крива обслуговування - швидкість обробки і обсяги вихідних даних.

Криві навантаження і крива обслуговування дозволяють уявити узагальнену динаміку зміни характеристик сервіс-орієнтованої системи, зокрема NFVI, без урахування приватних показників затримки і втрати пакетів при проходженні потоку даних через вузли-обробники. Для кількісної оцінки таких показників якості як затримка і втрата пакетів в теорії мережевого обчислення введені визначення функції прибуття і функції відправлення.

Система MANO, як єдиний центр управління та моніторингу, приймає управлінські рішення на підставі інформації про криву навантаження і криву обслуговування. У процесі аналізу фрагмента мережевої інфраструктури від Nf-Vi до Op-Vi типу вузли-споживачі - вузли-обробники - пристрій надання сервісу вважаються логічними кінцевими і ланками, а вузли  $2 < n < N-1$  є проміжними вузлами-обробниками NFVM, найчастіше їх роль виконують маршрутизатори, шлюзи доступу, керовані і некеровані комутатори (рисунок 2.5) [18].

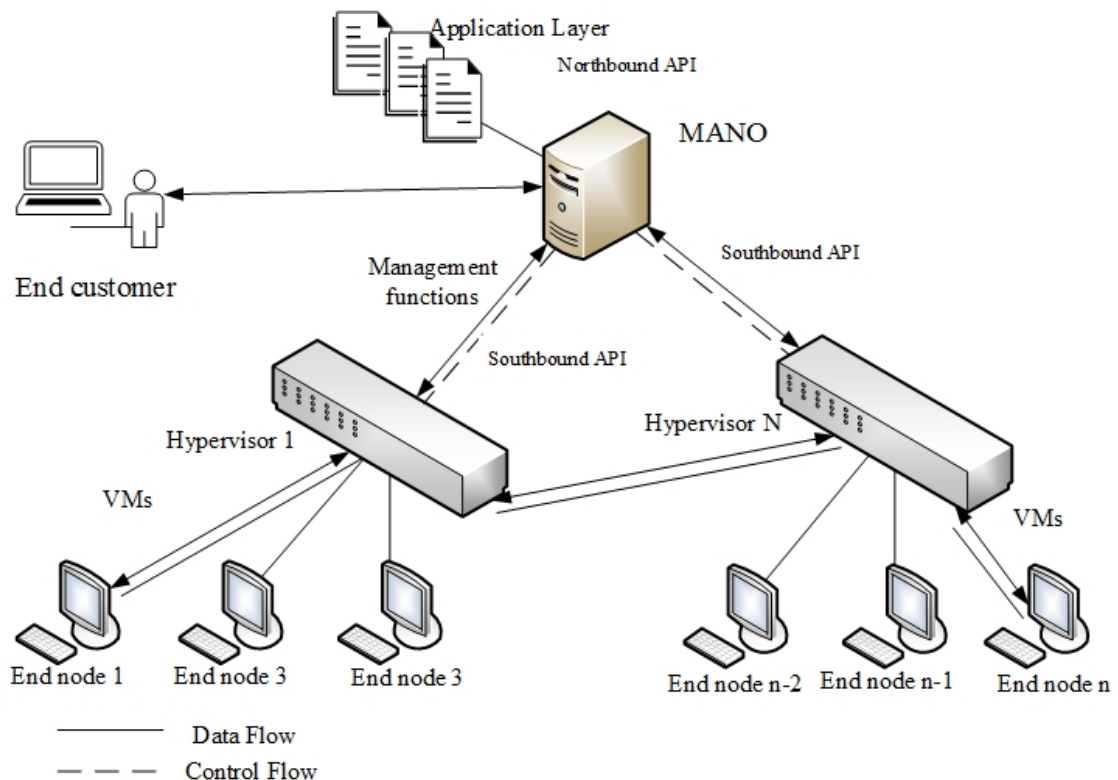


Рис. 2.5. Фрагмент мережі з підтримкою функцій віртуалізації

На рисунку 2.6 приведена модель вузла-обробника відповідно до визначень теорії мережевого обчислення.

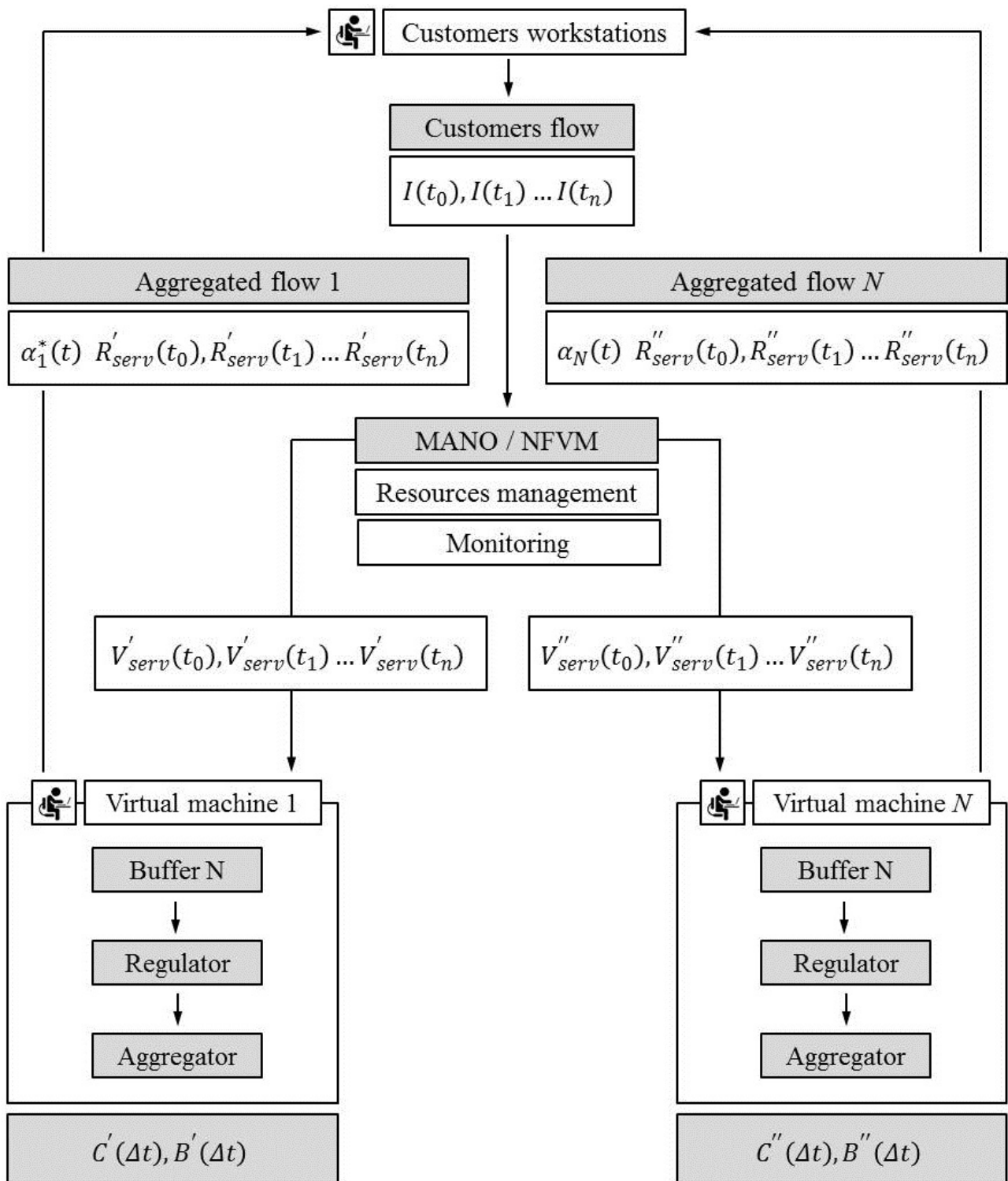


Рис. 2.6. Структурна схема фрагменту мережі: вузли- обробники з підтримкою функцій віртуалізації і елемент, що управляє



## 2.2. Моделювання процесу надання сервісів у мультисервісних мережах з підтримкою функцій віртуалізації за допомогою елементів теорії мережевого обчислення

З метою забезпечення гарантованої якості обслуговування в мережах з підтримкою NFV, системою MANO виконується безперервний моніторинг і контроль за станом інфраструктури NFV на всіх етапах життєвого циклу сервісу. У разі мультипоточності і агрегування каналів зв'язку системою MANO, передбачається застосування ряду механізмів резервування і розподілу ресурсів [41]. Однак досить важко оцінити якість обслуговування потоків всередині агрегованого каналу зв'язку. Таким чином, розробка методів аналізу QoS агрегованого потоку даних в мультисервісних мережах з підтримкою функцій віртуалізації є актуальною.

Мультисервісна мережа з підтримкою функцій віртуалізації представлена дворівневою архітектурою: рівнем доступу, який характеризується параметрами комутатора і рівнем управління, який характеризується параметрами контролера. Основною її особливістю є формування агрегованих потоків даних. Агрегація здійснюється відповідно до значення поля ToS вхідного трафіку [17]. Агрегування потоків даних здійснюється як на кінцевих управляючих пристроях - HyperVisor, так і на проміжних пристроях мережі - FlowVisor. Як FlowVisor, так і HyperVisor здійснюють перелік можливих дій лише на основі керуючих повідомлень, отриманих від MANO.

Віртуальні мережеві елементи здатні обслуговувати кілька потоків даних в паралель  $\sum_1^m F_j(t_i)$ . При цьому кожен потік  $j$ , який обслуговується цим пристроєм характеризується своїм власним набором характеристик:  $I'(t), C'(\square t), B'(\square t)$ .

Якість обслуговування потоків даних, а, отже, і надання сервісів кінцевим користувачам залежать від розподілу обчислювальних ресурсів вузла. Ефективність обчислювальних ресурсів (вузлів-обробників) може бути

представлена у вигляді  $Q(S(t)) \rightarrow \max_{CP_\alpha}$ , де  $Q()$  – цільова функція,  $CP_\alpha$  – дисципліна обслуговування.

Продуктивність вузлів-обробників при наявності декількох потоків даних у відповідності з [41, 46] залежить від наступних показників: максимальної швидкості обробки -  $p$ , середньої швидкості обробки -  $r$ , інтенсивності потоку провідних даних -  $b$ , максимального розміру пакетів -  $M$ , обсягу вхідного буфера -  $q$ , дисципліни обслуговування -  $Q$ .

Безліч  $(p, r, M, b)$  формує специфікацію потоку даних T-SPEC, безліч  $(R, q, Q)$  формує специфікацію обробника [47].

На рисунку 2.7 приведена модель фрагмента мережевої інфраструктури, для якої характерна передача агрегованих потоків даних.

Кожен потік даних  $(R1-R3)$ , що надходить з вузла-споживача або попереднього вузла-обробника має свої власні характеристики і задається індивідуальною кривою навантаження  $(I_1(t_i), I_2(t_i), I_3(t_i))$ . У разі агрегації сумарний мультипотік, що містить безліч потоків, які належать одному класу обслуговування, визначається узагальненими характеристиками і повинен бути описаний узагальненою кривою навантаження. Вид кривої прибуття залежить від специфікації класу трафіку і характеризується узагальненими показниками  $(\max p, \max r, \max M, \max b)$ :  $F_N(t_{setup}) = \max_{n=1,2,3} pt + \max_{n=1,2,3} b$ .

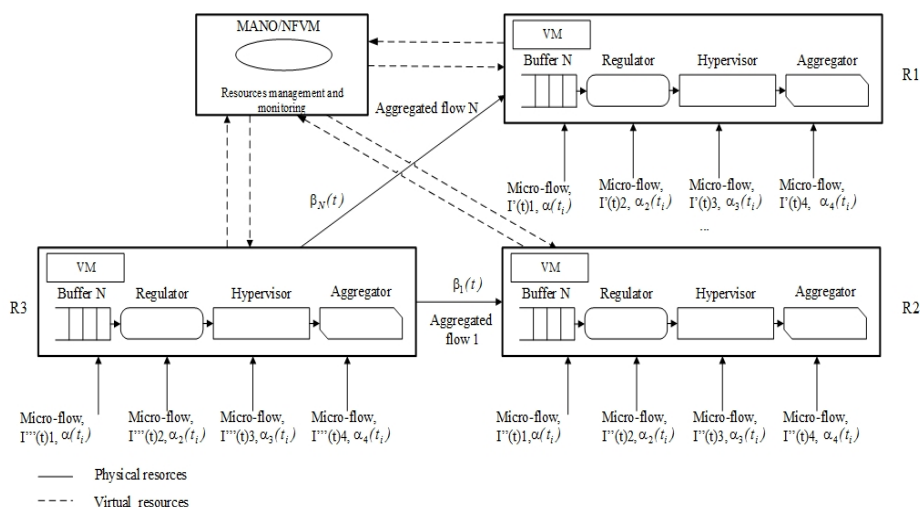


Рис.2.7. Структурна схема при агрегуванні декількох потоків даних

При обробці агрегованого потоку вузлом-обробником кожен потік має свій пріоритет обслуговування, який визначений ToS [41, 46]. Від пріоритету обслуговування залежить політика розподілу обчислювальних ресурсів при обробці кожного потоку (швидкість обробки, значення затримки, час запізнювання). Швидкість обробки вузлом-обробником потоку, в залежності від його пріоритету, може бути оцінена за допомогою наступного рівняння:

$$r_i = \frac{\phi_i B_N(t_i)}{\sum_{j=0}^n \phi_j} S_N(t_i), \quad (2.6)$$

де  $\phi_i$  - пріоритет потоку даних при обслуговуванні вузлом-обробником,  $S_N(t_i)$  - поточне значення продуктивності  $N$ -го вузла-обробника,  $B_N(t_i)$  - затримка при обробці потоків даних.

Зміна інтенсивності агрегованого потоку, що проходить через обробник кривої навантаження, розрахована по відношенню до швидкості обробки, може бути визначена за формулою:

$$Tb(s) = \frac{1}{S_N(t_i)} \sum_{i=1}^M b_i - b_{i-1}. \quad (2.7)$$

Максимальна зміна значення обробки визначається як  $\max Tb(s) = \max_N \{ Tb \}$ .

1. Максимальний розмір переданого через мережу пакета дорівнює  $M_{max}$ .

Для фрагмента мережі, наведеного на рисунку 2.7, характерні наступні твердження:

- максимальна затримка, яку буде відчувати потік найвищого класу сервісу, що передається через мережу, обмежена зверху значенням

$N \cdot \sum_{i=1}^n B_i(\square t)$ , де  $N$  – кількість вузлів-обробників,  $B_i(\square t)$  - максимальна затримка

$i$ -того вузла.

- мінімальний розмір буфера вузла-обробника, необхідний для передачі всіх пакетів, що надійшли в агрегованому потоці, обмежений знизу значенням  $P_N(t) \cdot T_b(t) + M_{max} p$ .

У разі, якщо результуюча максимальна продуктивність, перевищує граничне значення, то мережева інфраструктура може бути модифікована повторно.

Крива обслуговування мережі, що програмноконфігурується, формується за допомогою наступного рівняння:

$$\beta(\tau) = \sum_{k=1}^n \beta_k(\tau), \forall \tau > 0, \quad (2.8)$$

де  $\beta_k(\tau)$  - крива обслуговування, що формується кожним окремим вхідним потоком даних,  $n$  - кількість вхідних потоків.

Відповідно до (7) і (8) крива обслуговування для мікро-потoku може бути сформована за допомогою наступного рівняння:

$$\beta_i(\tau) = \beta(\tau) - \sum_{k=1, k \neq i}^n \alpha_k(\tau - \tau_k),$$

$$\forall \tau > 0, \tau > \tau_k > 0 \quad (2.9)$$

де  $\sum_{k=1, k \neq i}^n \alpha_k(\tau - \tau_k)$  - сумарна крива навантаження для вхідних мікро-потоків, за винятком потоку  $i$ ,  $\tau_k$  - час запізнювання, що вноситься при обробці вузлом сумарного потоку  $\alpha_k(\tau)$ .

Вихідна з вузла-обробника крива навантаження, яка в подальшому буде застосована як вхідна крива навантаження для подальших вузлів-обробників може бути представлена як [4]:

$$\alpha^*(\tau) = \begin{cases} \frac{b - \frac{M}{t}}{p - r} \leq T_h, & \alpha^*(\tau) = b + r(|D(t)| + t); \\ \frac{b - \frac{M}{t}}{p - r} \geq T_h, & \alpha^*(\tau) = \min \left\{ R(|D(t)| + t) + \frac{b - \frac{M}{t}}{p - r} (p - R)^+, b + r(|D(t)| + t) \right\}. \end{cases}$$

(2.10)

Допущення, що застосовуються при моделюванні: в якості регулятора розглянуто регулятор типу Leaky Bucket [4, 17, 18].

```

Flowclassification:
 $\alpha_{low}[l1] \leftarrow \infty$ 
 $\alpha_{high}[l1] \leftarrow 0$ 
 $\alpha_{low}[l2] \leftarrow 0$ 
 $\alpha_{high}[l1] \leftarrow \infty$ 
for all  $I_i \in g(l1, l2)$  do
  if  $l = \text{firstlink}(f)$  then
    if  $f \in I_1$  then
       $\alpha_{low}[l1] \leftarrow \alpha_{low}[l] + \alpha(f)$ 
    else
       $\alpha_{high}[l2] \leftarrow \alpha_{high}[l] + \alpha(f)$ 
    end if
  end if
end for
  for all  $l = \text{pred}(f)$  do
     $\alpha_{low}[l1] \leftarrow \alpha_{low}[l] + \alpha(f) + \max_{n=1,2,3} pt + \max_{n=1,2,3} b$ 
  else
     $\alpha_{high}[l2] \leftarrow \alpha_{high}[l] + \alpha(f) + \max_{n=1,2,3} pt + \max_{n=1,2,3} b$ 
  end for
  for all  $\beta$  do
     $\beta_{low}[l1] \leftarrow [\alpha_{low}[l] + \alpha(f) + \max_{n=1,2,3} pt + \max_{n=1,2,3} b]^+$ 
     $\beta_{high}[l1] \leftarrow [\alpha_{high}[l] + \alpha(f) + \max_{n=1,2,3} pt + \max_{n=1,2,3} b]^+$ 
  end for
output:  $(\alpha_{low}[l], \beta_{low}[l], \{l\}), (\alpha_{high}[l], \beta_{high}[l], \{l\})$ 

```

Приведена вище модель поведінки мультисервісної мережі з підтримкою функцій віртуалізації обчислює кореневий вузол-обробник (віртуально-обчислювальний прилад) і набір маршрутів, які застосовуються для передачі потоків даних з низьким і високим пріоритетом обслуговування відповідно.

Перевагою даного підходу є можливість обліку повторного використання і перевизначення обчислювальних ресурсів.

Спочатку обчислюються значення кривих навантаження, визначених як  $\alpha_{low}$  і  $\alpha_{high}$  для низького і високого пріоритету обслуговування відповідно, що дозволяє забезпечити комплексний аналіз розподілу навантаження, а, отже, своєчасно регулювати політику надання сервісів у MANO.

Далі розраховується маршрут надання сервісу з тривалістю надання сервісу на кожному обчислювальному пристрої - формується крива обслуговування. На підставі даних про криві навантаження і кривих обслуговування формується первинний сценарій надання сервісу. Наступні сценарії формуються шляхом рекурсивного розрахунку їх кривих прибуття в якості виходу, що межують з попередніми. В якості вихідних даних формуються криві обслуговування і набір маршрутів надання сервісів.

### 2.3. Аналіз затримки «від краю до краю» в процесі надання сервісів у мультисервісних мережах з підтримкою функцій віртуалізації

Аналіз життєвого циклу сервісів в мережах з підтримкою технології NFV показав, що ключовий вплив на якість надаваних сервісів мають затримки, що вносяться як на кожному з етапів надання сервісу, так і «затримки від краю до краю». Вплив затримок «від краю до краю» найбільш великий безпосередньо на етапі надання сервісу.

Верхнє граничне значення затримки обчислюється таким чином:

$$d = \alpha(t_1) - \beta(t) = (rt + b) - \frac{\sum_{k=1}^l b_k}{R - \sum_{k>l}^n r}. \quad (2.11)$$

Верхнє значення довжини буфера при певних характеристиках комутатора може бути розраховане наступним чином:

$$b_i = \sum_{k=1}^n (b_k + r_k t) - (R - \sum_{k=1, k \neq i}^n r_k) t. \quad (2.12)$$

Якщо OpenFlow комутатор здійснює перенаправлення вхідного запиту від вузла-споживача контролера, то значення кривої обслуговування змінюється. У цьому випадку має місце наступна формула:

$$\beta(\tau_+) = [\beta_c(\tau_+) - \sum_{k=1}^l \alpha^*(\tau)]^+, \quad (2.13)$$

де  $\tau_+$  - часовий інтервал з урахуванням отримання відповіді від контролера, як правило  $\tau_+ \equiv \tau$ , тобто час обробки OpenFlow запиту контролером мізерно малий,  $\beta_c(\tau_+)$  - крива обслуговування контролера,  $\sum_{k=1}^l \alpha^*(\tau)$  - крива навантаження, що виходить з вузла-обробника.

Максимальне значення затримки може бути розраховане наступним чином:

$$\begin{aligned} d &= \sup \{ \sum_{k=1}^n \alpha_k(\tau) - [\beta_c(\tau_+) - \sum_{k=1}^l \alpha^*(\tau)]^+ \} = \\ &= \frac{\sum_{k=1}^l b_k}{R - \sum_{k=1, k \neq i}^n \sum_{k=i}^l r_k} + \frac{\sum_{k=1, k \neq i}^n \sum_{k=i}^l b_k}{R_c} + (\tau - \tau_+) \end{aligned} \quad (2.14)$$

Значення кривої навантаження, що виходить з вузла-обробника, розраховується відповідно до формули [2, 19]:

$$\alpha(\tau) = \overline{Sp}(\alpha^0(\tau) \otimes \beta^0(\tau)) + \beta_c(\tau), \quad (2.15)$$

де  $\overline{Sp}$  - коефіцієнт масштабування, отриманий на основі роботи регулятора.

Припустимо, що значення запізнювання і максимальна довжина буфера не залежить від наявності зворотного зв'язку.

Нехай потік проходить через кілька вузлів-обробників. Загальний фрагмент мережі складається з  $N$  вузлів-обробників. Кожен обробник визначено індивідуальним набором характеристик, що впливають на його продуктивність.

Припустимо, що в якості регулятора потоку даних на кожному вузлі-обробнику застосуємо регулятор Leaky Bucket і крива навантаження, що виходить з вузла-обробника, характеризується рівнянням (2.12) дисципліна обслуговування - First In First Out [20].

Крива обслуговування кожного наступного вузла обробника підпорядковується наступній закономірності:

$$\begin{aligned}
\beta^0(\tau) &= [\beta(\tau) - \sum_{k=1}^N \alpha^{(k)}]^+; \\
\beta^1(\tau) &= [\beta(\tau) - \sum_{k=2}^N \alpha^{(k)}]^+; \\
&\dots \\
\beta^N(\tau) &= [\beta(\tau) - \sum_{k=2}^N \alpha^{(k)}]^+,
\end{aligned}
\tag{2.16}$$

де крива навантаження, що виходить з кожного вузла-обробника, визначена наступним чином:

$$\begin{aligned}
\alpha^0(\tau) &= \alpha(\tau); \\
\alpha^1(\tau) &= \overline{\text{Sp}}_1(\alpha^0(\tau) \otimes \beta^0(\tau)) + \beta_c(\tau); \\
&\dots \\
\alpha^N(\tau) &= \overline{\text{Sp}}_N(\alpha^{(N-1)}(\tau) \otimes \beta^{(N-1)}(\tau)) + \beta_c(\tau).
\end{aligned}
\tag{2.17}$$

Значення затримки «від краю до краю» в цьому випадку може бути розраховане за допомогою наступного рівняння:

$$\begin{aligned}
d &= ((rt + b) - \frac{\sum_{k=1}^1 b_k}{R - \sum_{k>1}^n r_0})^0 + (\alpha^1(\tau) - \frac{\sum_{k=1}^1 b_k}{R - \sum_{k>1}^n r_1}) + \\
&+ (\alpha^2(\tau) - \frac{\sum_{k=2}^1 b_k}{R - \sum_{k>1}^n r_2}) + \dots + (\alpha^3(\tau) - \frac{\sum_{k=2}^1 b_k}{R - \sum_{k>1}^n r_3})
\end{aligned}
\tag{2.18}$$

і не може перевищувати наступне значення:

$$\forall t : d = \sup \{ \sum_{k=1}^N \alpha_k(\tau) - [\beta_c(\tau_+) - \sum_{k=1}^N \alpha^N(\tau)]^+ \}.
\tag{2.19}$$

Значення буфера кожного вузла-обробника при цьому залишається постійним.

Наведена математична модель, заснована на теорії мережевого обчислення, дозволяє розрахувати і проаналізувати граничні значення затримки.



## 2.4. Висновки до другого розділу

1. Аналіз особливостей системи MANO і NFVC, NFVM на різних етапах життєвого циклу сервісу показав, що тимчасові показники якості, такі як затримка і її варіації, час відгуку мають критичний вплив на рівень наданої якості обслуговування. Відзначено, що аналіз і моделювання показників даного типу є досить трудомістким завданням.

2. Для моделювання динаміки мережевих характеристик в процесі надання сервісів кінцевим споживачам запропоновано застосовувати теорію мережевого обчислення. Теорія мережевого обчислення дозволяє оперувати як точними значеннями мережевих характеристик, так і визначати їх порогові значення. Подібний підхід дозволяє розраховувати порогові показники якості надання сервісів і відстежувати динаміку деградації якості.

3. Запропоновано моделювання вузлів-обробників NFVC в рамках мережевого обчислення, розглянуті сценарії обробки одного потоку даних і агрегованого типу трафіку. Наведено модель процесу надання сервісів з урахуванням агрегації потоків даних, що дозволяє уникнути зниження ефективності мережевої інфраструктури при релокації обчислювальних ресурсів і повторному використанні вузлів-обробників.

4. На основі особистої моделі мережі запропонована модель розрахунку граничних показників затримки і продуктивності «з кінця в кінець» при наявності декількох вузлів-обробників. Дана модель передбачає наявність одного агрегованого потоку даних, однак на базі рівнянь (2.11) - (2.17) може бути запропоновано її подальший розвиток, що дозволяє моделювати поведінку вузлів-обробників в разі формування декількох агрегованих потоків даних.

## РОЗДІЛ 3

### МОДЕЛІ І МЕТОДИ АНАЛІЗУ І УПРАВЛІННЯ ЯКІСТЮ СЕРВІСІВ У МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ

Основною особливістю мереж з віртуалізацією мережевих функцій є динамічне масштабування, швидка адаптація, автоматизоване конфігурування мережевого обладнання, а також підтримка механізмів віртуалізації різного виду сервісів. Результат аналізу функціонування мереж з підтримкою функцій віртуалізації, зокрема NFV, показав, що, механізми MANO не є досконалыми: обмеженість продуктивності серверів, розрізненість застосовуваних операційних систем і алгоритмів функціонування гіпервізора часто призводять до необґрунтованого збільшення обсягів даних, які зберігаються і передаються, що, в свою чергу, спричиняє виникнення мережевих затримок і збільшення часу простою обладнання.

#### **3.1. Оцінка якості надання комплексних сервісів в мережах з підтримкою віртуалізації функцій**

Відповідно до рекомендацій RFC 7149 і IETF P1904.3, а також ряду інших рекомендацій [44, 73] для оцінки показників якості комплексних сервісів запропоновано використовувати набір наступних показників:

- надійність сервісу  $Rel(s)$ ,
- час відгуку при формуванні комплексного сервісу  $T(s)$ ,
- доступність сервісу  $Av(s)$ .
- вартість надання комплексного сервісу  $Cost(s)$ ,

Аналіз значень даних показників і своєчасне забезпечення їх заданого рівня дозволяють задовольнити вимоги користувачів до якості обслуговування.

Аналіз показників якості комплексних сервісів, що надаються кінцевим користувачам, характеризується значенням функції ефективності. Функція ефективності може бути задана у вигляді наступного рівняння [105]:

$$Eff(s) = Rent(s) - W(N(c,m), L(c,m)) - P(N,L), \quad (3.2)$$

де  $Rent(s)$  - функція, що характеризує рентабельність надання сервісу,  $W(N(c,m), L(c,m))$  - функція витрат на обслуговування мережевої інфраструктури.  $W(N(c,m), L(c,m))$  залежить від таких показників, як вартість налаштування ( $c$ ) і підтримки ( $m$ ) мережевого обладнання  $N$  і каналів зв'язку  $L$ ,  $P(N,L)$  - штрафна функція, залежить від значення штрафу за порушення показників QoS, погоджених у договорі про надання послуг (сервісів).

У загальному випадку функція ефективності, наведена у рівнянні (3.2) залежить від таких показників, як:

- ефективність каналу передачі ( $R_{sum\_Mij}$ );
- інтенсивності запитів на надання комплексного сервісу ( $\lambda_{S_i}$ );
- показників якості обслуговування ( $ToS$ ).
- кількість обчислювальних елементів, які беруть участь у формуванні сервісу ( $\sum_{i=1}^m J_{si} + \sum_{j=1}^m J_{oj}$ );
- питома вага кожного обчислювального елементу в процесі надання комплексного сервісу ( $k_{N_{si}}$ );

Роль обчислювальних елементів, які беруть участь в процесі формування комплексного сервісу, може бути визначена за допомогою наступного рівняння:

$$k_{N_{si}} = \begin{cases} N_{si} = 1, \text{ якщо використовується в ролі основного сервісу;} \\ 0 < N_{si} < 1, \text{ якщо використовується в ролі джерела одиничного сервісу;} \\ N_{si} = 0, \text{ якщо сервіс не використовується.} \end{cases} \quad (3.3)$$

Ефективність передачі даних обраним каналом визначається за допомогою наступного виразу:

$$A^{IN} \rightarrow \max_{CP_{\alpha}} | R(\sum Mt_{ij}), 0 \leq Mt_{ij} \leq 1, \quad (3.4)$$

де  $A^{IN}$  – сумарна інтенсивність,  $R(\sum Mt_{ij})$  – сумарна пропускна здатність міжвузлами  $N_{si}$  и  $N_{sj}$ ,  $CP_{link}$  – політика управління каналом.

При цьому, вибір каналу передачі даних може бути реалізований за допомогою наступної системи рівнянь:

$$R_{sum\_Mt_{ij}} = \begin{cases} Mt_{ij} = 1, \text{ канал використовується в якості основного,} \\ 0 < Mt_{ij} < 1, \text{ канал використовується в якості допоміжного,} \\ Mt_{ij} = 0, \text{ канал не використовується.} \end{cases} \quad (3.5)$$

Максимізація значень функції ефективності композиції сервісів виконується за рахунок підвищення показників якості обслуговування відповідно до RFC 7149 [69] і IETF P1904.3 [41].

Вартість комплексного сервісу залежить від структури сервісу та принципів його формування. Так, при наявності послідовної взаємодії одиничних сервісів вартість буде визначатися сумою вартостей всіх складових компонентів:

$$Cost(s) = \sum_{i=1}^m Cost(s_i) . \quad (3.6)$$

де  $m$  – загальна кількість одиничних сервісів.

Можливий випадок, коли затребуваними будуть не всі сервіси. В такому випадку тільки один з доступних одиничних сервісів  $(s_1, s_2, \dots, s_n)$  буде

затребуваний. Тоді  $p_i$  – ймовірність того, що  $s_i$  буде вибрано відповідно до поточного стану мережі. У цьому випадку вартість комплексного сервісу визначається сукупною вартістю одиничних сервісів, обраних відповідно до асоційованої з нею ймовірністю:

$$\text{Cost}_{od}(s) = \sum_{i=1}^m \text{Cost}(s_i) \cdot p_i. \quad (3.7)$$

Сервіси також можуть надаватись у відповідності до певних розроблених циклів. В цьому випадку необхідно використовувати умови виходу із циклу. В цьому випадку вартість комплексного сервісу буде дорівнювати добутку вартості виконуваного сервісу на кількість звернень до нього в циклі:

$$\text{Cost}_{цикл}(s) = \text{Cost}(s_i) \cdot l \quad 0 \leq l \leq h, \quad (3.8)$$

де  $l$  – кількість звернень до сервісу,  $h$  – максимальна кількість звернень.

Також може використовуватись схема надання відмовостійкого сервісу. В цьому випадку еквівалентні сервіси будуть виконуватися паралельно, при цьому будуть використані результати тільки того сервісу, час виконання якого буде меншим. Тоді вартість комплексного сервісу буде дорівнювати сумі вартостей усіх виконуваних паралельно сервісів:

$$\text{Cost}_{нар}(s) = \sum_{i=1}^m \text{Cost}(s_i). \quad (3.9)$$

Одним із важливих параметрів, які необхідно визначати, є час відгуку, що характеризує інтервал часу між моментом відправки запиту на надання сервісу і моментом отримання відповіді на запит. У разі, якщо комплексний сервіс є послідовним виконанням декількох одиничних сервісів, час відгуку може бути задано, як:

$$T_{посл}(s) = \sum_{i=1}^n T(s_i). \quad (3.10)$$

При паралельному виконанні ряду одиничних сервісів час відгуку буде визначатися максимальною затримкою, що вноситься одним із сервісів.

$$T_{\text{нап}}(s) = \max_i \{T(s_i)\} . \quad (3.11)$$

У разі застосування умовного оператора час відгуку буде визначатися часом відгуку сервісу, обраним відповідно до асоційованої з ним ймовірності:

$$T_{\text{ум}}(s) = \sum_{i=1}^n T(s_i) \cdot p_i , \quad (3.12)$$

де  $p_1 + p_2 + \dots + p_n = 1$ . При цьому, тільки один із сервісів може бути відібраний.

У разі наявності циклів в структурі комплексного сервісу час відгуку буде залежати від кількості звернень  $l$  до атомарного сервісу:

$$T_{\text{цикл}}(s) = T(s_i) \cdot l, \quad 0 \leq l \leq h . \quad (3.13)$$

При використанні відмовостійкого оператора часу відгуку виконуваних паралельно сервісів можна визначити наступним чином:

$$T_{\text{відм}}(s) = \min_i \{T(s_i)\} . \quad (3.14)$$

При наявності послідовно виконуваних одиничних сервісів в структурі комплексного сервісу надійність може бути визначена наступним чином:

$$\text{Rel}_{\text{носл}}(s) = \prod_{i=1}^m \text{Rel}(s_i) . \quad (3.15)$$

У разі паралельного виконання надійність комплексного сервісу визначається добутком надійності одиничних сервісів, що беруть участь в її формуванні:

$$\text{Rel}_{\text{нап}}(s) = \prod_{1 \leq i \leq l} R(s) . \quad (3.16)$$

При використанні відмовостійкого оператора надійність буде визначатися за формулою:

$$Rel(s) = 1 - \prod_{i=1}^n (1 - Rel(s_i)). \quad (3.17)$$

З точки зору користувача дуже важливою характеристикою є доступність сервісу, що характеризує здатність мережевої інфраструктури коректно надання сервісу при всіх можливих мережевих обмеженнях [50]. Доступність характеризується відношенням часу, коли сервіс перебуває в стані готовності до використання до загального часу, протягом якого відбуваються вимірювання і може бути представлена у вигляді рівняння (1.1).

Так як доступність і надійність є ймовірностями виникнення певної події, то методика їх обчислення аналогічна.

Таким чином, маючи інформацію про показники якості, а також сценарії включення окремих сервісів в комплексний, можна визначити інтегральний показник  $QoS_{\text{інт}}$  для будь-якого комплексного сервісу.

### **3.2. Метод формування комплексних сервісів з урахуванням показників якості обслуговування**

Для того, щоб охопити весь спектр рішень щодо забезпечення необхідної якості обслуговування, необхідно розробити метод підвищення рівня QoS за відсутності послуг з необхідним рівнем якості обслуговування.

Забезпечення необхідних показників якості часто зводиться до вирішення оптимізаційної задачі при формуванні сервісів [77, 80, 83].

Для покриття всієї області рішень по забезпеченню заданої якості обслуговування необхідна розробка методу підвищення рівня QoS, в разі відсутності сервісів з необхідним рівнем якості обслуговування.

Для вирішення поставленого завдання в роботі запропоновано математичний метод формування розподіленої комплексного сервісу на базі інформації про наявні в мережі атомарні сервіси.

В основі запропонованого методу лежить ідея формування комплексного сервісу  $CS$  (Complex Service), що складається з декількох сервісів з безлічі доступних сервісів, показники якості яких нижче заданих норм, одночасне використання яких дозволить підвищити доступність сервісу:

$$CS_i \subset SS, \quad SS_i = \{CS_1, CS_2, \dots, CS_i, \dots, CS_n\}, \quad (3.19)$$

$$CS_i = \{S_w, S_{w+1}, \dots, S_{w+y}\},$$

де  $i$  – номер комплексного сервісу в множині  $CS$ ,  $w$  – номер сервісу у вибраній підмножині  $P_i$ ,  $z$  – кількість сервісів в підмножині  $P_i$ ,  $y$  – будь-яке число від 1 до  $(z-w)$ , при цьому можлива ситуація, коли  $CS_i \cap CS_{i+1} \neq \emptyset$ .

Розподілений комплексний сервіс може бути представлений у вигляді об'єднання безлічі сервісів  $P_i$ . При цьому, QoS параметри атомарних сервісів  $P_i$  можуть мати значення нижче необхідних, проте, за рахунок їх одночасного використання, сумарні показники продуктивності розподіленого сервісу зростають.

Метод формування розподіленого сервісу дозволяє сформувати сервіс з заданими значеннями параметрів якості у випадку, коли відсутній сервіс з необхідними показниками якості в мережі з підтримкою технології віртуалізації. Даний алгоритм зображено на рис. 3.1.

Значення врахованих параметрів QoS представлені безліччю  $L$ . Інформація про показники QoS атомарних сервісів витягується з модифікованої UDDI реєстру [25].

Безліччю  $Q$  представлено значення параметра QoS, яке не задовольняє вимогам користувача. Безліч  $Q$  належить множині  $L$ ,  $Q \subset L$ . Безліч  $D$  також належить множині  $L$ ,  $Z \subset L$  і включає в себе всі параметри даної множини, за виключенням параметра, що належить множині  $Q$ ,  $Z = L \setminus Q$ .

*Початок;*

*Вибірка параметрів QoS із SLA (Service level agreement):*

$$SLA_{req} = \{Cost_{req}, T_{req}, R_{req}, Rel_{req}, Av_{req}\};$$

*for  $w=1..N$*

*пошук  $S_w$ , що є складовою комплексного сервісу  $CS$ ;*



```

for w=1..N
    пошук  $S_w$ , що є функційно-еквівалентними;
for w=1..N
    формування множини сервісів  $S_w$ , що задовольняють вимогам до якості
    обслуговування QoS;
i=0;
i=i+1;
if I<N
    Розрахунок  $\max(P_i)$ ;
    If  $G \subset L$ 
    then if  $M_{K, \text{sum}} \geq M_K(\text{SLA}_{\text{req}})$ ;
        j=0;
        j=j+1;
        if j<d
            if  $H \subset L$  then if  $M_{D_j, \text{sum}} \geq M_{D_j}(\text{SLA}_{\text{req}})$ 
                if j<d then go to "j=j+1";
                else Збереження сервісу у множини QoS;
                f=f+1;
            else if  $M_{D_j, \text{sum}} < M_{D_j}(\text{SLA}_{\text{req}})$ 
                if j<d then go to "j=j+1";
                else Збереження сервісу у множини QoS;
                f=f+1;
        else
            else if  $M_{K, \text{sum}} < M_K(\text{SLA}_{\text{req}})$ 
                j=0;
                j=j+1;
                if j<d
                    if  $H \subset L$  then if  $M_{D_j, \text{sum}} \geq M_{D_j}(\text{SLA}_{\text{req}})$ 
                        if j<d then go to "j=j+1";
                        else Збереження сервісу у множини QoS;
                        f=f+1;
                    else if  $M_{D_j, \text{sum}} < M_{D_j}(\text{SLA}_{\text{req}})$ 
                        if j<d then go to "j=j+1";
                        else Збереження сервісу у множини QoS;
                        f=f+1;
                    else go to "i=i+1";
            else if f>0

```

*then if  $f > 1$*

*Формування QoS для множини сервісів  $S$ , таких, що відповідають вимогам  $SLA_{req} = \{Cost_{req}, T_{req}, R_{req}, Rel_{req}, Av_{req}\}$ ;*

*Нормування значень  $R$ ;*

*Прийняття рішення щодо надання комплексного сервісу  $CS$ ;*

*Формування інформації щодо доступу до комплексного сервісу  $CS$ ;*

*else Формування повідомлення про неможливість надання сервісу користувачу*

*Кінець.*

Рис. 3.1. Псевдокод формування комплексного сервісу CS

Множина  $H$  і  $G$  належать множині  $L$ ,  $G \subset L$ ,  $H \subset L$ ,  $Z \cup H = L$ . Множина  $H$  включає в себе показники, які необхідно максимізувати, а множина  $G$  включає в себе показники якості, які необхідно мінімізувати.

Вимоги користувача до якості обслуговування представлені множиною  $SLA_{req} = \{Cost_{req}, T_{req}, R_{req}, Rel_{req}, Av_{req}\}$ .

Пошук всіх CS даного сервісу без урахування показників якості необхідно проводити у випадку, коли за запитом користувача сервіс з необхідними ПК не виявлений в мережі.

Далі інтегральні показники QoS обраного комплексного сервісу порівнюються з вимогами користувача за якістю обслуговування, представленими множиною  $SLA_{req}$ .

У разі, якщо параметр QoS прагне до максимізації своїх значень, то порівняння виконується наступним чином:

$$M_{K_{sum}} \geq M_K(SLA_{req}) \quad (3.20)$$

Або,  $K \subset G$  :

$$M_{K_{sum}} < M_K(SLA_{req}) \quad (3.21)$$

Якщо  $M_{K_{sum}}$  задовольняє вимогам  $SLA_{req}$  по параметру QoS, який належить множині  $K$ , то проводиться перевірка інших параметрів QoS, які

належать множині  $D$ . Дана дія необхідно для того, щоб відібраний комплексний сервіс мав не тільки показники якості по шуканого параметру не нижче необхідних, а й зберігав значення інших параметрів в необхідних межах. У випадку, якщо  $Z_j$  - параметр QoS належить множині  $H$  ( $Z_j \subset H$ ), то порівняння відбувається наступним чином:

$$M_{D_j, \text{sum}} \geq M_{D_j}(\text{SLA}_{\text{req}}), \quad (3.22)$$

або,  $Z_j \subset G$ :

$$M_{D_j, \text{sum}} < M_{D_j}(\text{SLA}_{\text{req}}) \quad (3.23)$$

У разі, якщо інтегральні значення показників якості з множини  $Z$  нижче необхідних, то далі проводиться перевірка наступної підмножини сервісів  $CS_i$ , відібраних з множини  $\text{SLA}_{\text{req}}$ . У іншому випадку, обирається сервіс  $CS_i$  зберігається в безлічі  $CS1$ .

Якщо в множини  $CS$  залишилися неперевірені підмножини, то процес перевірки триває. Інакше, проводиться аналіз сформованої безлічі  $CS1$ .

Якщо множина  $CS1$  є порожньою, то користувачеві відправляється повідомлення про відсутність в мережі сервісу з необхідними характеристиками. Якщо множина  $CS1$  складається з одного сервісу, то користувачеві надається адреса даного сервісу.

У разі, якщо безліч  $CS1$  включає в себе більше одного сервісу  $S$ , то виконується процедура обчислення сервісу з найкращими параметрами.

Робота методу в загальному випадку описується наступною послідовністю етапів.

На першому етапі створюється матриця якості QoSM сервісів, які включені в множину  $CS1$  і вектор  $\text{SLA\_VEC}$  з вимогами користувача за якістю обслуговування:

$$\text{SLA\_VEC} = \text{SLA}_{\text{req}} = \{\text{Cost}_{\text{req}}, T_{\text{req}}, R_{\text{req}}, S_{\text{req}}, Av_{\text{req}}\}, \quad (3.24)$$

де  $f$  – кількість сервісів в множині  $CS1$ ,  $i$  – номер сервісу в множині  $CS1$ ,  $j$  – номер параметру QoS в множині  $L$ ,  $l$  – кількість параметрів QoS умножині  $L$ .

Далі виконується нормування елементів матриці QoSM і вектора SLA\_VEC [89, 93].

На наступному етапі проводиться обчислення евклідової відстані [93] між векторами  $CS1_t$  і  $SLA\_VEC$ :

$$\text{dis}(CS1_t, SLA\_VEC) = \sqrt{\sum_{a=1}^l (CS1_{t_a} - SLA\_VEC_a)^2} \quad (3.25)$$

У разі, якщо для користувача одні параметри є більш важливими ніж інші, то при порівнянні сервісів дана пріоритетність враховується за допомогою вагового коефіцієнта, представленого вектором  $K$ ,  $K = \{k_1, k_2, \dots, k_a, \dots, k_l\}$ :

$$\text{dis}(CS1_t, SLA\_VEC) = \sqrt{\sum_{i=1}^l (CS1_{t_i} - SLA\_VEC_i)^2 \cdot K_i} \quad (3.26)$$

На наступному етапі визначається сервіс  $CS1_t$  з мінімальною евклідовою відстанню. Далі адреса відібраного сервісу надаються користувачеві.

Запропоновані методи дозволяють підвищити надійність і продуктивність запитуваних сервісів. Застосування методу формування розподіленого сервісу та методу підвищення надійності сервісу дозволить доповнити і вдосконалити механізм формування сервісів з необхідними показниками якості, збільшити кількість більш якісних сервісів в мережі, знизити завантаженість сервісів з високими ПК за рахунок залучення сервісів з більш низькими значеннями параметрів QoS.

### 3.3. Метод динамічного розподілу пропускної здатності каналів зв'язку в мережах з віртуалізацією мережевих функцій

Завдання ефективного розподілу пропускної здатності каналів зв'язку можна відобразити у формулі [116, 122]:

$$Q(\Lambda^{IN}, R_{in\alpha}, R_{out\alpha}) \rightarrow \max_{CP_{NL}}, \quad (3.27)$$

де  $Q()$  – цільова функція чи функція оптимізації,  $\Lambda^{IN}$  – сумарний потік вхідного трафіку,  $R_{in\alpha}$  и  $R_{out\alpha}$  – пропускні спроможності каналів зв'язку зовнішньої мережі і внутрішньої мережі,  $CP_{NL}$  – політика управління мережевим навантаженням (вибір каналу зв'язку). В даному випадку завдання оптимізації можна звести до наступного вигляду [116]:

$$\Lambda^{IN} \rightarrow \max_{CP_{NL}} | R_{in\alpha}, R_{out\alpha}. \quad (3.28)$$

Оптимального розподілу пропускної здатності можна досягнути, відповідно до запропонованого методу, шляхом розподілу потоків трафіку всіма допустимими каналами зв'язку пропорційно доступним пропускним здібностям цих каналів в поточний момент часу або у момент часу надходження заявки.

Передбачається, що для кожного потоку доступні  $n$  каналів зв'язку. Вибір каналу зв'язку ґрунтується на значенні його пропускної здатності. Вибір політики розподілу пропускних здатностей каналів зв'язку залежить від початкових QoS вимог до потоку даних (пріоритетність запиту).

У загальному випадку загальний потік можна розподілити по каналах зв'язку, з огляду на таку залежність:

$$\frac{\lambda_1}{\mu_{ch1}} = \frac{\lambda_2}{\mu_{ch2}} = \dots = \frac{\lambda_n}{\mu_{chn}} = \rho \mid \sum_{i=1}^{N_{доет}} \lambda_i = \Lambda^{IN}, \sum_{i=1}^{N_{доет}} \mu_{chi} = C, \quad (3.29)$$

де  $\rho$  – завантаженість каналів зв'язку,  $N_{\text{дост}}$  – кількість доступних каналів зв'язку,  $\lambda_i$  – інтенсивність трафіку для  $i$ - того каналу зв'язку,  $\mu_{chi}$  – пропускна здатність  $i$ - того каналу зв'язку.

За умови, що всі завантаження каналів зв'язку рівні  $\rho$  між собою, вираз (3.29) може бути зведено до наступного виразу:

$$\rho = \frac{\sum_{i=1}^{N_{\text{дост}}} \lambda_i}{\sum_{i=1}^{N_{\text{дост}}} \mu_{chi}} = \frac{\Lambda^{IN}}{\sum_{i=1}^n \mu_{chi}}, \quad (3.30)$$

де  $\frac{\Lambda^{IN}}{\sum_{i=1}^n \mu_{chi}} \rightarrow \min$ .

Так як верхня сума  $\sum_{i=1}^{N_{\text{дост}}} \lambda_i$  завжди зводиться до  $\Lambda^{IN}$ , незалежно від кількості каналів зв'язку, для вирішення поставленого завдання необхідно максимізувати нижню суму  $\sum_{i=1}^{N_{\text{дост}}} \mu_{chi}$ .

Досягнення максимального значення  $\mu_{chi}$  при розподілу потоків вхідного трафіку всіма допустимими каналами зв'язку, допустимо лише в разі пропорційного розподілу потоку трафіку щодо пропускних здатностей. При цьому враховується, що сумарна пропускна здатність декількох каналів зв'язку також може бути пропорційна частці  $i$ -го вхідного потоку:

$$\mu_{ch=x} = \sum_{j=x}^y \mu_{chj} \frac{\lambda_i}{\Lambda^{IN}}, 1 < x < y, x < y < N_{\text{дост}}. \quad (3.31)$$

Пропонований метод розподілу пропускної здатності полягає в первинному виборі шляху максимальної пропускної здатності для пріоритетного потоку і шляхів, що доповнюють, для розподілу непріоритетних потоків. В принципі функціонування пропонованого алгоритму покладено алгоритми Едмондс-Карпа [49] і максимізації пропускної здатності [59] з деякими припущеннями.

Метод складається з наступних етапів:

1. Побудова матриці зв'язності або визначення топології накладеної мережевої інфраструктури контролером. Матриця зв'язності включає в себе всі можливі взаємозв'язки між множиною  $N_\lambda$  - вузлів-відправників і множиною  $VND$  - обслуговуючих пристроїв:

$$NC = \begin{bmatrix} 0,1,0,\dots,1 \\ 1,0,1,\dots,1 \\ \dots\dots\dots \\ 1,1,0,\dots,1 \end{bmatrix}.$$

Рядки матриці відповідають обчислювальним пристроям мережі  $1 \leq N_{\lambda i} \leq N$ , стовпці матриці - обслуговуючим пристроям  $1 \leq VND_{\lambda i} \leq VND$ , де  $N$  - кількість обчислювальних пристроїв,  $VND$  - кількість віртуалізованих обслуговуючих пристроїв, які можуть бути задіяні в процесі надання сервісів. Наявність взаємозв'язку між обчислювальним і обслуговуючим пристроєм позначається символом «1», відсутність взаємозв'язку - символом «0».

Обмеження безлічі можливих каналів зв'язку між вузлом-відправником і обслуговуючим пристроєм. В подальшій роботі алгоритму розглядаються лише ті канали зв'язку, значення яких в матриці  $NC = 1$ .

2. Визначається безліч прямих каналів зв'язку – зв'язністю «точка-точка» і обхідних каналів. Найкоротші обхідні канали зв'язку відносяться до резервних каналів.

3. Побудова правил розподілу пропускної здатності при наданні сервісів з різним рівнем пріоритету.

Значення максимальної і дійсної пропускної здатності каналів зв'язку можна визначити на основі побудованої матриці зв'язності.

На керуючому пристрої формується два типи матриць розмірністю  $N \times N$ .

$$R_1 = \begin{bmatrix} 0, R_{12}, R_{13}, \dots, R_{1N} \\ R_{21}, 0, R_{23}, \dots, R_{2N} \\ \dots\dots\dots \\ R_{N1}, R_{N2}, R_{N3}, \dots, 0 \end{bmatrix} \quad R_2 = \begin{bmatrix} 0, r_{12}, r_{13}, \dots, r_{1N} \\ r_{21}, 0, r_{23}, \dots, r_{2N} \\ \dots\dots\dots \\ r_{N1}, r_{N2}, r_{N3}, \dots, 0 \end{bmatrix},$$

де  $N$  – кількість вузлів в мережі, задіяних в процесі надання сервісу,  $R_1$  - матриця поточних пропускних здатностей  $R_{ij}$  каналів зв'язку міжвузлами,  $R_2$  - матриця встановлених порогових значень пропускних здатностей  $I_{ij}^*$  каналів зв'язку.

4. У загальному випадку наданий сервіс маркується двома типами маркерів з високим рівнем пріоритету, і сервіс з низьким рівнем пріоритету. Поділ здійснюється на підставі вимог, запропонованих в [41]. Таким чином, можуть бути складені додаткові динамічні матриці, що включають поточний розподіл пріоритетного і непріоритетного трафіку (запитів) в мережі:

$$M^p_{\lambda_{inp}} = \begin{bmatrix} 0, \lambda^p_{inp12}, \lambda^p_{inp13}, \dots, \lambda^p_{inp1N} \\ \lambda^p_{inp21}, 0, \lambda^p_{inp23}, \dots, \lambda^p_{inp2N} \\ \dots \\ \lambda^p_{inpN1}, \lambda^p_{inpN2}, \lambda^p_{inpN3}, \dots, 0 \end{bmatrix} \quad M^{np}_{\lambda_{inp}} = \begin{bmatrix} 0, \lambda^{np}_{inp12}, \lambda^{np}_{inp13}, \dots, \lambda^{np}_{inp1N} \\ \lambda^{np}_{inp21}, 0, \lambda^{np}_{inp23}, \dots, \lambda^{np}_{inp2N} \\ \dots \\ \lambda^{np}_{inpN1}, \lambda^{np}_{inpN2}, \lambda^{np}_{inpN3}, \dots, 0 \end{bmatrix}.$$

Загальний потік вхідного трафіку визначається як  $\lambda_{inij} = \lambda^p_{inij} + \lambda^{np}_{inij}$ , де  $\lambda^p_{inij}$  - запити з високим рівнем пріоритету,  $\lambda^{np}_{inij}$  - запити з низьким рівнем пріоритету.

Надалі ключове значення матиме агрегований або комбінований потік даних, який передається по каналу зв'язку  $\lambda_{agrij} = \sum \lambda^{p/np}_{inij}$ .

До розгляду приймається або канал зв'язку з максимальним значенням поточної пропускної здатності (або прямий канал зв'язку) для потоку від вузла  $i$  до вузла  $j$ .

Розподіл пропускної здатності каналів зв'язку здійснюється за наступним принципом:

- визначається поточне завантаження даного каналу зв'язку:

$$\begin{cases} Ch_{R_{ij}} = Th_{ij}, \text{overloaded} \\ Ch_{R_{ij}} < Th_{ij}, \text{loaded} \\ 0 < Ch_{R_{ij}} < Th_{ij}, \text{underloaded} \end{cases}, \quad (3.32)$$

де  $Th_{R_{ij}}$  - поточне порогове значення



Нехай в деякий момент часу  $t_s$  в кожний  $i$ -й вузол надходять потоки вхідного трафіку, які можна розділити на два класи обслуговування: потік з високим рівнем пріоритету  $\lambda_{inp}^p(t_s) = \{\lambda_{inp1}^p(t_s), \dots, \lambda_{inpj}^p(t_s)\}$  і потік з низьким рівнем пріоритету  $\lambda_{inp}^{np}(t_s) = \{\lambda_{inp1}^{np}(t_s), \dots, \lambda_{inpj}^{np}(t_s)\}$ .

Контролер визначає ненавантажений канал зв'язку і проводить перевірку можливості передачі трафіку з високим пріоритетом з даного каналу зв'язку:  $Ch_{R_{ij}}(t_s) \geq \lambda_{inpij}^p(t_s)$ .

- якщо твердження вірне, то потік даних передається даним каналом зв'язку і визначається залишкова пропускна здатність  $Ch_{c_{ij}}(t_s) - \lambda_{inpij}^p(t_s) = \Delta Ch_{c_{ij}}(t_s)$ . Проводиться повторна перевірка завантаженості для новоприбулого запиту або сформованого потоку даних  $\lambda_{inpij}^p(t_{s+1})$ , якщо твердження вірне, то потік даних передається даним каналом зв'язку.

- якщо твердження не вірне, то вибирається інший канал, здатний передати трафік типу  $\lambda_{inpij}^p(t_s)$  і для даного каналу визначається як:

$$Ch_{R_{ij}}(t_s) - \lambda_{inpij}^p(t_s) = |\Delta Ch_{R_{ij}}(t_s)| \quad (3.33)$$

Якщо  $|\Delta Ch_{R_{ij}}(t_s)| \geq \lambda_{inpij}^{np}(t_s)$ , то даним каналом зв'язку передається частка трафіку з низьким пріоритетом в співвідношенні:  $LP = \frac{\lambda_{inpij}^{np}(t_s)}{\eta}$ ,  
(3.43)

де  $\eta$  – коефіцієнт завантаженості, вводиться для обмеження повного завантаження каналу зв'язку і можливості його використання як резервного.

Оцінка залишкового каналного ресурсу.

На основі зібраних статистичних даних розраховується залишкова пропускна здатність для даного каналного ресурсу:

$$Th_{R_{ij}} = \begin{cases} R_{ij} - \sum_{i,j=1}^n \lambda_{ij}^p(t_s) & | \lambda_{ij}^p(t_s) \leq R_{ij} \\ 0, \lambda_{ij}^p(t_s) = R_{ij} \end{cases}, \quad (3.34)$$

$Th_{R_{ij}}$  виступає граничним значенням для даного каналу зв'язку при подальшому розподілі навантаження.

У відповідності з формулою (  $AL = \frac{\lambda^{np}_{inpij}(t_s)}{k}$  ) обчислюється максимальне значення завантаження каналу:

$$C_{ij\max}(t_s) = \sum_{i=1}^n \lambda^{p}_{inpij}(t_s) + \frac{\lambda^{np}_{inpij}(t_s)}{k} .$$

(3.45)

Таким чином, застосування даного алгоритму не призводить до повної завантаженості каналу зв'язку. Залишаючи мінімальне значення в якості резервного. Структурна схема розробленого методу приведена на рис. 3.2.

*Початок;*

*Введення сформованого комплексного сервісу CS;*

*Місцезнаходження сервісів  $S_i$  встановлено;*

*Визначення пріоритету сервісів  $S_i$ ;*

*$i = 1$ ;*

*$i = i + 1$ ;*

*if  $i < N$*

*then if Сервіс  $S_i$  визначено як пріоритетний*

*then Визначення поточної пропускної здатності  $Th_{R_{ij}}$  ;*

*if  $0 < Ch_{R_{ij}} < Th_{R_{ij}}$*

*Канал  $Ch_{ij}$  обирається для передачі даних з високим пріоритетом;*

*Розрахунок залишкової пропускної здатності;*

*if  $0 < |\Delta Th_{R_{ij}}| < Th_{R_{ij}}$*

*then Передача трафіку з низьким рівнем пріоритету;*

*Розрахунок залишкової пропускної здатності  $AL$ ;*

*If  $AL < Th_{R_{ij}}(SLA)$  then Передача інформації з низьким рівнем пріоритету;*

*else Визначення поточної пропускної здатності  $Th_{R_{ij}}$  ;*

*go to "i=i+1";*

*else Формування рішення щодо завантаженості каналу та пошук іншого;*

*else if  $Th(SLA) < Th_{R_{ij}}$  then Передача трафіку з низьким рівнем пріоритету;*

*Розрахунок залишкової пропускної здатності  $AL$ ;*

*If  $AL < Th_{r_{ij}}(SLA)$  then Передача інформації з*

*низьким рівнем пріоритету;*

*else Визначення поточної пропускної здатності  $Th_{R_{ij}}$  ;*

*go to “ $i=i+1$ ”;*

*else Канал перевантажено;*

*Формування рішення щодо завантаженості каналу та*  
*пошук іншого;*

*else Визначення поточного стану пропускної здатності каналу;*

*if  $0 < Ch_{R_{ij}} < Th_{r_{ij}}$  then Канал перевантажено;*

*Формування рішення щодо завантаженості каналу та*  
*пошук іншого;*

*else Передача трафіку з низьким рівнем пріоритету;*

*Розрахунок залишкової пропускної здатності  $AL$ ;*

*If  $AL < Th_{r_{ij}}(SLA)$  then Передача інформації з низьким рівнем*

*пріоритету;*

*else Канал перевантажено;*

*go to “ $i=i+1$ ”;*

*Формування рішення щодо завантаженості каналу та*  
*пошук іншого;*

Рис. 3.2. Псевдокод реалізації динамічного розподілу пропускної здатності каналів зв'язку в мережах з підтримкою функцій віртуалізації

### 3.4. Оцінка ефективності удосконаленого методу динамічного розподілу пропускної здатності каналів зв'язку

Для оцінки ефективності запропонованого алгоритму балансування навантаження для мереж з підтримкою технології NFV використовувався інструмент моделювання мережі під назвою mininet [36, 37, 47]. Фрагмент експериментальної мережі наведено на рис. 3.3.

Керовані комутатори, що використовуються в процесі проведення експерименту, мають наступні характеристики: Sw1 - 45 Мбит/с, Sw2 - 45 Мбит/с, Sw3 - 45 Мбит / с, Sw4 - 60 Мбит/с, Sw5 - 60 Мбит/с, Sw6 - 60 Мбит/с.

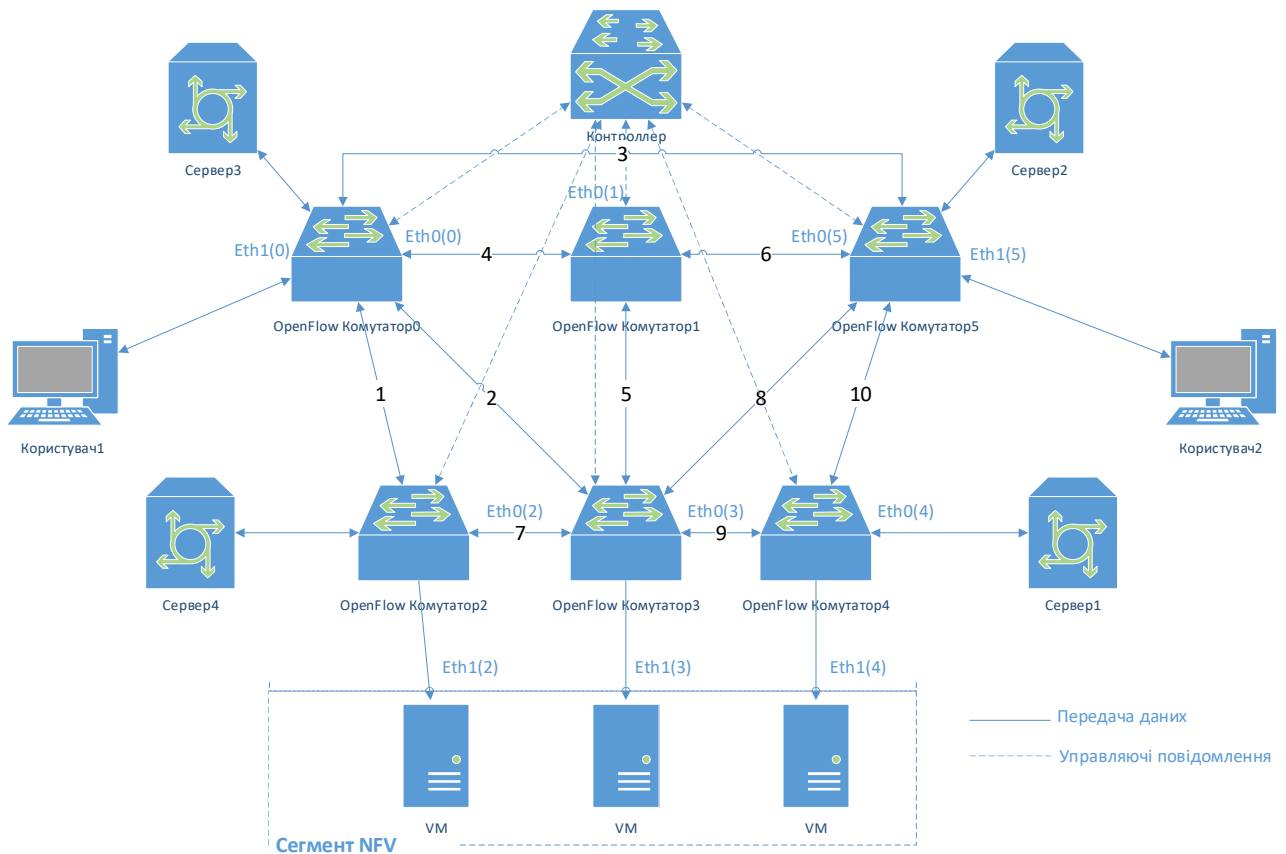


Рис. 3.3. Фрагмент експериментальної мережі

Фрагмент конфігурації мережі в середовищі mininet наведений на рис. 3.4:

```

*** Shutting down stale SimpleHTTPServers
*** Shutting down stale webservers
server1 192.168.2.2
server2 172.64.3.10
client 10.0.1.100
sw0-eth1 192.168.2.1
sw0-eth2 172.64.3.1
sw0-eth3 10.0.1.1
*** Successfully loaded ip settings for hosts
{'server1': '192.168.2.2', 'sw0-eth3': '10.0.1.1', 'sw0-eth1': '192.168.2.1', 'sw0-eth2': '172.64.3.1', 'client': '10.0.1.100', 'server2': '172.64.3.10'}
*** Creating network
*** Creating network
*** Adding controller
*** Adding hosts:
client server1 server2
*** Adding switches:
sw0
*** Adding links:
(client, sw0) (server1, sw0) (server2, sw0)
*** Configuring hosts
client server1 server2
*** Starting controller
*** Starting 1 switches
sw0
*** setting default gateway of host server1
server1 192.168.2.1
*** setting default gateway of host server2

```

```

server2 172.64.3.1
*** setting default gateway of host client
client 10.0.1.1
*** Starting SimpleHTTPServer on host server1
*** Starting SimpleHTTPServer on host server2
*** Starting CLI:
mininet>

```

Рис. 3.4. Конфігурація експериментального фрагменту мережі

В процесі надання сервісів між обчислювальними пристроями Host1 і Host3 розглянута організація наступних маршрутів передачі даних:

Path 1: Sw1 – Sw2 – Sw5;

Path 2: Sw1 – Sw3 – Sw5;

Path 3: Sw1 – Sw4 – Sw6 – Sw5.

В процесі експерименту були розглянуті наступні алгоритми розподілу пропускної здатності: Round Robin [1], Central Queuing [40] і пропонований метод.

Передача даних здійснювалася через мережеві елементи Sw1, Sw2, Sw3, Sw4, Sw5 і Sw6, що дозволило згенерувати навантаження на всі канали передачі даних. Після того, як навантаження було розподілено між каналами передачі, вузол Host1 почав передавати певний набір даних вузлу Host3 з метою імітації нового вхідного потоку даних в експериментальному фрагменті NFV. Результати експерименту показаного розподілу пропускної здатності при застосуванні різних методів балансування навантаження наведені на рис. 3.5, рис. 3.6 і рис. 3.7 відповідно.

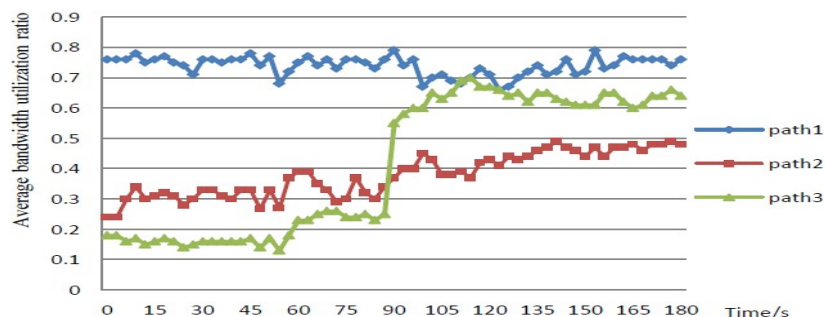


Рис. 3.5. Усереднене значення використання пропускної здатності каналів зв'язку при застосуванні запропонованого методу розподілу пропускної здатності

При необхідності передачі даних на новий вхідний потік, запропонована стратегія балансу навантаження вибирає як шлях передачі Path 3, який в хорошому стані навантаження (рис. 3.6). І всі альтернативні шляхи (Path 1 і Path 2) підтримуються в хорошому співвідношенні коефіцієнта використання смуги пропускання.

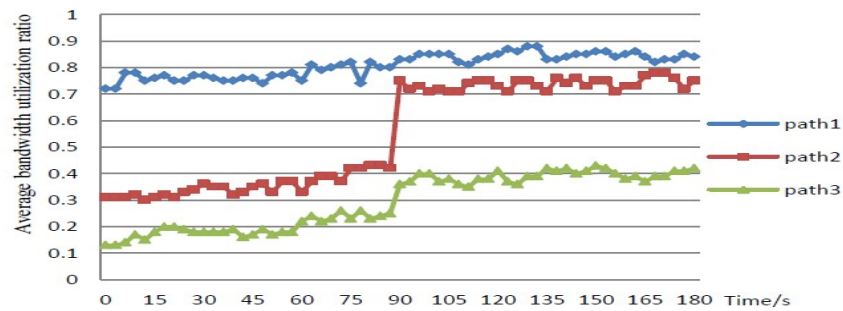


Рис. 3.6. Усереднене значення використання пропускну здатності при застосуванні методу Central Queuing

Проаналізувавши графіки (рисунок 3.7), можна зробити висновок про те, що запропонований в [18] алгоритм найменшого з'єднання при передачі нового вхідного потоку даних може розглядати тільки умову зв'язку наступного переходу. Таким чином, цей метод помилково вибирає Path 2 як шлях передачі. Через низьку пропускну здатність Path 2 коефіцієнт використання смуги пропускання швидко збільшується. Тим часом, оскільки існує загальне посилення в Path 2 і Path 3, ця стратегія балансу навантаження також впливає на Path 3 і збільшує ймовірність перевантаження.

На рис. 3.7 наведені експериментальні результати, отримані при функціонуванні алгоритму Round Robin. Основний недолік Round Robin полягає в тому, що він не може брати до уваги стан каналів зв'язку в реальному режимі часу. Застосування Round Robin призводить до гіршої умови навантаження в Host 1 і збільшує ймовірність перевантаження мережі в цілому.

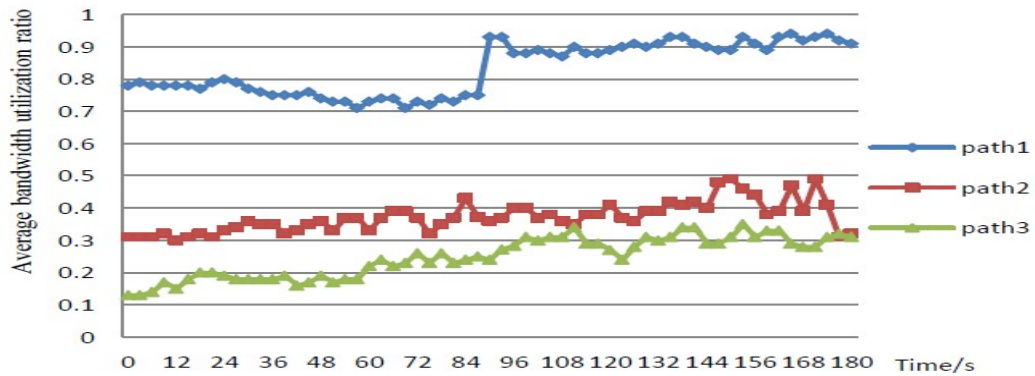


Рис. 3.7. Усереднене значення використання пропускної здатності при застосуванні методу Round Robin

На рис. 3.8 наведено графік зміни часу відгуку при передачі даних між обчислювальними елементами Host1 і Host3 для трьох аналізованих методів.

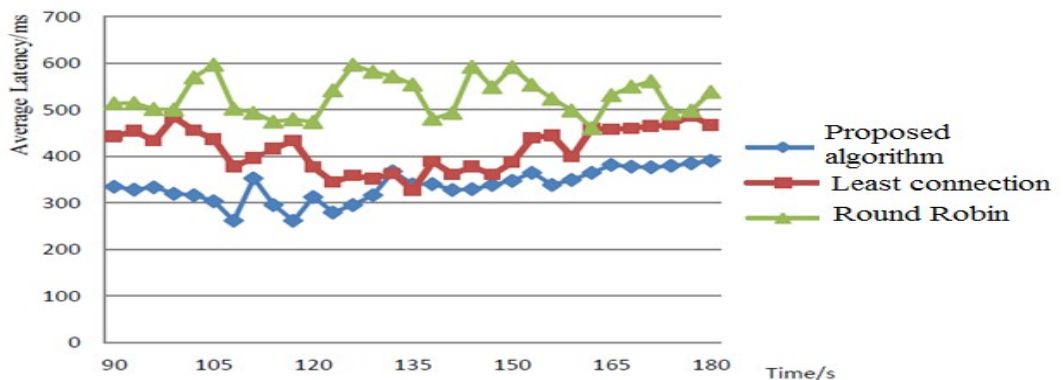


Рис. 3.8. Значення часу відгуку при передачі даних між обчислювальними елементами Host1 to Host3

Узагальнені результати експерименту наведені в табл. 3.1, 3.2.

Таблиця 3.1

Порівняння завантаження каналів передачі при використанні різних алгоритмів балансування (від OpenFlow комутатора 0 до серверу 1),%

	Round Rabin	Central Queuing	Розроблений метод
Path 1	0,31	0,380	0,301
Path 2	0,410	0,744	0,383
Path 3	0,899	0,842	0,668

Таблиця 3.2

Порівняння середньої затримки при використанні різних алгоритмів балансування (від OpenFlow комутатора 0 до Сервер 1), мс

	<b>Round Robin</b>	<b>Central Queuing</b>	<b>Розроблений метод</b>
<b>Path 1</b>	0,301	0,392	0,321
<b>Path 2</b>	0,319	0,440	0,330
<b>Path 3</b>	0,420	0,541	0,469

Беручи до уваги отримані результати, можна стверджувати, що розроблений метод динамічного балансування навантаження дозволяє відстежувати поточний стан мережі і своєчасно реагувати на зміни параметрів мережі в режимі реального часу.

Застосування запропонованого методу дозволяє уникнути перевантаження каналів передачі за рахунок того, що залишкова вартість смуги пропускання каналу використовується для передачі даних з низьким пріоритетом. Отримані результати вказують на те, що запропонований алгоритм балансування навантаження дозволяє підвищити ефективність використання каналів передачі на 7% і зменшити середню затримку в середньому на 16% в порівнянні з динамічним алгоритмом Central Queuing.

### **3.5. Висновки до третього розділу**

1. Проведено аналіз оцінки показників якості комплексних сервісів, заснований на інформації про показники якості атомарних сервісів. Розроблені математичні формалізми, що дозволяють оцінити вплив атомарних сервісів на сумарний показник якості результуючого комплексного сервісу в разі різних сценаріїв включення і використання атомарних.

2. Удосконалено метод формування комплексних сервісів, що дозволяє динамічно формувати склад комплексного сервісу шляхом своєчасної заміни або додавання атомарних сервісів. Розроблений метод дозволяє



використовувати кілька ідентичних атомарних сервісів для досягнення необхідного рівня якості обслуговування. Застосування даного методу дозволяє забезпечити необхідні показники продуктивності запитуваного сервісу.

3. Удосконалено метод динамічного розподілу пропускної здатності каналів передачі в NFV мережах, заснований на принципах розподілу залишкової пропускної здатності. Застосування даного методу дозволяє уникнути повного завантаження і перевантаження каналу передачі.

4. З метою аналізу ефективності запропонованої методики динамічного розподілу пропускної здатності каналів зв'язку проведено експериментальне дослідження, за результатами якого визначено, що запропонований метод розподілу пропускної здатності дозволяє підвищити ефективність використання каналів передачі на 7% і зменшити середню затримку в середньому на 16% в порівнянні з динамічним алгоритмом Central Queuing, а застосування розробленого методу реплікації сервісів дозволяє підвищити доступність сервісу в 14 разів в разі перенавантажень.

## РОЗДІЛ 4

### РОЗРОБКА МЕТОДУ РЕПЛІКАЦІЇ СЕРВІСІВ В МУЛЬТИСЕРВІСНИХ МЕРЕЖАХ З ПІДТРИМКОЮ ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ

Поряд з ефективністю функціонування системи управління на показники доступності послуг в NFV мережах також впливають такі фактори [57, 68, 72]:

- відмова системи зберігання. Відмова системи зберігання має на увазі вихід з ладу модулів зберігання інформації або їх слабку масштабованість, яка не дозволяє обробляти існуючу кількість запитів;

- недостатність мережевих ресурсів. Недостатність мережевих ресурсів має на увазі падіння продуктивності обчислювальних ресурсів і завантаженість каналів зв'язку, які впливають на виникнення затримок, а також можуть бути причиною відмови в обслуговуванні;

- відмова обчислювальних вузлів – фактор, при якому погіршення якості надаваних послуг, або ж відмова в обслуговуванні спричинені механічною помилкою чи програмним збоєм.

Непрогнозованість вище вказаних факторів може призводити до ряду несподіваних проблем. Для швидкої ліквідації проблем і підтримки доступності послуг в режимі 24/7 застосовується стратегія реплікації.

#### 4.1. Аналіз методів реплікації сервісів в мультисервісних мережах

Більшість великих провайдерів послуг, таких як Google, Amazon, IBM, RedHeat, Microsoft і ін. [74, 78, 91] на сьогоднішній день широко використовують механізми як кластерної, так і глобальної реплікації, як для реплікації даних, так і реплікації шаблонних послуг з метою підвищення їх надійності, відмовостійкості і продуктивності. Реплікація добре зарекомендувала себе як техніка, що дозволяє підвищити продуктивність і доступність послуг, а також забезпечити гнучку масштабованість послуг на безлічі розподілених обчислювальних пристроїв і географічно віддалених зон.

Механізм реплікації включає в себе обмін інформацією між серверами, що надають послуги, з подальшим копіюванням даних або послуг, розподілом і узгодженням їх на розподілених обчислювальних вузлах.

Структурна схема розподіленої мультисервісної мережі з підтримкою реплікації приведена на рис. 4.1.

Нині велика кількість робіт направлена на пошук оптимальних механізмів реплікації даних в хмарній мережевій інфраструктурі [36, 37]. В цілому, в залежності від топології і масштабів мережі, а також географічної розподіленої серверів найбільш широке застосування знайшли такі стратегії реплікації: статичні і динамічні реплікації даних, централізовані і децентралізовані, синхронні і асинхронні.

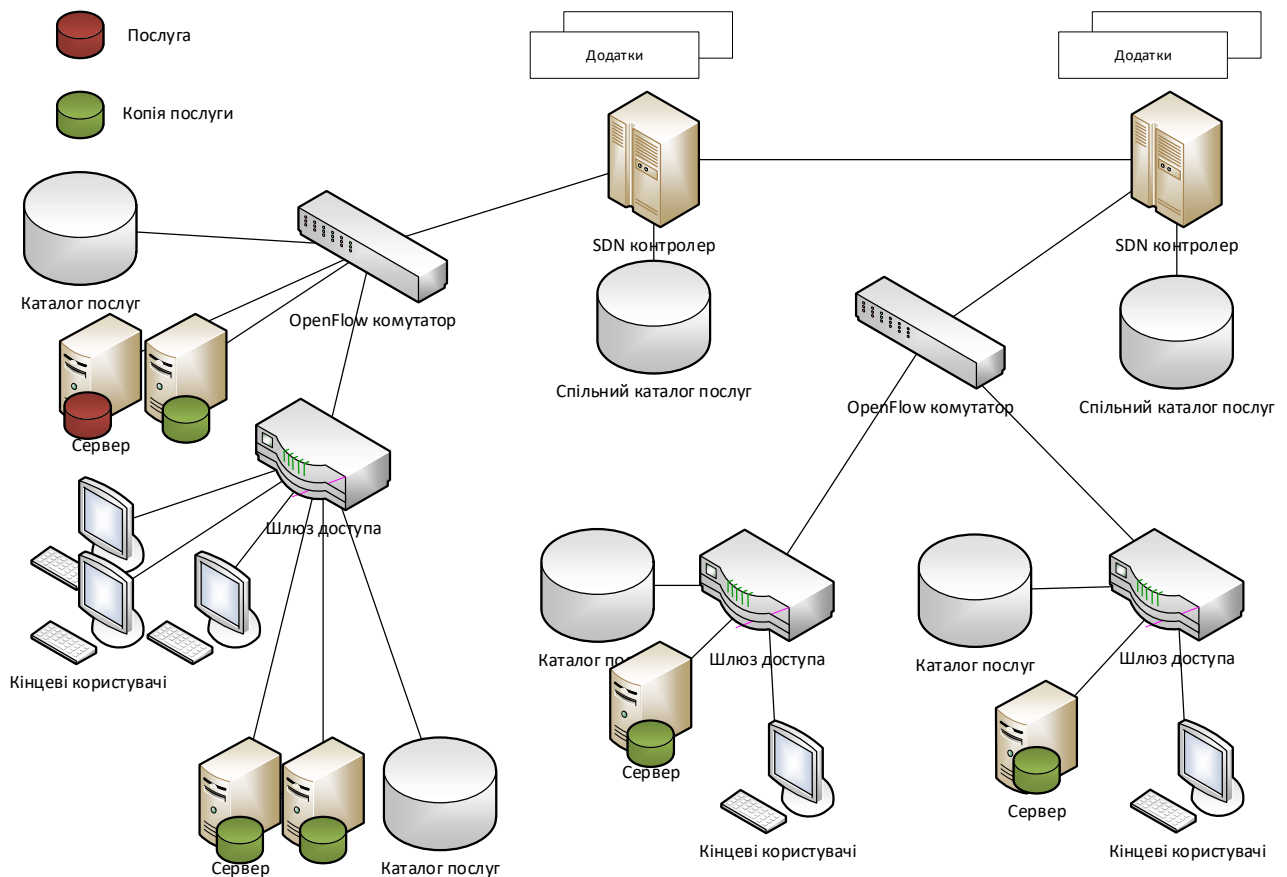


Рис. 4.1. Розташування реєстрів сервісів у фрагментіNFVмережі

### **Метод статичної реплікації даних**

При використанні статичної реплікації правила формування і розподілу реплік послуги визначені заздалегідь і не змінюються протягом усього життєвого циклу. Одним із прикладів статичної реплікації є алгоритм реплікації на основі заздалегідь отриманої інформації про кількість необхідних реплік і правила їх розподілу без суворого обліку поточного стану системи (кількості використовуваних ресурсів сервера, кількість запитів і локація запитів).

Нова копія в разі застосування статичної реплікації створюється тоді, коли кількість реплік конкретної послуги або блоку даних, необхідних для формування комплексної послуги, стає нижче межі, яка задана при ініціалізації процесу реплікації. Таким чином, виконується реплікація з фіксованою кількістю копій послуг, які рівномірно розподілені по всій мережі. Даний вид реплікації підтримується в рішеннях, запропонованих Windows Azure, Amazon S3, GFS [28, 29].

Застосування статичного методу реплікації в програмно-конфігурованих мережах спричинює зниження ефективності використання мережевих ресурсів, так як кожен користувач має різні вимоги до продуктивності, тривалості затребуваності послуги та її надійності, а, отже, має різні потреби щодо процедури реплікації.

Однією з модифікацій статичної реплікації, яка дозволяє адаптуватися до змін вимог, є OPERA (OPEn Replication Architecture) [17]. Сервери з підтримкою реплікації OPERA формують рішення про необхідність реплікації послуги на підставі декількох критеріїв: ступеня доступності послуги, ступеня затребуваності послуги та локалізацію послуги. Однак стратегія реплікації OPERA також, як і інші методи статичної реплікації, не дозволяє робити гнучкий вибір в разі неоднорідної природи послуг (мінливої популярності у користувачів).

### **Метод динамічної реплікації даних**

Стратегія динамічної реплікації або адаптивна стратегія реплікації дозволяє створювати репліки послуг, розподіляти і управляти ними на основі

гнучких поведінкових шаблонів, що відповідають кожній копії [ ]. В цьому випадку рішення про формування і розміщення реплік засновані на зміні робочого навантаження серверів, їх доступності та продуктивності, кількості і локалізації запитів користувачів. У разі динамічної замість постійної підтримки копій на віддаленому обчислювальному вузлі, реплікація послуги виконується тільки тоді і в тому фрагменті мережі, де цього вимагають користувачі.

Підмножини даних, що належать популярній послугі, можуть бути збережені як на найближчий сервер, так і на локальний вузол в якості репліки. Таким чином, при використанні динамічної стратегії, репліки створюються і видаляються відповідно до зміни структури доступу [108, 109].

У мережах, де мережеві ресурси є вузьким місцем (низька пропускна здатність каналів зв'язку), для зберігання реплік вибираються локальні фрагменти мережі, де дані репліки мають найбільшу популярність. Найчастіше в процесі динамічної реплікації формується рішення про необхідність створення репліки послуги і вибір її місця розташування, який ґрунтується на розрахунку відношення реальної кількості реплік, що зберігаються в даному фрагменті мережі, і кількості реплік, запитуваних у даний момент, а також аналізі мережевих характеристик сервера, на якому дані репліки повинні бути розміщені.

Одна з реалізацій стратегії динамічної реплікації запропонована в механізмі FIFO [27]. В цьому випадку, планувальник Fair, з метою формування рішення щодо вибору місця розташування реплік послуги, які дозволять скоротити витрати мережевих ресурсів, використовує ряд певних критеріїв. В даному випадку копія популярного блоку даних або послуги зберігається на локальному вузлі або сервері.

#### Централізована і децентралізована стратегії реплікації даних

У разі застосування централізованої (горизонтальної) стратегії реплікації вибирається центральний (master) сервер, який є провідним, на кожному віддаленому фрагменті мережі розміщується підлеглий (slave) сервер [61]. До підлеглого сервера надходять запити кінцевих користувачів, на підставі

результатів аналізу запитів користувача формується висновок про вибір репліки і її розміщенні. Центральний сервер обробляє і надсилає дані підлеглим серверам, таким чином, кожен з них містить актуальні дані. Всі дані синхронізовані. Структурна схема централізованої реплікації сервісів приведена на рис. 4.2.

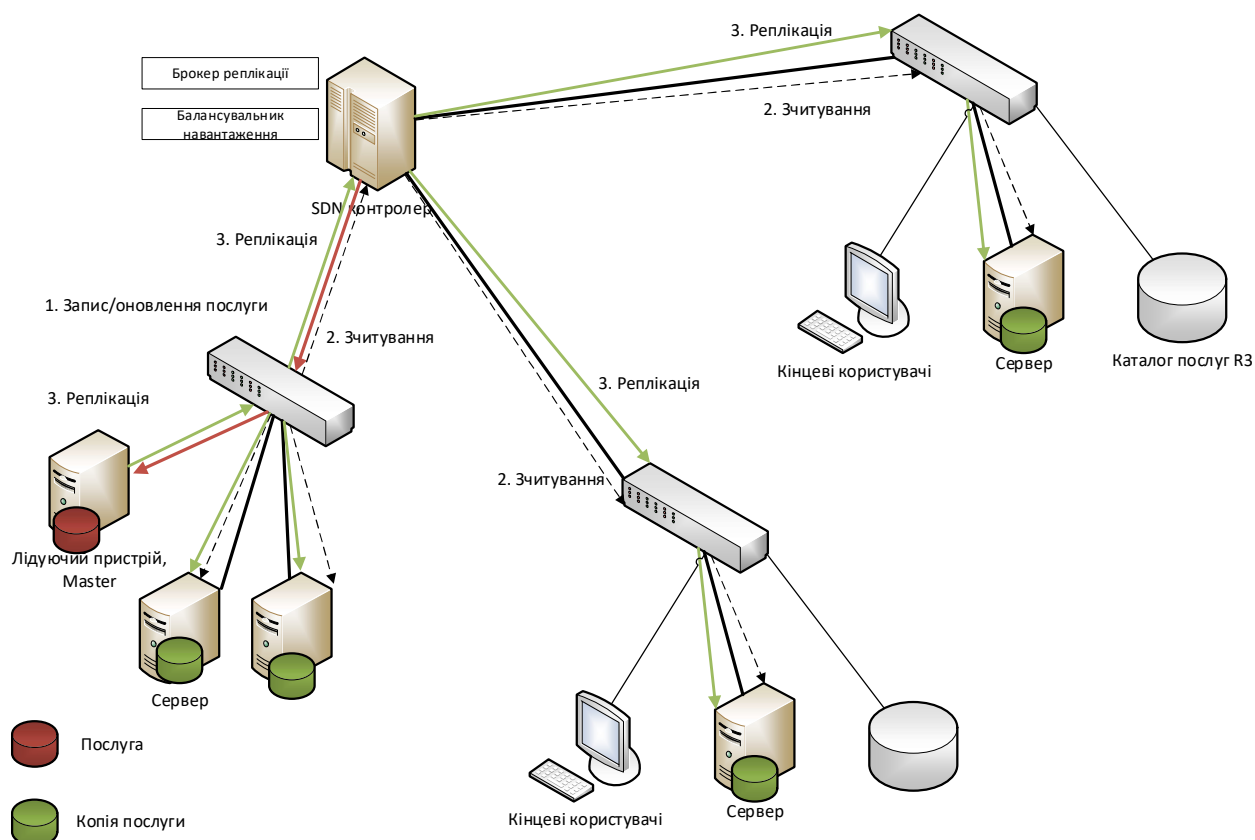


Рис. 4.2. Структурна схема мережі з використанням централізованої реплікації сервісів

У разі застосування децентралізованої (вертикальної) стратегії вибирається кілька географічно розподілених ведених серверів, репліки яких синхронізовані між собою і безліч підлеглих серверів, підключених до кожного ведучого [61, 65]. Децентралізована стратегія реплікації полягає в формуванні реплікаційних наборів і подальшому обміні ними між двома видами серверів: провідними серверами і підлеглими серверами. При цьому існує кілька рангів підлеглих серверів. Підтримка синхронізації не є обов'язковою. На рис. 4.3 наведена структурна схема децентралізованої стратегії реплікації.

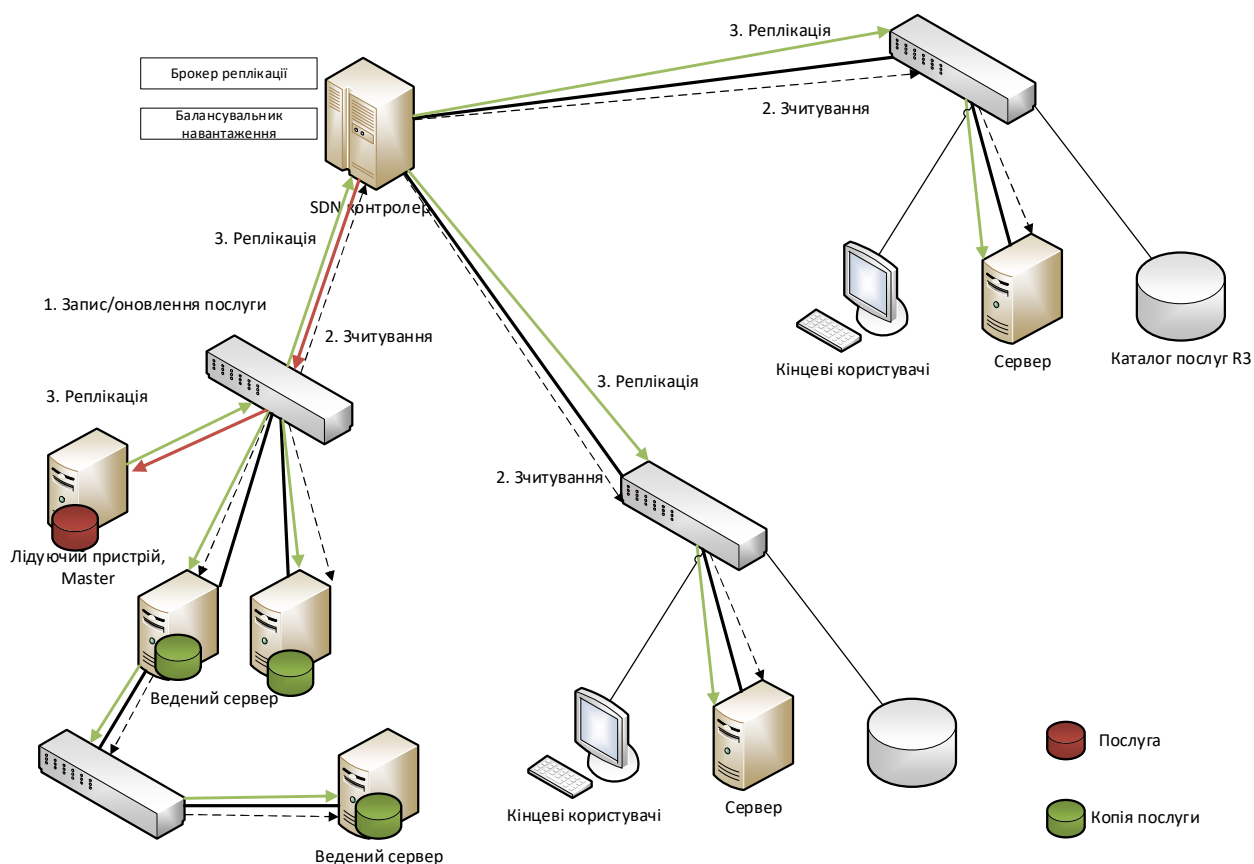


Рис. 4.3. Структурна схема мережі з використанням децентралізованої реплікації сервісів

Централізована реплікація передбачає однорідну структуру вузлів мережі і підтримує узгоджений огляд даних усіма користувачами. Відповідно до принципів функціонування мереж з підтримкою віртуалізації мережевих функцій централізований механізм реплікації ефективний при відносно невеликому числі учасників обміну даними. Синхронізація реплік при використанні централізованого механізму реплікації часто призводить до підвищення накладних витрат. Застосування децентралізованої стратегії реплікації дозволяє забезпечити ефективний розподіл мережевих ресурсів.

### Стратегія синхронної і асинхронної реплікації

У разі синхронної реплікації оновлення однієї репліки обов'язково призводить до оновлення всіх інших реплік, що містять цей же блок даних або послугу. При цьому, оновлення реплік автоматично здійснюється в одній і тій же транзакції, тобто синхронно. Це забезпечує існування єдиної ідентичної репліки послуги на всіх серверах.

Стратегія синхронної реплікації має істотний недолік - формування додаткового навантаження при виконанні всіх транзакцій, необхідність в якій часто відсутня. Також, в разі збільшення кількості реплік, тривалість такої транзакції може значно зрости, що негативно вплине на доступність ресурсів, задіяних при синхронізації.

Перевага асинхронної реплікації полягає в тому, що додаткові витрати реплікації не пов'язані з транзакціями оновлень. Істотним недоліком асинхронної реплікації є можливість несумісності або суперечливості, наявної і зміненої репліки.

У разі застосування асинхронної реплікації оновлення однієї репліки послуги не спричиняє оновлення всіх реплік, що містять один і той же фрагмент, в одній і тій же транзакції. Найчастіше, оновлення інших копій виконується пізніше в результаті зростання потреб користувачів. Таким чином, при асинхронній реплікації вводиться затримка, протягом якої окремі репліки послуги можуть бути фактично неідентичними [77, 93].

Порівняльні характеристики, переваги та недоліки синхронної і асинхронної стратегій реплікації наведені в табл. 4.1.

Таблиця 4.1

#### Порівняння синхронної асинхронної стратегій реплікації

Синхронна реплікація	Асинхронна реплікація
Виконання реплікації в одній транзакції	Виконання реплікації в кількох транзакціях(допускається відкат учасі)
Обмеження на відстань між репліками	Необмежена відстань між репліками
Найбільш високий сервісний рівень захисту інформації	Проміжнийсервісний рівень захисту інформації
Продуктивність рішення залежить від відстаней і обсягів змін даних	Продуктивність не залежитьвідвідстані
Працює при максимальній пропускній здатності каналів передачі даних	Працюєпри середній пропускній здатності комунікаційних каналів



На сьогоднішній день існує безліч методів, що дозволяють нівелювати виникнення протиріч в існуючій і зміненої копії. Основними стратегіями реплікації, які спрямовані на усунення суперечностей реплік є [70]:

1. реплікація на основі первинної копії,
2. активна реплікація,
3. реплікація на основі кворумів.

Стратегія реплікації на основі первинної копії заснована на формуванні єдиної еталонної копії даних. Сервер або безліч серверів (в разі децентралізованого підходу) містять дану еталонну репліку, що виступають в ролі координатора для всіх інших серверів. Всі операції поновлення репліки, що містить ті ж дані, що і еталонна репліка, і зміна однієї копії призводить до перезапису еталонної репліки. Після чого відповідно до одного з існуючих протоколів розповсюдження оновлень [103], вони пересилаються на інші репліки.

При активній реплікації і реплікації на основі кворуму, формування еталонної репліки не виконується. Стратегія активної реплікації схожа з синхронною реплікацією: модифікована репліка направляється на всі сервери, що містять дану репліку. Недоліком активної реплікації є формування надлишкового навантаження і відсутність масштабованості.

Стратегія реплікації на основі протоколів кворуму базується на запиті та підтвердженні можливості читання і запису модифікованої копії. Однак для оновлення існуючої копії локального сервера необхідно направити запит на запис або читання копії і отримати дозвіл на це. Рішення про відмову або дозвіл запису оновленої репліки приймається в результаті голосування більшості серверів, що містять репліки об'єкта. Голосування здійснюється за принципом кворуму: кворуму запису і кворуму читання.

Якщо існує всього  $N$  копій певних даних, з яких  $N_{read}$  число копій, до яких направляється запит на читання даних, і  $N_{write}$  число реплік, для яких необхідно виконання операції запису. То рішення кворуму читання і запису приймається істинним, тільки при наступній умові:

$$\begin{aligned}
 N_{read} &> I \\
 N_{read} + N_{write} &> N; \\
 N_{write} &> N/2.
 \end{aligned}
 \tag{4.1}$$

Виконання наведених вище умов необхідне для обмеження можливості конфлікту подвійного запису, перше вказує на те, що хоча б один сервер містить актуальну репліку.

Порівняльна характеристика механізмів реплікації, спрямованих на виявлення суперечностей між репліками, приведена в табл. 4.2

Таблиця 4.2

Порівняльні характеристики механізмів реплікації, спрямованих на виявлення суперечностей між репліками

Реплікація на основі первинної копії	Активна реплікація	Реплікація на основі кворуму
Простота реалізації	Складність протоколів взаємодії	
Можливість підтримувати працездатність при виході з ладу n-1 реплік	Можливість підтримувати працездатність при виході з ладу (N / 2-1) реплік	
Зберігання інформації про місце розташування реплік сервісу на єдиному локальному реєстрі первинної копії сервісу	Необхідність зберігання інформації про місцезнаходження всіх реплік в локальних реєстрах реплік. Забезпечення механізму обліку змін розташування реплік	
Наявність єдиної точки відмови (в разі виходу з ладу первинної копії сервісу, процес оновлення не зможе бути виконаний, поки не буде призначено нову первинну копію з існуючих реплік сервісу)	Відсутність єдиної точки відмови	
	Погана масштабованість	
	Необхідність забезпечення повністю впорядкованого механізму групової розсилки	При кожному запиті клієнта необхідно погоджувати версію оновлень з більш ніж половиною реплік примірника сервісу
Низька продуктивність через наявність вузького місця у вигляді первинної копії	Проблема реплікації оновлень	Проблема конфлікту подвійного запису

Порівняльна характеристика комплексних стратегій реплікації, що застосовуються з метою підвищення доступності послуг в розподілених мультисервісних мережах, приведена в табл. 4.3.

Таблиця 4.3

Порівняльна характеристика стратегій реплікації, що застосовуються у розподілених мережах

№, п/п	Стратегії	Цілі впровадження	Результат впровадження
1	Динамічна децентралізована реплікація з аналізом ресурсоємності мережі [52]	Зниження споживання енергоресурсів в процесі надання сервісу; економія смуги пропускання; зменшення часу відгуку	Ефективний розподіл залишкової пропускної здатності каналів зв'язку; скорочення енергоспоживання.
2	Динамічна централізована реплікація з оцінкою часу відгуку [58]	Зменшення часу відгуку	Удосконалення нефункціональних характеристик, таких як доступність і продуктивність
3	Статична централізована реплікація [73]	Рівномірне завантаження серверів; розподіл кількості запитів; нормалізація щільності розподілу реплік	Зменшення навантаження на окремі канали передачі даних; розподіл навантаження на сервери; синхронізація реплік
4	Статична децентралізована стратегія реплікації [91]	Підвищення рівня доступності, Балансування запитів (оптимізація розподілу навантаження)	Зменшення часу відгуку, Удосконалення нефункціональних характеристик

Результат аналізу розглянутих комплексних стратегій реплікації, показав, що основними факторами, які необхідно враховувати в процесі формування алгоритму реплікації з метою підвищення доступності послуг є:

- 1) Можливість визначення як глобальної, так і локальної популярності реплік. У разі, якщо популярність перевищує сформований поріг, то виконується операція реплікації.
- 2) Розрахунок кількості необхідних копій і вибір їх місця розташування. Нова копія створюється в тому фрагменті мережі, де згенеровано найбільшу кількість запитів даної копії.
- 3) Розрахунок кількості серверів або обчислювальних пристроїв (VM), яким надано дозвіл на процес запису даних копій. Розрахунок кількості пристроїв, для яких можливе розміщення репліки, ґрунтується на значенні їх поточної продуктивності і ймовірності відмови в доступі.

- 4) Забезпечення несуперечності реплік послуг. Кожен раз при зміні однієї з копій, вона починає відрізнятися від інших. Відповідно для підтримки несуперечності копій сервісів, оновлення необхідно перенести на інші копії.

Як було відзначено, мережі з підтримкою функцій віртуалізації є великомасштабними розподіленими системами, з динамічно мінливою кількістю кінцевих користувачів. Для швидкого реагування і адаптації мережевої інфраструктури до вимог користувачів бажано використовувати динамічні методи реплікації.

#### **4.2. Розробка стратегії динамічної реплікації сервісів**

Архітектура мережі з підтримкою функцій віртуалізації є багат шаровою і включає ряд механізмів, що дозволяють підтримувати коректний розподіл і ефективно взаємне використання обчислювальних ресурсів. Розподіл ресурсів, структурування фрагментів мережі і забезпечення їх ефективної взаємодії є основними завданнями системи MANO. При використанні такого підходу часто формується не цілком пов'язана топологія всієї мережі, а лише окремі її логічні об'єднання - віртуальні локальні мережі (VLAN). При цьому мережі зберігання даних (SAN) або віртуальної мережі SAN (VSAN) виступають узагальненим мережевим ресурсом, доступним для будь-якого пристрою NFV. Логічна схема поділу програмно-конфігурованої мережі на віртуальні локальні мережі наведена на рис.4.4.

Процес надання послуги характеризується наступними взаємодіями: споживач сервісу запитує певний набір послуг, далі NFV контролер, на підставі інформації про поточну кількість послуг і їх місцезнаходження, формує відповідь, в якому вказує дані для доступу до послуги. У разі впровадження нової послуги постачальник сервісу передає інформацію про неї контролеру, який в свою чергу заносить дані про послугу до реєстру сервісів [67].

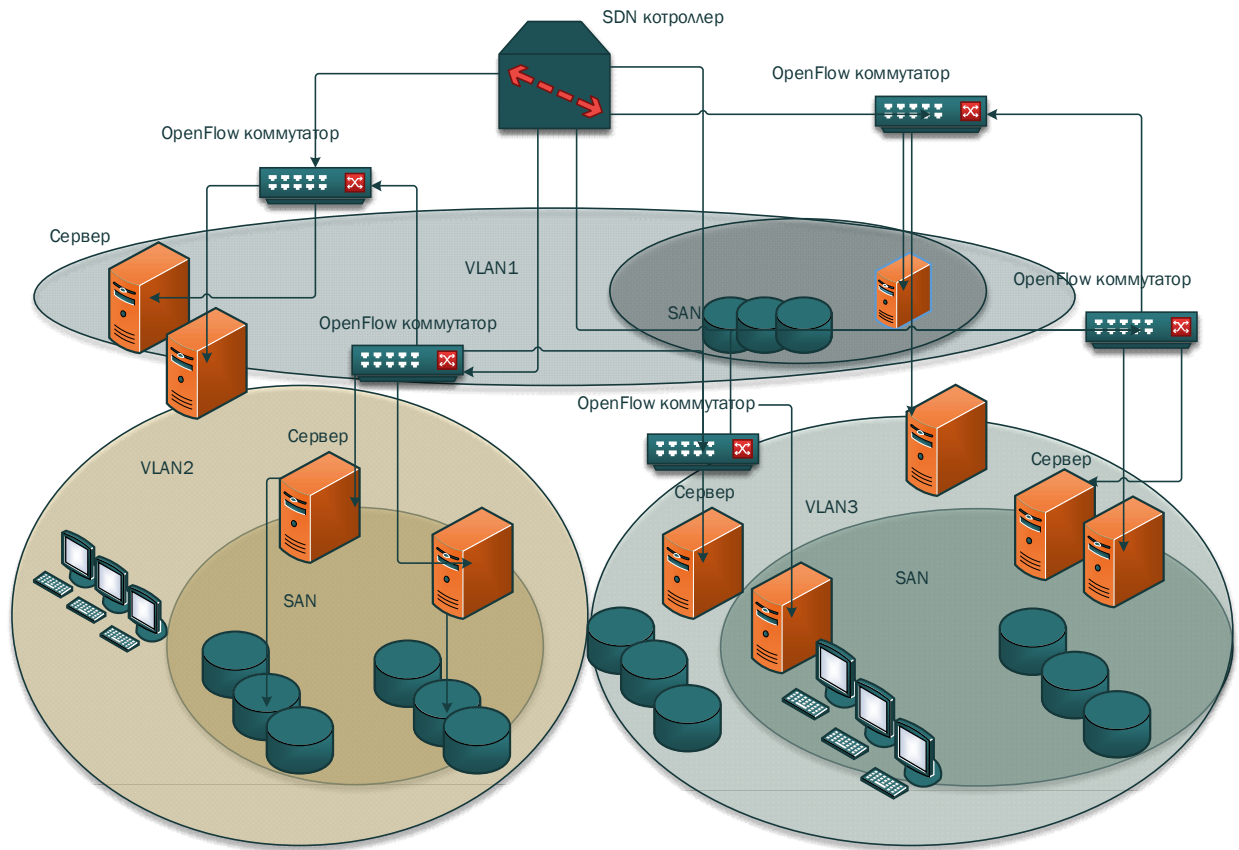


Рис.4.4. Розподіл зон в мережах з підтримкою віртуалізації функцій

В такому випадку, множину споживачів послуг може бути визначено як множина  $User = \{user_1, user_2, \dots, user_m\}$ , безліч комплексних послуг, що надаються сервіс провайдером, задається як  $CS = \{CS_1, CS_2, \dots, CS_n\}$ , де  $CS_i$  - послуга, сформована шляхом композиції окремих задач  $cs_{ki}$ :  $CS_i = \{cs_{1i}, cs_{2i}, \dots, cs_{ki}\}$ . Склад комплексної послуги може варіюватися. Наприклад, в разі використання послуг IP-телефонії з підтримкою відео (IaaS) користувач може запросити або повний склад послуги -  $CS_{videoi} = \{cs_{datai}, cs_{voicei}, \dots, cs_{videoi}\}$ , або частковий -  $CS_{voicei} = \{cs_{datai}, cs_{voicei}\}$ . Сценарій формування комплексної послуги в разі розподілу атомарних сервісів між різними фрагментами мережі наведений на рис.4.5.

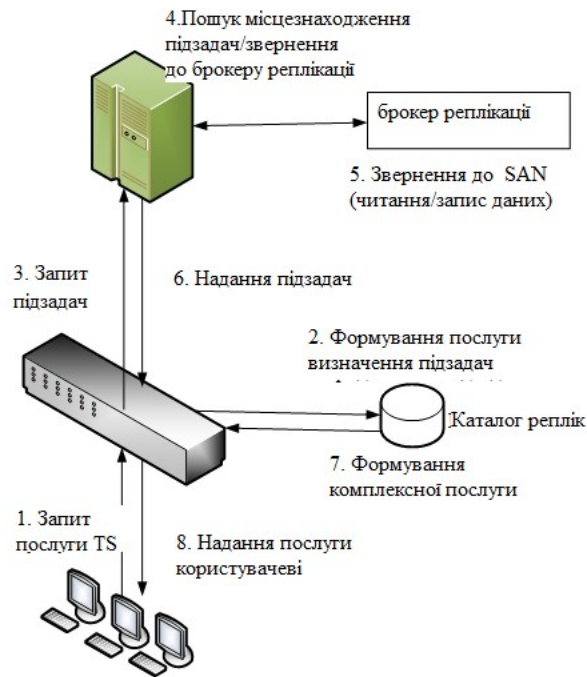


Рис.4.5. Сценарій надання комплексної послуги в мережах з підтримкою віртуалізації мережевих функцій

Контролер формує рішення про формування та надання комплексної послуги на підставі інформації про локації послуги  $loc(CS_i)$  чиї складових  $cs_{ki}$ ,  $loc$  має на увазі ір адресу бази даних чи сервери, які містять послугу), вартості надання послуги  $CS_i$  чиї складових  $cs_{ki}$ ,  $c$ , часу відгуку  $t$ , продуктивність вузла, на якому розміщена запитувана послуга  $p$ , надійність  $r$  і інших опціональних компонент.

Таким чином, каталог послуг  $LIST(CS_i)$ , що надаються сервіс провайдером, характеризується множиною наступного вигляду:

$$\begin{aligned}
 LIST(CS_i) &= \{loc(CS_i), c(CS_i), t(CS_i), p(CS_i), r(CS_i), \dots\}; \\
 \forall tk_i, loc(tk_i) &\in loc(CS_i); \\
 p(CS_i) &= \min p(cs_{ki}); \\
 c(CS_i) &= \sum_1^j tk_k.
 \end{aligned} \tag{4.2}$$

де  $\{c(CS_i), t(CS_i), p(CS_i), r(CS_i)\} \in Q$ ,  $Q$  – безліч показників якості послуги  $CS_i$ . Значення даних показників є вирішальним при виборі чи формуванні репліки послуги.

Види послуг, що надаються в межах однієї зони обслуговування можуть істотно відрізнятися одна від одної [89], що обумовлює наявність кількох каталогів послуг  $LIST[CS_i]$  з різним змістом. Кожний каталог послуг  $LIST_i[CS]$  закріплений за певною VLAN чи LAN, сформованої за допомогою об'єднання декількох OpenFlow комутаторів [15] і має свою область обслуговування  $SA_i$ , межі якої визначаються постачальником послуг в процесі преконфігурації:

$$\begin{aligned} LIST_i[CS] &= \{LIST_1[CS_{i_1}], LIST_2[CS_{i_2}], \dots, LIST_n[CS_{i_n}]\}, \\ SA_{ij}[CS] &= \{SA_{i_1}[CS_{i_1}], SA_{i_2}[CS_{i_2}], \dots, SA_{i_n}[CS_{i_n}]\}, \\ SA_i[CS], SA_i[CS] &\supset \{LIST_1[CS_{i_1}], LIST_2[CS_{i_2}], \dots, LIST_n[CS_{i_n}]\}. \end{aligned} \quad (4.3)$$

де  $LIST_i[CS]$  – множина реєстрів послуг, що надаються в мережі,  $AS_i[CS]$  – безліч логічних зон обслуговування мережі,  $CS$  – безліч комплексних сервісів, що надаються в мережі,  $i$  – номер області обслуговування,  $n$  – кількість реєстрів сервісів і областей обслуговування в мультисервісній мережі  $q$  – номер реєстру сервісів. Номер реєстру сервісів збігається з областю обслуговування.

У зв'язку з тим, що мережі з підтримкою функцій віртуалізації можуть включати в себе безліч географічно розподілених підмереж різних розмірів, область обслуговування кожного каталогу послуг може значно варіюватися: вона може включати як об'єднання декількох OpenFlow комутаторів, так і декількох контролерів. Приклад розміщення каталогів сервісів в мультисервісній мережі наведено на рис. 4.6.

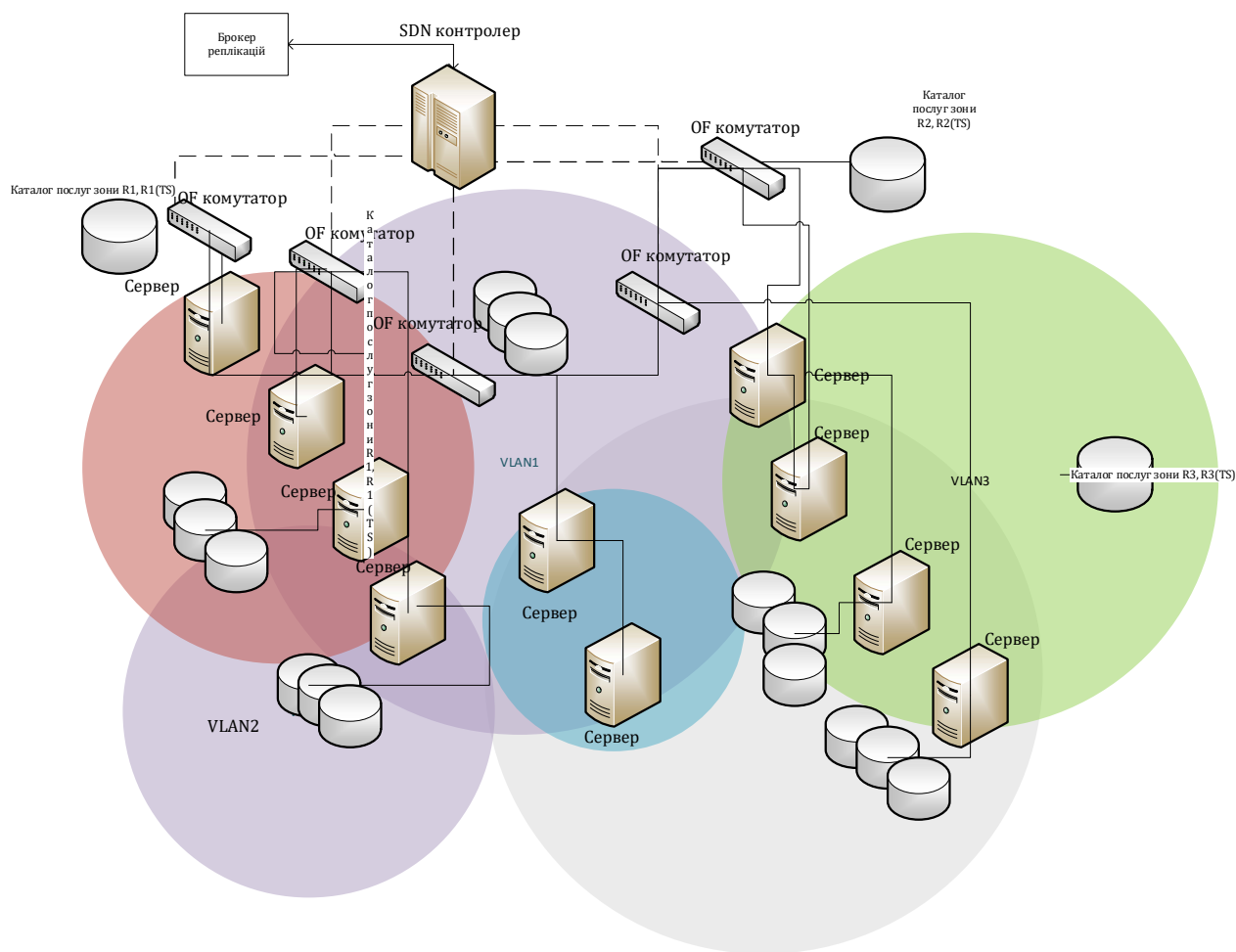


Рис.4.6. Розміщення каталогів послуг в мережах з підтримкою функцій віртуалізації

При реєстрації в каталозі нової послуги  $LIST_j[CS]$ , їй присвоюється певний порядковий номер  $i$ , при цьому  $i$  є локальним порядковим номером послуги. Глобально послуга визначена за допомогою двох індексів  $i$  і  $j$  та представлена як  $CS_{ji}$ . При видаленні послуги, порядковий номер наступних послуг може змінитися. Порядковий номер послуги може бути змінений у разі видалення однієї з попередніх послуг з реєстру. Однак номер області обслуговування  $j$ , за яким закріплена послуга, залишається постійним протягом життєвого циклу послуги. Такий тип індексації послуг застосуємо до всіх реєстрів сервісів, розмножених в мережі.

Процес пошуку послуг в мережах з підтримкою функцій віртуалізації наведено в розділі 1.



Зростання популярності послуги, розташованої на віддаленому сервері, часто призводить до зниження рівня її якості. Зниження рівня якості обумовлено перевантаженням каналів зв'язку, затримками в процесі обробки і надання послуги. Збільшення кількості запитів призводить до перевантаження сервера і виникнення відмов в обслуговуванні, а, отже, і зменшення рівня доступності послуги.

Таким чином, розробка стратегії реплікації, яка дозволяє враховувати вимоги користувачів і забезпечувати гнучкий перерозподіл реплік відповідно до рівня їхньої популярності і поточних характеристик мережі (завантаженістю серверів і каналів передачі даних) є актуальною.

#### **4.3. Визначення ключових показників якості в процесі формування репліки комплексного / атомарного сервісу**

Відповідно до п. 4.1 уявлення про доступність сервісів формується на підставі значення таких показників, як  $\{t(CS_i), p(CS_i), r(CS_i)\}$ . Виходячи з можливих змін значень компонент безлічі  $\{t(CS_i), p(CS_i), r(CS_i)\}$ , сформовані такі визначення доступності:

- миттєва доступність  $Av_{CS_i}(t)$  ;
- стаціонарна доступність  $Av_{CS_i}(T_{lc})$ , де  $T_{lc}$  - час життя послуги.

Миттєва доступність  $Av_{CS_i}(t)$  послуги визначається як ймовірність того, що мережі або фрагмент мережі знаходиться в працездатному стані в момент часу і дозволяє коректно надавати послугу, навіть в разі наявності ряду обмежень:

$$a_{TS_i}(t) = \begin{cases} p\{a_i(t) = p(up) | a_0(t) = p(down)\}, \text{ при наявності збоїв,} \\ r_{TS_i}(t), \text{ у випадку відсутності збоїв.} \end{cases} \quad (4.4)$$

де  $r_{CS_i}(t)$  - надійність мережі в момент часу  $t$ ,  $p(up)$  - ймовірність того, що система знаходиться в працездатному стані,  $p(down)$  - ймовірність того, що послуга не може бути надана цим вузлом.

Стаціонарна доступність послуги визначається як ймовірність того, що система знаходиться в стані, здатному надати послугу з заданою якістю сервісу протягом всього життєвого циклу послуги:

$$Av_{CS_i}(T_{lc}) = \lim_{t \rightarrow T_{lc}} a_i(t) = \lim_{t \rightarrow T_{lc}} \frac{\int_0^t a_i(t) dt}{t}, \quad (4.5)$$

У процесі формування стратегії реплікації первинною вимогою є визначення поняття доступності реплік і вибір критеріїв реплікації. Відповідно до особливостей процесу надання послуг в розподілених мультисервісних мережах введено два додаткових визначення доступності: доступність репліки атомарної підзадачі і доступність репліки комплексної послуги.

Наявність одного екземпляра даних або підзадачі на сервері визначено як  $loc(cs_{ki})$ . Ймовірність того, що підзадача доступна на даному сервері, визначена як  $P(loc(cs_{ki}))$ , а ймовірність відмови в доступі  $1 - P(loc(cs_{ki}))$ . У процесі реплікації підзадачі  $ts_{ki}$  може бути створено  $n$  реплік  $n \cdot cs_{ki}$ , в такому випадку узагальнена ймовірність доступності послуги може бути визначена наступним чином:

$$P(cs_{ki}) = (1 - P(loc(cs_{ki})))^n, \quad (4.6)$$

а ймовірність відмови в обслуговуванні:

$$P(cs_{ki}) = 1 - (1 - P(loc(cs_{ki})))^n. \quad (4.7)$$

Доступність комплексного сервісу - це здатність системи коректно формувати і надавати різні сервіси шляхом об'єднання кількох атомарних

підзадач в єдине ціле із заданим рівнем якості при всіх можливих обмеженнях системи.

У разі, якщо комплексна послуга  $CS_i$  складається з безлічі атомарних підзадач  $\{cs_{1i}, cs_{2i}, \dots, cs_{ki}\}$ , розподілених між різними вузлами і ймовірність доступності репліки кожної підзадачі для всіх обчислювальних пристроїв з однаковою конфігурацією однакова, то ймовірність доступності комплексної послуги може бути розрахована наступним чином:

$$P(CS_i) = ((1 - P(\text{loc}(cs_{ki})))^n)^k, \quad (4.8)$$

а ймовірність відмов у доступі:

$$P(CS_i) = (1 - (1 - P(\text{loc}(cs_{ki})))^n)^k. \quad (4.9)$$

Таким чином, з метою забезпечення доступності послуг в процесі формування рішень про реплікації, необхідно вирішити такі завдання:

1. Момент прийняття рішення про формування реплік;
2. Визначення кількості реплік і їх розташування в мережі;
3. Момент прийняття рішення про видалення репліки.

#### 4.3.1. Визначення моментів часу формування реплік

Ухвалення рішення про реплікації виконується на підставі аналізу рівня популярності послуги. З метою індикації зміни популярності послуги пропонується введення такого атрибута брокера реплік, як лічильник запитів послуги з урахуванням джерела запиту –  $Cnt(\sum_{i=1}^n Cl_n(\text{SourceIP}), CS_{ij})$ , де в якості джерела розглядається адреса сервера (NAT). Введення лічильника дозволяє підрахувати кількість запитів, що надійшли від користувачів, з урахуванням їх розташування в мережі  $\sum_{i=1}^n Cl_n(\text{SourceIP})$ , до каталогу послуг  $LIST_j[CS]$  для

отримання інформації про місце розташування послуги  $TS_i$  чи множині підзадач  $\{CS_{1i}, CS_{2i}, \dots, CS_{ki}\}$ . В такому випадку, брокер реплік володіє такою інформацією:

$$BR_{SDN} : [loc(CS_i), c(CS_i), t(CS_i), p(CS_i), r(CS_i), CNT, \dots], \quad (4.10)$$

де  $cnt(\sum_{i=1}^n Cl_n(SourceIP), CS_{ij}) \in CNT$  – множина лічильників сервісів каталогу послуг.

Ключове значення при цьому має поріг реплікації  $Th(\sum CS_{ij})$  [120]. Вибір значення порога реплікації впливає на частоту формування і розподілу копій послуги. Так як в мережах, з підтримкою концепції NFV, вимоги до якості послуг можуть суттєво відрізнятися [104], це призводить до значних розбіжностей порогів реплікації для різного типу послуг. Вибір порога реплікації також залежить від обсягу даних, які повинні бути передані в процесі надання послуги, значення поля ToS і поточного значення доступності сервісу:

$$Th(\sum CS_{ij}) = (CS_{ij}(length), CS_{ij}(ToS), Av(CS_{ij})), \quad (4.11)$$

де  $CS_{ij}(length)$  - об'єм послуги чи блоку даних у бітах,  $CS_{ij}(ToS)$  - пріоритет надання послуги,  $Av(CS_{ij})$  - поточний рівень доступності послуги.

Змінним параметром, на підставі якого можна зробити висновок про необхідність реплікації, для кожного типу послуги виступає показник доступності послуги. Доступність послуги залежить від поточної продуктивності сервера  $Perf_{loc}(CS_{ij})$  чи кластера серверів  $n \cdot Perf_{loc}(CS_{ij})$ . Взаємозв'язок між порогом реплікації і продуктивністю сервера, на якому розміщена послуга, може бути задана в такий спосіб:

$$Th(\sum CS_{ij}) = k \cdot Perf_{loc}(CS_{ij}), \quad (4.12)$$

де  $k$  – коефіцієнт, що визначає вимоги до реплікації для послуги  $CS_{ij}$ . Якщо значення  $k < 1$ , то реплікація виконується із запобіганням відмов в обслуговуванні. Якщо  $k > 1$ , то реплікація виконує для розвантаження сервера при значному збільшенні потоку заявок.

Брокером реплік з періодичністю  $T_{update}$  проводиться підрахунок числа запитів кожного типу сервісу. У разі, якщо виявлено, що кількість запитів на надання сервісу, що надійшли з одного фрагмента мережі, перевищує поріг реплікації  $cnt(\sum_{i=1}^n Cl_n(SourceIP), CS_{ij}) > Th(\sum CS_{ij})$  і/чи призводить до погіршення якості послуги, то брокер приймає рішення про необхідність реплікації сервісу. З метою підвищення доступності, сервіс реплікується на локальний сервер.

Реплікація великих обсягів даних автоматично не виконується: рішення про реплікації великих обсягів даних приймається на основі політик QoS, прийнятих сервіс провайдером [43, 47].

#### 4.3.2. Визначення кількості реплік і їх розташування в мережі

Найчастіше запитувана користувачем послуга знаходиться на віддаленому сервері. Таким чином, серверу необхідно обслужити як ряд локальних запитів, так і ряд віддалених запитів без прогнозованого рівня доступності. Поряд з цим, при зростанні кількості запитів, продуктивність віддаленого сервера може значно скоротитися, що може спричинити за собою деградацію якості послуг.

Кількість реплік, створення яких необхідно для підтримки заданого рівня доступності сервісу запропоновано розраховувати наступним чином:

$$Av_{new}(CS_i) = Av_{old}(CS_i) + K(1 - Av_{old}(CS_i)), K \in [0, 1], \quad (4.13)$$

де  $Av_{old}(CS_i)$  - доступність послуги до реплікації,  $Av_{new}(CS_i)$  - доступність послуги після реплікації,  $K$  - ваговий коефіцієнт.

Значення вагового коефіцієнта розраховується наступним чином:

$$K = \frac{F(CS_i)}{\sum_{k=1}^n F(CS_k)}, CS_i \notin \{CS_k\}, \quad (4.14)$$

де  $n$  – загальна кількість реплік послуг, не включаючи послугу  $CS_i$ .

З метою забезпечення гарантованої якості послуг в процесі реплікації, дані переносяться в той фрагмент мережі, де популярність послуги перевищує граничну межу. На підставі територіальної ознаки репліки можуть бути розділені на локальні і віддалені.

Під локальною реплікацією мається на увазі створення копій і збереження копій послуги в тому ж фрагменті мережі, де розміщується еталонна репліка. Формування необхідної кількості реплік в локальному фрагменті мережі призводить до скорочення службового трафіку, а також знижує навантаження на мережеві елементи. Також, створення додаткової репліки при локальній реплікації дозволяє підвищити доступність і продуктивність запитуваного сервісу, шляхом перерозподілу запитів між репліками.

Якщо під час надходження запиту послуга  $CS_{ij}$  є недоступною, то запит перенаправляється до іншої репліки, при цьому дана дія буде прозорою для кінцевого користувача.

Дистанційна реплікація передбачає формування копій послуги та їх перенесення в той фрагмент мережі, де її популярність висока. Необхідність у вилученій реплікації виникає в тому випадку, коли кількість запитів з однієї ділянки мережі до певного віддаленого ресурсу значно зростає. Тому, з метою зменшення трафіку переданого мережею, а також скорочення часу доступу до ресурсу, запускається процес реплікації.

Процес вилученої реплікації призводить до зниження зайнятості каналів передачі даних, а також зменшення значення затримок поширення і обробки,

завдяки чому вдається підвищити масштабованість мережі і доступність запитуваних послуг.

Таким чином, визначення кількості реплік, які необхідно залишити в локальній зоні обслуговування і кількість реплік, які необхідно перемістити на віддалений фрагмент мережі дозволить підвищити ефективність розподілу ресурсів і забезпечити гарантований рівень QoS.

Кількість реплік послуги, розподілених між фрагментами мережі, визначається наступним чином:

$$n_{R_i}(CS_i) = \left\lfloor \frac{F_{R_i}(CS_i)}{\sum F_R(CS_i)} \times mCS_i \right\rfloor, \quad (4.15)$$

де  $F_{loc_k}(CS_i)$  - фактор реплікації, що відповідає певному фрагменту мережі,  $\sum F(CS_i)$  - сумарний фактор реплікації,  $mCS_i$  - загальна кількість нових реплік послуги.

Приклад розподілу реплік послуги наведено на рис. 4.7. У даному сценарії передбачається, що існує кілька реплік послуги  $CS_i$  в тому ж фрагменті мережі, де розташована послуга. На підставі аналізу популярності сервісу сформовано рішення про необхідність формування чотирьох додаткових копій послуги. В рамках наведеного прикладу фактор реплікації  $F(CS_i)$  послуги для різних зон склав:

$$F_{R_1}(TS_i) = 3,14,$$

$$F_{R_2}(TS_i) = 2,57,$$

$$F_{R_3}(TS_i) = 1,43,$$

...

$$F_{R_n}(TS_i) = 0.$$

Розподіл реплік на підставі даної інформації здійснюється наступним чином:

$$n_{R_1}(TS_i) = \left\lfloor \left( \frac{3,14}{3,14 + 2,57 + 1,43} \right) \times 4 \right\rfloor = 2,$$

$$n_{R_2}(TS_i) = \left\lfloor \left( \frac{2,57}{3,14 + 2,57 + 1,43} \right) \times 4 \right\rfloor = 1,$$

$$n_{R_3}(TS_i) = \left\lfloor \left( \frac{1,43}{3,14 + 2,57 + 1,43} \right) \times 4 \right\rfloor = 1,$$

$$n_{R_m}(TS_i) = 0.$$

Виходячи з отриманих результатів, дві сформовані репліки залишаються в локальному фрагменті мережі, а три копії розподіляються між  $R_2$  и  $R_3$ .

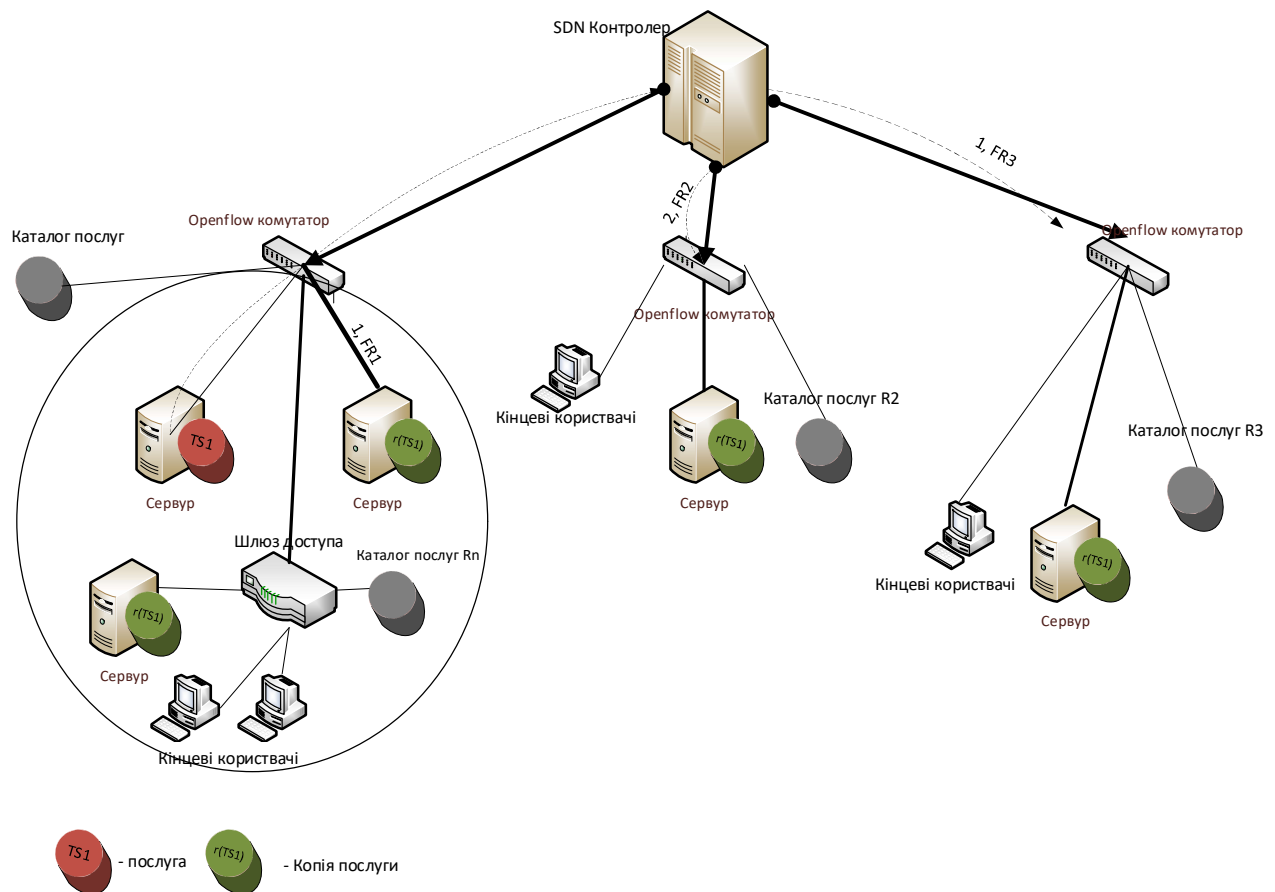


Рис.4.7. Розподіл копій послуги в залежності від фактора реплікації

Перед початком проведення операції реплікації, реєстр переглядає свою базу даних на наявність сервера, який має ресурси, необхідні для розміщення даної репліки. У випадку наявності такого серверу, реєстр надсилає запит до сервісу  $CS_i$  на створення копії, при цьому в запиті він вказує місце розташування для майбутньої копії.



Після проведення операції реплікації, реєстру відправляється відповідь, в якій вказується, чи була успішно виконана дана процедура і адреса для реєстрації репліки.

Реплікація може бути скасована і реєстр направлений до перевірки лічильника наступного сервісу у випадку, коли в базі даних реєстру не виявлені доступний сервер.

### 4.3.3. Визначення моменту прийняття рішення про видалення репліки сервісу

Серед факторів, які впливають на вибір порогу реплікації відповідно до (4.11) і (4.12) можна виділити наступні: популярність послуги, тип послуги, задані параметри QoS, поточний стан мережевих елементів.

З метою підтримки високого рівня продуктивності мережі поряд із завданням формування реплік необхідно розглянути завдання їх видалення. Завдання видалення реплік є протилежним завданню прийняття рішення про необхідність створення.

Поріг видалення репліки  $Tresh(CS_{ij})$  визначається брокером реплікації динамічно, на підставі інформації про зниження продуктивності мережі, відсутності популярності послуги, блокування обчислювальних пристроїв через неефективне використання RAM. З метою визначення рівня затребуваності послуги брокер реплікації поряд з визначенням порогу формування репліки встановлює поріг видалення. Поріг визначення встановлюється за допомогою відповідного лічильника  $cnt_{del}(\sum_{i=1}^n Cl_n, CS_{ij}(DestIP))$ , где  $\sum_{i=1}^n Cl_n$  - сумарна кількість запитів на надання, як комплексної послуги, так і її фрагментів,  $CS_{ij}(DestIP)$  - місцезнаходження послуги (адреса сервера або кластера серверів, де послуга не користується популярністю). Таким чином, брокер реплікації містить наступну інформацію про послуги:

$$BR_{NFV} : [loc(CS_i), c(CS_i), t(CS_i), p(CS_i), r(CS_i), CNT, CNT_d] , \quad (4.16)$$

де  $cnt_{del}(\sum_{i=1}^n Cl_n, CS_{ij}(DestIP)) \in CNT_d$  – множина лічильників запитів послуги, що ідентифікують її популярність на певному сервері чи кластері серверів.

Якщо в ході перевірки значення лічильника  $cnt_{del}(\sum_{i=1}^n Cl_n, CS_{ij}(DestIP))$  буде менше порогу віддалення  $Th_d(\sum CS_{ij})$ :  $cnt_{del}(\sum_{i=1}^n Cl_n, CS_{ij}(DestIP)) < Th_d(\sum CS_{ij})$ , то брокер приймає рішення про необхідність віддалення репліки сервісу.

Як і поріг реплікації, значення порога видалення залежить від ряду факторів:

$$Th_d(\sum CS_{ij}) = (TS_{ij}(length), CS_{ij}(ToS), Av'_{TS_{ij}}(t), CS_{ij}(r\_length)), \quad (4.17)$$

де  $CS_{ij}(length)$  - об'єм послуги чи блоку даних у бітах,  $CS_{ij}(ToS)$  - тип послуги,  $Av'_{TS_{ij}}(t)$  - доступність послуги в даний момент часу на певному сервері,  $CS_{ij}(r\_length)$  - трафік, що утворюється в процесі реплікації послуги. Значення  $CS_{ij}(r\_length)$  залежить від частоти оновлення послуги та обсягів реплікованих даних.

При видаленні репліки комплексної послуги або атомарного сервісу слід проаналізувати залежність між  $CS_{ij}(length)$  и  $CS_{ij}(r\_length)$ . Зважаючи на вказані значення, необхідно слідкувати за тим, щоб видалення послуги не нівелювало вигоду процесу реплікації. Наприклад, якщо послуга знаходиться на віддаленому сервері і в процесі її реплікації повинні бути передані великі обсяги даних, то видалення послуги не бажане і навпаки, якщо кількість даних, переданих під час синхронізації реплік послуги велика, а популярність послуги невелика, то репліка видалається.

Співвідношення, що дозволяє визначити поріг видалення репліки, задане наступним чином:

$$\sum n(CS_{i_{\text{mod}}}(\text{lenght})) < \sum m(CS(\text{lenght})), \quad (4.18)$$

де  $\sum n(CS_{i_{\text{mod}}}(\text{lenght}))$  – сумарний об'єм даних, які будуть надсилатися на репліку послуги  $TS_{ij}$  при необхідності синхронізації реплік;  $\sum m(CS(\text{lenght}))$  – сумарний обсяг даних, які будуть пересилатися між віддаленим обчислювальним пристроєм або користувачем і первинним екземпляром послуги в разі відсутності репліки.

Для визначення  $\sum n(CS_{i_{\text{mod}}}(\text{lenght}))$  і  $\sum m(CS(\text{lenght}))$  вводяться наступні позначення:

$N_{rep}$  – кількість реплік первинного примірника послуги;  $T_{dt}$  – тривалість блокування доступу до послуги через процес оновлення,  $n_{mp}$  – кількість транзакцій в секунду;  $Op$  – безліч змін для репліки, породжуваної транзакцією;  $d$  – потужність довільної кількості  $Op$ ;  $f$  – частота синхронізації реплік.

$$\sum n(CS_{i_{\text{mod}}}(\text{lenght})) = T_{dt} \cdot n_{mp} \cdot d \cdot f, \quad (4.19)$$

$\sum m(CS(\text{lenght})) | r = 0$  – загальна кількість змін, які будуть надсилатися на репліку первинного примірника сервісу;

$$(\sum m(CS(\text{lenght})) | r = 0) = N_{rep} \cdot T_{dt} \cdot n_{mp} \cdot d \cdot f, \quad (4.20)$$

$V$  – множина операцій, які повинен виконати первинний примірник сервісу для виконання запиту користувача;  $v$  – потужність довільної кількості  $V$ ;

$$\sum n(CS_{i_{\text{mod}}}(\text{lenght})) = N_{cl} \cdot T_{dt} \cdot n_{mp} \cdot v, \quad (4.21)$$

$$\left(\sum m(CS(lenght)) \mid r = 0\right) = \sum_{l=1}^{N_{MR}} N_{cl}^l \cdot T_{dt} \cdot n_{mp} \cdot v, \quad (4.22)$$

де  $N_{cl}$  – кількість користувачів області  $ER_j$ , що бажають отримати доступ до послуги;  $N_{MR}$  – кількість областей, з яких приходять запити користувачів на отримання доступу до первинного примірника сервісу.

Маючи інформацію про обсяг трафіку, створюваного оновленнями між репліками і трафіком, який буде виникати за рахунок виконання запитів віддалених користувачів в разі відсутності реплік сервісу можна оцінити необхідні рівні порога реплікації і видалення з метою отримання максимальної ефективності методу реплікації.

Алгоритм динамічної реплікації (стадія формування і видалення реплік) наведено на Рис.4.8.

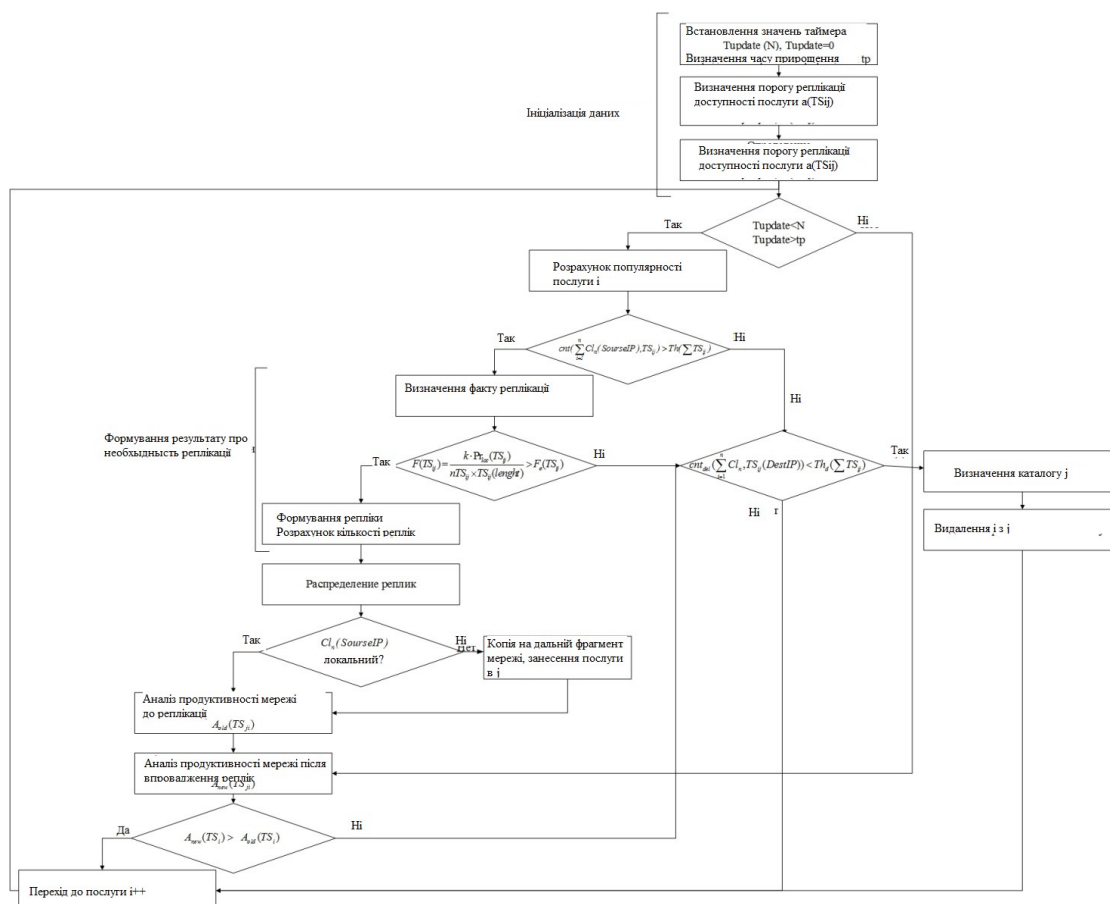


Рис. 4.8. Алгоритм динамічної реплікації (процес формування і видалення реплік) в мультисервісних мережах з підтримкою концепції NFV

Параметри, введені в рамках алгоритму:  $N$  - тривалість тимчасового інтервалу протягом якого брокер реплікації зберігає інформацію значення лічильника;  $tp$  – час збільшення при зборі інформації про популярність послуги,  $t_p \leq N$ ;  $i$  – номер послуги в реєстрі;  $j$  – номер каталога послуг;  $cnt(\sum_{i=1}^n Cl_n(SourceIP), CS_{ij})$  – лічильник сервісу  $S_i$ ;  $F(CS_i)$  - фактор реплікації,  $Th(\sum CS_{ij})$  - поріг реплікації,  $Th_d(\sum CS_{ij})$  - поріг віддалення репліки,- фактор реплікації,  $Cl_n(SourceIP)$  - джерела запитів найбільшої інтенсивності, - поріг  $F_e(CS_{ij})$  - фактор реплікації,  $cnt_{del}(\sum_{i=1}^n Cl_n, CS_{ij}(DestIP))$  - лічильник запитів на послуги, які передані на обслуговування серверу  $CS_{ij}(DestIP)$ .

Якщо репліка є віддаленою, то реєстр LIST відправляє повідомлення про видалення репліки на віддалений реєстр  $R_{rem j}$ , в області обслуговування якого зберігається первинний примірник сервісу  $S_i^{rem j}$ . Після отримання повідомлення, віддалений реєстр  $R_{rem j}$  видалає з опису первинного екземпляра сервісу інформацію про позначки копії і її місце розташування. По закінченню процедури перевірки всі лічильники обнуляються  $cnt(R, F_i) = 0$ , а таймеру  $T$  присвоюється значення  $N$ .

#### 4.4. Висновки до четвертого розділу

1. В даному розділі проведений аналіз існуючих методів підвищення доступності сервісів, визначено, що реплікація є одним з найбільш ефективних методів її підвищення. Застосування даного методу дозволяє підвищити доступність і продуктивність сервісів, а також поліпшити масштабованість географічно розподіленої мультисервісної мережі.

2. Розроблено метод динамічної реплікації, який дозволяє підвищити доступність послуг в NFV мережах, в разі зростання їхньої популярності. Застосування методу передбачає можливість виконання двох видів реплікації:

локальної і віддаленої. Основними регуляторами алгоритму динамічної реплікації виступають поріг реплікації і поріг видалення.

3. Для визначення рівня порогу реплікації та рівня порогу видалення проведена оцінка таких показників, як продуктивність сервісу, тип послуги, політика провайдера сервісу, а також відношення обсягу трафіку, створюваного пересиланням оновлень між реплікою і первинним екземпляром сервісу та обсягу трафіку, створюваного пересиланням даних між користувачами і первинним екземпляром сервісу, в разі відсутності реплік.

4. Метод первинного архівування дає можливість реалізувати сувору несуперечливість копій сервісів з можливим зниженням доступності послуг у разі зростання кількості реплік. Цей метод був модифікований з метою застосування в NFV мережах, а також для виконання вимог щодо забезпечення несуперечності копій сервісів.

5. Розроблено комплексний алгоритм реплікації сервісів, що покриває всю область завдань з реалізації реплікації і забезпечення несуперечності реплік сервісів.

## ВИСНОВКИ

У дисертаційній роботі запропоновано вирішення актуального науково-прикладного завдання, яке полягає в удосконаленні механізмів забезпечення необхідного рівня якості обслуговування в мультисервісних мережах з підтримкою віртуалізації мережевих функцій за рахунок удосконалення моделей і методів аналізу, формування та надання сервісів.

Відповідно до результатів, отриманих в ході вирішення поставленого науково-прикладного завдання, зроблений ряд наступних висновків:

1. Аналіз поточного стану та перспектив розвитку мультисервісних мереж з підтримкою функцій віртуалізації показав, що ключовими в процесі надання сервісів є наступні чинники: ефективність системи управління і фізичні характеристики мережевого обладнання та каналів передачі даних. Надання сервісів заданого рівня якості можливо лише в разі узгодженої взаємодії компонентів мережевої інфраструктури і коректного розподілу мережевих ресурсів.

2. Аналіз нефункціональних показників, що впливають на якість надання сервісів у мережах з підтримкою віртуалізації показав, що такі нефункціональні характеристики, як затримка і час відгуку істотно впливають на якість обслуговування. Існуючий на сьогоднішній день інструментарій не дозволяє забезпечити повноцінний аналіз даних характеристик для мереж з підтримкою віртуалізації мережевих функцій. Для моделювання динаміки мережевих характеристик в процесі надання сервісів запропоновано застосовувати теорію мережевого обчислення.

3. Розроблено метод динамічної реплікації сервісів в мережах з підтримкою функцій віртуалізації мережевих ресурсів, який за рахунок використання теорії мережевого обчислення дозволяє підвищити доступність і ефективність надання сервісів за рахунок своєчасного формування, видалення копій різних типів сервісів і їх раціонального розподілу між фрагментами мережі.

4. Проведений аналіз методів формування комплексних сервісів показав широкі можливості в управлінні рівнем якості обслуговування за рахунок зміни складу комплексних сервісів. Тому було удосконалено метод формування та управління складом комплексних сервісів в мережах з підтримкою функцій віртуалізації шляхом урахування поточного стану ресурсів мережі під час формування і управління складом сервісів, що дозволяє забезпечити необхідний рівень якості обслуговування у відповідності до погодженого договору про якість обслуговування. Удосконалений метод дозволяє формувати і управляти складом комплексного сервісу за рахунок обліку показників якості складових комплексного сервісу, моніторингу та своєчасної їх заміни, що дозволяє забезпечити необхідний рівень QoS.

5. Удосконалено метод балансування навантаження в мережі з підтримкою віртуалізації мережевих ресурсів за рахунок ранжування сервісів по якості їх надання та необхідній пропускній здатності та послідовного перерозподілу наявних вільних мережевих ресурсів, що дозволяє на основі аналізу поточного навантаження як на віртуальну, так і на фізичну складову мультисервісної мережі забезпечити заданий рівень якості обслуговування з меншими накладними витратами.

6. Набув подальшого розвитку метод динамічного розподілу пропускної здатності каналів передачі даних у NFV мережах, що базується на принципах розподілення залишкової пропускної спроможності. Застосування даного методу дозволяє уникнути повного завантаження та перевантаження каналу передачі.

7. З метою забезпечення необхідного рівня доступності сервісів, згідно з розробленим комплексом моделей і методів, запропонована методика управління запитами користувачів в мережах з підтримкою віртуалізації мережевих функцій.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abdulghafoor Raed Yahya Performance analysis of load balancing mechanisms in Software-Defined Networking / O.B.Tkachova, Abdulghafoor Raed Yahya, Hassan Mohamed Muhi-Aldeen // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2016. – №187– P.197–203.
2. Al-Shaer E., Al-Haj S. FlowChecker: Configuration Analysis and Verification of Federated Openflow Infrastructures // Proceedings of the 3rd ACM Workshop on Assurable and Usable Security Configuration, (SafeConfig '10). – Chicago, Illinois, USA, 2010. – pp. 37-44.
3. An Approach to QoS-Aware Service Selection in Dynamic Web Service Composition / Yang, Y., Tang, S., Xu, Y., Zhang, W., Fang, L.// In: 3rd IEEE International Conference on Networking and Services (ICNS), – New York, USA, 2007 –pp.18-23.
4. Andi B. Introduction to OpenFlow [Електроннийресурс] / Bean Andi // Open Networking Foundation. –2011. –Режим доступа: <https://www.opennetworking.org/news-and-events/latest-news/195-introduction-to-openflow-introduction-to-openflow>.
5. Anis Yousefi. Request Replication: An Alternative to QoS Aware Service Selection / Anis Yousefi and Douglas G. Down // Proceedings of the 2011 IEEE International Conference on Service-Oriented Computing and Applications, SOCA '11, – pp. 1-14
6. Ardagna, D. Global and Local QoS Constraints Guarantee in Web Service Selection / Ardagna, D., Pernici, B. // IEEE International Conference on Web Services, 2005.– pp. 805-806.
7. Bengtsson J., Larsen K., Larsson F., Pettersson P., Yi W. UPPAAL – a tool suite for automatic verification of real-time systems // Hybrid Systems III, Lecture Notes in Computer Science. Vol. 1066. – Springer Berlin Heidelberg, 1996. – Pp. 232–243.

8. Boucadair M. [Электронный ресурс] Service Function Chaining: Framework & Architecture / Boucadair M., Jacquenet C. // Internet Draft (Intended Status: Standards Track) – 2014. – Режим доступа к статье: <https://tools.ietf.org/search/draft-boucadair-sfc-framework-02>
9. Bouillard A., Jouhet L., Thierry E. Service curves in Network Calculus: dos and don'ts: tech. rep. – 2009. – URL: <https://hal.inria.fr/file/index/docid/431674/filename/RR-7094.pdf>.
10. Bouillard A., Junier A. Worst-case Delay Bounds with Fixed Priorities Using Network Calculus // Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools, (VALUETOOLS'11). – Paris, France: ICST (Institute for Computer Sciences, Social-Informatics, Telecommunications Engineering), 2011. – pp. 381–390.
11. Bouillard A., Steay G. Exact worst-case delay for FIFO-multiplexing tandems // Proceedings of the 6th International Conference on Evaluation Methodologies and Tools (VALUETOOLS'12). – Oct. 2012. – Pp. 158–167. – URL: <http://www.di.ens.fr/~bouillard/Publis/Valuetools-fifo.pdf>.
12. Cai Z. Maestro: Balancing Fairness, Latency and Throughput in the OpenFlow Control Plane / Cai Z, Cox A.L, Eugene NgT.S. // Technical Report, Rice University: Houston, TX, USA–2011 – 78 p.
13. Chemeritskiy E. On QoS management in NFV by multipath routing / Chemeritskiy E., Smelansky R. // Proceedings 2014 international science and technologic conference «Modern Networking Technologies (MoNeTec)». Moscow, Russia, June 27–29, 2014. – P. 41–46.
14. Chemeritsky E., Smeliansky R., Zakharov V. A formal model and verification problems for Software Defined Networks // Automatic Control and Computer Sciences. – 2014. – Vol. 48, no. 7.
15. Chen P. The entity relationship model – Towards a unified view of data / P. Chen // ACM Trans. On Database System. – 1976. – №1(1). – P. 9-45.
16. Chung L., Prado Leite J. do On Non-Functional Requirements in Software Engineering // Conceptual Modeling: Foundations and Applications,

Lecture Notes in Computer Science. Vol. 5600. – Springer Berlin Heidelberg, 2009. – Pp. 363–379.

17. Cisco corporation. Cisco Enterprise Network Functions Virtualization [Электронный ресурс] // Сайт Cisco. – 2015 Режим доступа: <http://www.cisco.com/c/en/us/solutions/collateral/enterprise-networks/enterprise-network-functions-virtualization-nfv/white-paper-c11-736783.pdf>

18. Cisco IOS NetFlow [Электронный ресурс] // Сайт Cisco. – 2016 – Режим доступа: <http://www.cisco.com/c/en/us/products/ios-nx-os-software/ios-netflow/index.html>.

19. Ciucu F., Schmitt J. Perspectives on Network Calculus: No Free Lunch, but Still Good Value // ACM SIGCOMM Computer Communication Review (SIGCOMM'12). –2012. – Aug. – Vol. 42, no. 4. – Pp. 311–322.

20. Cloud Sourcing NFV VoIP and IMS. Hosted Model Delivers Superior Agility and Business Model Today [Электронный ресурс]/ Alianza Company. – 2015. Режим доступа: [http://www.alianza.com/hubfs/Collateral/Cloud\\_Sourcing\\_NFV\\_VoIP\\_Alianza.pdf?t=1473364112924](http://www.alianza.com/hubfs/Collateral/Cloud_Sourcing_NFV_VoIP_Alianza.pdf?t=1473364112924)

21. Costas N. OpenFlow and NFV Technical Report / Atanasova I. K., Costas N. // Technical Report CESGA–2014–001 – 2014. – 64 p.

22. Cruz R. A Calculus for network delay. II. Network analysis // IEEE Transactions on Information Theory. – 1991. – Jan. – Vol. 37, no. 1. – Pp. 132–141.

23. Curtis A. Devoflow: scaling flow management for high-performance networks / A. Curtis, J. Mogul, J. Tourrilhes, P. Yalagan-Dula, P. Sharma // In Proceedings of the ACM SIGCOMM, 2011. –vol. 1. –p. 127-133.

24. David A. Chappell Enterprise Service Bus / A. David – USA: O'Reilly & Associates, Inc. – 2004. – 27 p.

25. Effect of Weighting Scheme to QoS Properties in Web Service Discovery/ Agushaka J. O., Lawal M. M., Bagiwa, A. M. and Abdullahi B. F // International Journal of Computer Science and Information Security Vol. 7 No. 3 March 2010. – pp.92–100.

26. Egawa T. NFV standardization Landscape from ITU–T Study Group /13  
Egawa T. // ITU Workshop on NFV. – Geneva, Switzerland, 4 June 2013. – 22 p.

27. Enns R. NETCONF configuration protocol [Электронный ресурс] / R. Enns // Internet Draft, Internet Engineering Task Force. – December 2004. – Режим доступа: <http://tools.ietf.org/html/rfc4741>

28. ETSI Industry Specification Group (ISG) NFV, “ETSI GS NFV–MAN 001 V1.1.1: Network Functions Virtualisation (NFV); Management and Orchestration,” [http://www.etsi.org/deliver/etsi\\_gs/NFV–MAN/001\\_099/001/01.01.01\\_60/gs\\_NFV–MAN001v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/NFV–MAN/001_099/001/01.01.01_60/gs_NFV–MAN001v010101p.pdf), December 2014.

29. ETSI Industry Specification Group (ISG) NFV, “ETSI GS NFV–PER 001 V1.1.1: Network Functions Virtualisation (NFV); NFV Performance & Portability Best Practises,” [http://www.etsi.org/deliver/etsi\\_gs/NFV–PER/001\\_099/001/01.01.01\\_60/gs\\_nfv–per001v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/NFV–PER/001_099/001/01.01.01_60/gs_nfv–per001v010101p.pdf), June 2014.

30. ETSI Industry Specification Group (ISG) NFV, “ETSI GS NFV–SEC 001 V1.1.1: Network Functions Virtualisation (NFV); NFV Security; Problem Statement,” [http://www.etsi.org/deliver/etsi\\_gs/NFV–SEC/001\\_099/001/01.01.01\\_60/gs\\_NFV–SEC001v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/NFV–SEC/001_099/001/01.01.01_60/gs_NFV–SEC001v010101p.pdf), October 2014.

31. ETSI Industry Specification Group (ISG) NFV, ETSI GS NFV 002 V1.2.1: Network Functions Virtualisation (NFV); Architectural Framework [online]. Available at: [http://www.etsi.org/deliver/etsi\\_gs/NFV/001\\_099/002/01.02.01\\_60/gs\\_NFV002v010201p.pdf](http://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf), December 2014.

32. ETSI, European Telecommunications Standards Institute, Industry Specification Groups (ISG) – NFV [Электронный ресурс] –2015. –Режим доступа: <http://www.etsi.org/technologies–clusters/technologies/nfv>.

33. Feamster N. The Road to NFV: An Intellectual History of Programmable Networks / Feamster N., Rexford J., Zegura E. // ACM Queue.– 2013. – P. 20-40.

34. Franken L. Quality of Service Management: A Model–Based Approach / Leonard Franken. – Centre for Telematics and Information Technology, 1996. – 267p.

35. Gardey G., Lime D., Magnin M., Roux O. Romeo: A Tool for Analyzing Time Petri Nets // Computer Aided Verification, Lecture Notes in Computer Science. Vol. 3576. –Springer Berlin Heidelberg, 2005. – pp. 418–423.

36. Guoping Z. A QoS-Based Web Services Selection Method for Dynamic Web Service Composition / Guoping, Z. Huijuan, Z. & Zhibin, W. // Proc. 2009 First International Workshop on Education Technology and Computer Science, Hubei, IEEE Computer Society, – pp. 832-835.

37. Haerder T. Principles of Transaction-Oriented Database Recovery / T. Haerder, A. Reuter. // ACM Computing Surveys. – 15(4). – 1983.– pp. 287–317.

38. Handigol N. Reproducible network experiments using container-based emulation / N.Handigol, B.Heller, V.Jeyakumar // In Proceedings of the 8th International Conference on emerging Networking Experiments and Technologies – CoNEXT'12, December 10-13, 2012, Nice, France. – pp. 253- 264.

39. Hassan Mohamed Muhi-Aldeen. A network load balancing algorithm for overlay-based NFV solutions / Hassan Mohamed Muhi-Aldeen, Abdulghafoor Raed Yahya, O.B. Tkachova // Second International IEEE Conference «Problems of Infocommunications. Science and Technology» PICS&T–2016, 4–7 October 2016: proc. of the conf. – Kharkiv, Ukraine, 2016. – Pp.1-4.

40. Hassan Mohamed Muhi-Aldeen. A method of modeling a behavior of NFV infrastructure based on network calculus theory / Hassan Mohamed Muhi-Aldeen // Scholars Journal of Engineering and Technology (SJET) ISSN 2321–435X (Online) Sch. J. Eng. Tech. – Vol 4(11). – 2017. – Pp. 517-524.

41. Hassan Mohamed Muhi-Aldeen. A method of network resource allocation in the networks supporting NFV technology / Hassan Mohamed Muhi-Aldeen, O.B.Tkachova // Науковий журнал "Комп'ютерно-інтегровані технології: освіта, наука, виробництво" Луцьк, 2018. Випуск № 33 – С.5-11

42. Hassan Mohamed Muhi-Aldeen. An evaluation of load distribution between servers' platforms inside a data center / Hassan Mohamed Muhi-Aldeen, Abdulghafoor Raed Yahya, O.B. Tkachova // 14th International Conference The

Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 21–25 February, 2017 – P. 357–359.

43. Henzinger T. A., Nicollin X., Sifakis J., Yovine S. Symbolic Model Checking for Realtime Systems // *Inf. Comput.* – 1994. – June. – Vol. 111, no. 2. – pp. 193–244.

44. Henzinger T., Ho P.–H., Wong–Toi H. HyTech: A model checker for hybrid systems // *Computer Aided Verification, Lecture Notes in Computer Science*. Vol. 1254. –Springer Berlin Heidelberg, 1997. – Pp. 460-463.

45. Hoare C. A. R. Communicating Sequential Processes / C. A. Hoare // *Commun. ACM.* – 1978. –Vol. 21, no. 8. – P. 666-677.

46. Hoffman L. Talking Model–Checking Technology / L. Hoffman // *Communications of the ACM.* – 2008. – V. 51. 77. – P. 110-112.

47. Holzmann G. J. The Model Checker SPIN // *IEEE Trans. Softw. Eng.* – 1997. –May. – Vol. 23, no. 5. – Pp. 279–295.

48. HP Network Functions Virtualization (NFV) [Электронныйресурс] // Сайт HP. –2015. –Режимдоступу: <http://www8.hp.com/ru/ru/cloud/nfv-overview.html>.

49. IBM Systems and Technology. IBM Software Defined Network for Virtual Environments[Электронныйресурс]. – Нью–Йорк, 2013.– 6 с.– Режим доступу:

<https://researchweb.watson.ibm.com/haifa/dept/stt/papers/QCW03028USEN.PDF>

50. IEEE Std 1348–1995. IEEE Recommended Practice for the Adoption of Computer-Aided Software Engineering (CASE) Tools.

51. ITU-T J.241. Quality of service ranking and measurement methods for digital video services delivered over broadband IP networks, 2005. – 47 p.

52.ITU-T Recommendation H.248. Series H: Audiovisual and Multimedia Systems: Infrastructure of audiovisual services – Communication procedures: Gateway control protocol [Электронныйресурс]. – Geneva: International Telecommunications Union. – 2000. – Режим доступу: <http://www.itu.int/ITU-T/recommendations/index.aspx>

53. ITU-T Recommendation Y.2001 Series Y: Global information infrastructure, internet protocol aspects and next-generation networks – General overview of NGN. Geneva: International Telecommunications Union. – 2005.

54. ITU-T Y.1541. Network Performance objectives for IP-based services, Amendment 3, 2008. – 50 p.

55. ITU-T Recommendation E.802. Series E: Overall network operation, telephone service, service operations and human factors – Framework and methodologies for the determination and application of QoS parameters. Geneva: International Telecommunications Union. – 2007.

56. ITU-T Recommendation G.1000. Series G: Transmission systems and media, digital systems and networks. – Communications quality of service: A framework and definitions. Geneva: International Telecommunications Union. – 2002.

57. Kang N. Policy Transformation in Software Defined Networks / Kang N., Reich J., Rexford J., Walker D. // Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, (SIGCOMM '12), Helsinki, Finland, 2012. – P. 309-310.

58. Khurshid A., Zhou W., Caesar M., Godfrey P. B. VeriFlow: Verifying Network-wide Invariants in Real Time // Proceedings of the First Workshop on Hot Topics in Software Defined Networks, (HotNFV '12). – Helsinki, Finland, 2012. – 3p. 49–54.

59. Kim H., Feamster N. Improving network management with software defined networking / H. Kim, N. Feamster // Communications Magazine, IEEE, 2013. – pp. 114-119.

60. Lafuente A. Directed Search for the Verification of communication protocols / A. Lafuente // Doctorial thesis, Institute of computer science, University of Freiburg. – 2003. – 157 p.

61. Lawrence Shari. Software Engineering: Theory and Practice (2nd ed.) / Shari Lawrence. – Upper Saddle River, NJ: Prentice Hall, 2001. – 630 pages.

62. Le Boudec J.-Y., Thiran P. Network Calculus: A Theory of Deterministic Queuing Systems for the Internet. – Springer-Verlag, 2001. – URL: [http://icalwww.epfl.ch/PS\\_files/netCalBookv4.pdf](http://icalwww.epfl.ch/PS_files/netCalBookv4.pdf).

63. Mattos D. OMNI: OpenFlow MaNagement Infrastructure / D. Mattos // In Proceedings of the International Conference on the Network of the Future (NOF), – Paris, France, 28–30 November 2011. – P. 52–56.

64. Milner R. Algebras for communicating systems // Proceedings of AFCET/SMF Joint Colloquim in Applied Mathamatics. – 1978.

65. Mininet: An Instant Virtual Network on your Laptop (or other PC) [Электронныйресурс] – Режим доступа: <http://mininet.org/>

66. Mininet: Introduction to Mininet [On-line] – Available at: <https://github.com/mininet/mininet/wiki/Introduction-to-Mininet>

67. Misra J. Proofs of networks of process / Misra J., Chandy K. // IEEE Transactions on Software Engineering SE-7. – Vol(4). – 1981. – P.417–426.

68. Murata T. Petri Nets: Properties, Analysis and Applications / T. Murata // Proceedings of the IEEE. – April 1989, vol. 77. –P. 541 – 580.

69. Murata T. Petri Nets: Properties, Analysis and Applications // Proceedings of the IEEE, April 1989, vol. 77, 541 – 580.

70. Nutt G. Evaluation Nets for Computer Systems Performance Analysis / G. Nutt // FJCC, AFIPSPRESS. – 1972. – P. 279 – 286.

71. Palmer C. Generating network topologies that obey power laws / C.Palmer, J.Steffan // Proceedings of the Global Telecommunications Conference, (IEEE GLOBECOM '00). – Vol (1). – 2000. – Pp. 434–438.

72. Pana F., Put F. A Survey on the Evolution of RSVP // Communications Surveys & Tutorials. – 2013. – Vol. 15, no. 4. – Pp. 1859–1887.

73. PRISM. Probabilistic Symbol Model Checker [Электронныйресурс]. – Режим доступа: <http://www.prismmodelchecker.org/>.

74. Process Algebras. Transactions on Petri Nets and Other Models of Concurrency // Lecture Notes in Computer Science. Vol. 5100. – 2008. – Pp. 54–70.



75. Qadir J., Hasan O. Applying Formal Methods to Networking: Theory, Techniques, and Applications // IEEE Communications Surveys and Tutorials. – 2015. – Т. 17, № 1. – С. 256-291.

76. QoS-aware Middleware for Web Services Composition/ Zeng L., Benatallah B., et al // IEEE Transactions on Software Engineering, Vol. 30, No. 5, 2004, pp. 311–327.

77. Quintero D. IBM Software Defined Environment / Quintero D., Genovese W, Waon K. // IBM Redbook, IBM Corp. – 2015-820 с.

78. Recourse ONF. OpenFlow specification. Series [Электронный ресурс] // Open Network Foundation – Режим доступа: <https://www.opennetworking.org/images/stories/downloads/NFV-resources/onf-specifications/openflow/openflow-switch-v1.5.0.noipr.pdf>

79. Ros F. J. Five nines of southbound reliability in software-defined networks / F. J. Ros, P. M. Ruiz // In Proceedings of the Third Workshop on Hot Topics in Software Defined Networking, ser. HotNFV '14. – New York, NY, USA: ACM – 2014. – P. 31–36.

80. Rosenberg J. SIP: Session Initiation Protocol. RFC 3261 / J. Rosenberg, H. Schulzrinne, U. Columbia, G. Camarillo. – East Hanover, June 2002. – 269 p.

81. NFV–NFV Reference Architecture v1.0. Verizon Network Infrastructure Planning [Электронный ресурс] // Verizon – 2016 – 220p.– Режим доступа: [http://innovation.verizon.com/content/dam/vic/PDF/Verizon\\_NFV-NFV\\_Reference\\_Architecture.pdf](http://innovation.verizon.com/content/dam/vic/PDF/Verizon_NFV-NFV_Reference_Architecture.pdf)

82. Sherwood R. FlowVisor: A Network Virtualization Layer [Электронный ресурс] / Sherwood R., Gibb G , Yap K. K. , Appenzeller G., Casado M., McKeown N. // Open Networking Foundation.–CA, USA, 2009. – 18 p. – Режим доступа: <http://archive.openflow.org/downloads/technicalreports/openflow-tr-2009-1-flowvisor.pdf>.

83. Single Root I/O Virtualization (SR–IOV) [Электронный ресурс] // Сайт VMware, Inc. – 2015. – Режим доступа: <http://pubs.vmware.com/vsphere->

51/index.jsp?topic=%2Fcom.vmware.vsphere.networking.doc%2FGUID-CC021803-30EA-444D-BCBE-618E0D836B9F.html

84. Stantchev V. Effects of Replication on Web Service Performance in WebSphere / Vladimir Stantchev // ICSI Technical Report TR-08-003.- 2008

85. Stirling C. P. Modal and temporal logics / C. P. Stirling // Handbook of Logic in Computer Science. –Oxford University Press, 1992. – Vol. 2. – P. 477–563.

86. Stirling C. P. Modal and temporal logics for processes / C. P. Stirling // LNCS 1043. – 1996. –P. 149–237.

87. Szigeti T. Quality of Service Design Overview / T. Szigeti, C. Hattingh. – Cisco Press, 2004. – 145 p.

88. Tarjan R. Depth first search and linear graph algorithms / Tarjan R. // SIAM Journal on Computing. –Vol. 1. – № 2. –1972. – P.146–160.

89. The Infonetics Global Service Provider Survey. NFV and NFV Strategies: Global Service Provider Survey [Электронныйресурс]. –2014. – Режим доступа: <http://alu.us.neolane.net/res/img/286758382c7e061c52883e873cee02e6.pdf>

90. The Internet Engineering Task Force (IETF) Service Function Chaining (SFC) Working Group (WG), <https://datatracker.ietf.org/wg/sfc/charter/>, 2015, Accessed: February, 06 2015.

91. Tkachova O. An Method of Service Composition with Optimal Resources Allocation in Software-Defined Networking / O.Tkachova, Ie. Duravkin, Muhi-Aldeen Hassan Mohamed // Forth International IEEE Conference «Problems of Infocommunications. Science and Technology» PICS&T-2017, 10–13 October 2017 proc. of the conf. – Kharkiv, Ukraine, 2017. – Pp.1–4.

92. Tsai W. T. A software reliability model for Web services/ W. T. Tsai , D. Zhang , Y. Chen , H. Huang , R. Paul , N. Liao // The 8th IASTED International Conference on Software Engineering and Applications, 2004.

93. Vegesna S. IP Quality of Service / Srinivas Vegesna. – USA, 2003. – 335p.

94. Vida Ahmady, Anders Carlsson. ReSeLa: Remote Security laboratory. [Электронныйресурс] // TEMPUS project ENGENSEC 544455-TEMPUS-1-2013-

1-SE-TEMPUS-JPCR. – 2016. – Режим доступу: <http://engensec.eu/2015/04/19/report-on-remote-security-laboratory-cloud-system/>

95. Web Services Quality Factors Version 1.0. [Електронний ресурс] / Committee Specification OASIS. – 2011. – Режим доступу: <http://docs.oasis-open.org/ws-qm/WS-Quality-Factors/v1.0/cs01/WS-Quality-Factors-v1.0-cs01.html>.

96. Wilkinson N. Next Generation Network Services: Technologies & Strategies / Neill Wilkinson. – Wiley, 2002. – 210 p.

97. Yang H., Lam S. Real-Time Verification of Network Properties Using Atomic Predicates // IEEE/ACM Transactions on Networking. – 2015. – Vol. PP, no. 99. – Pp. 1-14.

98. Yeganeh S. On scalability of Software-Defined Networking / Yeganeh S., Tootoonchian A., Ganjali Y. // Communications Magazine, IEEE, vol. 51, no. 2, 2015. – P. 136-141.

99. Yeh R. T. Requirements Analysis – A Management Perspective // Proceedings of COMPSAC' 82. – Nov. 1982. – Pp. 410-416.

100. Yu T. A broker-based framework for QoS-aware web service composition in e-Technology, e-Commerce and e-Service / T. Yu and K. J. Lin //, 2005. IEEE'05. Proceedings. the 2005 IEEE International Conference on, pp. 22–29.

101. Yu T. Service Selection Algorithms for Composing Complex Services with Multiple QoS Constraints / Yu T., Lin K.J // International Conference on Service-Oriented Computing, 2005, pp. 130–143.

102. Zheng Wang. Internet QoS: architectures and mechanisms for Quality of Service / Zheng Wang. – USA: Lucent Technologies. – 2001. – 240p.

103. Алиев Т.И. Основы моделирования дискретных систем / Алиев Т.И. – СПб: СПбГУ ИТМО, 2009. – 363 с.

104. Багатоканальний електрозв'язок та телекомунікаційні технології: підручник у 2-х частинах. Ч.1 / О.В. Лемешко, В.А. Лошаков, В.В. Поповський та ін. за заг. ред. проф. Поповського В.В. – Х.: ТОВ "Компанія СМІТ", 2010. – 470 с.

105. Брукс П. Метрики для управления ИТ-услугами / П. Брукс. – Москва, 2008. – 270 с.

106. Буханько А.Н. Алгоритмы управления каналами связи интеллектуального агента участка сети / А.Н. Буханько, В.М. Безрук, Е.В. Дуравкин // Вісник національного університету «Львівська політехніка». – 2009. – № 645. – С. 68 – 72.

107. Гладкий А. В. Формальные грамматики и языки / А. В. Гладкий. – М.: Наука, 1973. – 368 с.

108. Гольдштейн Б.С. Проблемы перехода к мультисервисным сетям [Текст] / Б.С. Гольдштейн // Вестник связи. – 2002. – № 12. – С. 26–31.

109. Грекул В. И Проектирование информационных систем / В. И. Грекул, Г. Н. Денищенко, Н. Л. Коровкина. – М.: Интернет-университет информационных технологий, 2005. – 304 с.

110. Гросс М. Теория формальных грамматик / М. Гросс, А. Лантен. – М.: Мир, 1971. – 294 с.

111. Дмитриева Е.А. Система E-сетевого имитационного моделирования EVA / Е.А. Дмитриева. – Томск, Издательство ТПУ, 1994. – 31 с.

112. Карпов Ю.Г. Model Checking. Верификация параллельных и распределенных программных систем / Ю.Г. Карпов. – СПб.: БХВ-Петербург, 2010. – 552 с.

113. Коберн А. Современные методы описания функциональных требований к системам. / А. Коберн.: Пер. с англ. – М.: Издательство ЛОРИ, 2002. – 266 с.

114. Конев Б. Ю. Введение в моделирование и верификацию аппаратных и программных систем [Электронный ресурс] // Computer Science клуб при ПОМИ РАН. – Режим доступа.: <http://logic.pdmi.ras.ru/~infclub/?q=courses/verification>.

115. Контроль показателей качества обслуживания с учетом перехода к сети связи следующего поколения / Гольдштейн Б.С., Маршак М.Д., Мишин Е.Д. и др. / Техника связи. – 2009. – №1. – С.14–21.

116. Котов В.Е. Сети Петри / Котов В.Е. – М.: Наука, 1984. – 160с.
117. Кривый С.Л. Формальные методы анализа свойств систем / С.Л. Кривый, Л.Е. Матвеева // Кибернетика и системный анализ. – №2. – 2003. – С.15–36.
118. Кульгин М.В. Технология корпоративных сетей. /М.В. Кульгин. – СПб: Питер. – 2000. – 704с.
119. Кулямин В.В. Методы верификации программного обеспечения [Электронная версия] / В.В. Кулямин // Всероссийский конкурсный отбор обзорно–аналитических статей по приоритетному направлению "Информационно–телекоммуникационные системы". – 2008. – 117 с. Режим доступа: <http://www.viva64.com/go.php?url=282>
120. Лебедев А. Удаленная репликация: критерии выбора [Электронный ресурс]/ Александр Лебедев// StorageNews.–№ 2 (27).– 2006. Режим доступа: [http://storagenews.ru/27/NvisionGr-Remote\\_PKB.pdf](http://storagenews.ru/27/NvisionGr-Remote_PKB.pdf).
121. Лосев Ю.И. Применение методов анализа E–сетей к моделям СОД / Ю.И. Лосев, С.И. Шматков, Е. В. Дуравкин // Радиотехника: Всеукр. межвед. науч.–техн. сб. – 2002. – Вып.132. – С. 149 – 151.
122. Мухи–Алдин Хассан Мохамед. Модель оценки показателей качества обслуживания в облачных решениях // Информатика, математика, автоматика. Матеріали науково–технічної конференції, 17–21 квітня 2017 р. – Суми: Сумський державний університет, 2017. – С. 104.
123. Мухи–Алдин Х.М. Комплексный метод повышения доступности сервисов в программно–конфигурируемых сетях, основанный на динамической репликации [Электронный ресурс] / Х.М. Мухи–Алдин, Е.Б. Ткачова, М.Т. Салах // Проблемы телекоммуникацій. – 2016. – № 2 (19). – С. 81 – 93. – Режим доступа к журн.: [http://pt.journal.kh.ua/2016/2/1/162\\_tkachova\\_rep.pdf](http://pt.journal.kh.ua/2016/2/1/162_tkachova_rep.pdf).
124. Мэри Т. Интеграция данных и композитные бизнес–сервисы: Часть 3. Создание уровня данных множественной принадлежности с контролем доступа и системой защиты // Мэри Тейлор, Чанг Джи Гуо [Электронная

версия]: IBM, 2010 – режим доступа: <http://www.ibm.com/developerworks/ru/library/dm-0712taylor/>.

125. Питерсон Дж. Теория сетей Петри и моделирование систем / Дж. Питерсон. – М.: Мир, 1984. – 264с.

126. Поповский В. В. Математические основы управления и адаптации в телекоммуникационных системах / В. В. Поповский, В. Ф. Олейник. – Х. : СМИТ, 2011. – 362 с. – ISBN 978-966-2028-72-0

127. Рекомендация МСЭ–Т Y.110 «Принципы и архитектура глобальной информационной инфраструктуры» (ITU–T Rec. Y.110 «Global Information Infrastructure principles and framework architecture»). – 1998. – 47 с.

128. Смелянский Р. Программно–конфигурируемые сети [Электронный ресурс] / Р.Смелянский // Открытые системы. – 2012. – № 9. – Режим доступа: [www.osp.ru/os/2012/09/13039421/](http://www.osp.ru/os/2012/09/13039421/).

129. Советов Б.Я. Моделирование систем / Б.Я. Советов, С.А. Яковлев. – М.: Высш. шк., 1985. – 271 с.

130. Таненбаум Э. Распределенные системы. Принципы и парадигмы. / Таненбаум Э., М. ван Стеен. – СПб.: Питер, 2003. – 877с.

131. Технологии NFV и NFV: новые возможности для телекоммуникаций [Электронный ресурс] // Центр прикладных исследований компьютерных сетей. – 2015. –Режим доступа: <http://arccn.ru/media/1132>.

132. Технология Brocade Virtual Cluster Switching [Электронный ресурс] // Brocade Communications Systems. –2014. –Режим доступа: <http://brocade.ocs.ru/-vcs>.

133. Ткачева Е.Б. Адаптивный алгоритм перераспределения сетевых ресурсов в сетях с поддержкой технологии NFV / Е.Б. Ткачева, Хассан Мохамед Мухи–Алдин// Радиотехника: Всеукр. межвед. науч.–техн. сб. – 2017. – Вып. 191 – С. 135–141.

134. Ткачева Е.Б. Применение теории сетевого исчисления при моделировании программно–конфигурируемых сетей / Е.Б. Ткачева, Е.В.

Дуравкин, Х.М. Мухи–Алдин // Системы управління, навігації та зв'язку. – 2017 – №6(46) – С. 136–141.

135.Ткачева Е.Б., Мухи–Алдин Х.М. Анализ сценариев оркестровки сервисов в облачных технологиях / Е.Б. Ткачева, Мухи–Алдин Х.М. // Інформаційна безпека та комп'ютерні технології: Збірник тез доповідей II Міжнародної науково–практичної конференції, 20–22 квітня 2017 року, м. Кропивницький: ЦНТУ, 2017. – С. 173–175.

136.Харитонов В.В. Качество обслуживания и эффективное использование ресурсов в мультисервисных сетях / В.В. Харитонов // Вестник связи. – 2004. – №12. – С. 48–55.

137.Хассан Мохамед Мухи-Алдин. Методологія побудови кореляційних алгоритмів класифікація мережевого трафіку / Хассан Мохамед Мухи-Алдин // Міжвузівський збірник "Наукові нотатки".– 2018, – Вып.64, – С. 128-135.

138.Хассан Мохамед Мухи-Алдин. Комплексный метод динамической репликации данных в программно-конфигурируемых сетях / Хассан Мохамед Мухи-Алдин, Е.Б. Ткачева // Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки» – 2018. – Том 29 (68). – 2018. – Вып. 5 – С. 18-26.

139.Царьков Д. Верификация распределённых программ методом проверки на модели: Дис. канд. физ.–мат. наук: 05.13.11 / Московский Государственный Университет им. М.В. Ломоносова. – 2002. – 185 с.

140.Шахгельдян К.И. Автоматическая репликация данных в корпоративной информационной среде / Шахгельдян К.И. // Научно–технические ведомости Санкт–Петербургского государственного политехнического университета.–2008.– №3. –С. 25 – 33.

## ДОДАТОК А

## Акти впровадження

ЗАТВЕРДЖУЮ  
 Директор Державного підприємства  
 «Науковий центр точного машинобудування»  
 д.т.н., с.н.с.

« 4 » квітня 2018 р.



О.В. СЕЛЮКОВ

## Акт № 013

## про впровадження результатів дисертаційної роботи

## МУХІ-АЛДІН ХАССАН МОХАМЕДА

Комісія у складі: голови комісії – заступника директора ДП «НЦТМ» з наукової роботи Якимов В.В та членів комісії: Іванілова А.О, Яценка Н.Г., встановила, що в Державному підприємстві «Науковий центр точного машинобудування» Державного космічного агентства України реалізовані результати наукових досліджень, які розроблені особисто Мухі-Алдіні Хассан Мохамедом, а саме подальший розвиток отримав метод балансування навантаження, який дозволяє на основі аналізу поточного навантаження як на віртуалізовані канали зв'язку, так й обчислювальні пристрої побудувати маршрут передачі даних, що забезпечить надання послуг з відповідною якістю обслуговування;

Визначено, що запропонований метод балансування навантаження дозволяє підвищити ефективність використання каналів передачі на 7% та зменшити середню затримку загалом на 16% у порівнянні з динамічним алгоритмом Central Queuing, а застосування розробленого методу реплікації сервісів дозволяє підвищити доступність сервісу в 14 разів у випадку перевантажень.

Запропонований автором метод балансування навантаження є підставою для теоретичної проробки та обґрунтування вимог до перспективних систем оцінки якості функціонування телекомунікаційних систем та мереж зв'язку на основі інтелектуальних технологій.

Голова комісії:

В.В. ЯКИМОВ

Члени комісії:

А.О. ІВАНІЛОВ

4 квітня 2018 р.

Н.Г. ЯЦЕНКО





МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БУДІВНИЦТВА І АРХІТЕКТУРИ**

Повітрофлотський пр. 31, м. Київ-37, 03037, тел. (044)241-55-80, факс (044) 248-32-65

E-mail: knuba\_admin@ukr.net, Web: http:// www.knuba.edu.ua, код ЄДРПОУ 02070909

11.12.19 № 132

На № \_\_\_\_\_ від \_\_\_\_\_

**Довідка**

щодо впровадження результатів дисертаційного дослідження здобувача  
**Мухі-Алдіна Хассана Мохамеда**  
 за темою “ Моделі та методи забезпечення якості обслуговування у мережах з  
 підтримкою функцій віртуалізації ”  
 Науковий керівник д.т.н., професор Хлапонін Ю.І.

Результати дисертаційного дослідження здобувача Мухі-Алдіна Хассана Мохамеда за темою “Моделі та методи забезпечення якості обслуговування у мережах з підтримкою функцій віртуалізації” на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.12.02 – телекомунікаційні системи та мережі впроваджені в діяльність Київського національного університету будівництва і архітектури.

Пропозиції, розроблені дисертантом використано у навчальному процесі Київського національного університету будівництва і архітектури, зокрема щодо можливості впровадження нових дисциплін “Технології віртуалізації” та “Хмарні технології”, які дозволили підвищити якість викладання дисциплін.

Також впроваджено в процес навчання студентів в діючій в КНУБА Академії Cisco методику управління мережевими ресурсами, що дозволяє підвищити рівень доступності комплексних сервісів.

Проректор з наукової роботи Київського  
 національного університету будівництва і архітектури:

**В.О. ПЛОСКИЙ**



ЗАТВЕРДЖУЮ  
 Директор ТОВ «СВІТ ІТ»  
  
 КОВАЛЬОВ  
 « 12 » березня 2019 р.

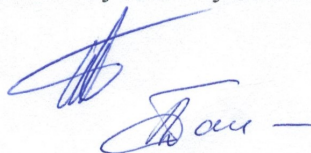
**Акт № 1203-1**  
**про впровадження результатів дисертаційної роботи**  
**МУХІ-АЛДІН ХАССАН МОХАМЕДА**

Комісія у складі: голови комісії – заступника директора ТОВ «СВІТ ІТ» Балаби О.В. та членів комісії: Бакалова Л.Д., Артеменка А.О., встановила, що в ТОВ «СВІТ ІТ» реалізовані результати наукових досліджень, які розроблені особисто Мухі-Алдін Хассан Мохамедом, а саме: вдосконалено існуючі та розроблено нові моделі і методи управління та розподілу мережевих ресурсів, що дозволяють підвищити рівень якості обслуговування у мережах з підтримкою технології NFV, зокрема, дозволяють забезпечити високий рівень доступності сервісів, що надаються кінцевому користувачу.

Визначено, що розроблений метод динамічного управління складом комплексних сервісів за рахунок зміни їх складових у NFV мережах дозволяє формувати та управляти складом комплексного сервісу з урахуванням поточного стану складових сервісів при формуванні комплексного сервісу, що призводить до забезпечення необхідного рівня QoS.

Запропоновані методи і моделі аналізу та оцінки якості надання послуг було використано під час розробки процесів пошуку та резервування системою MANO каналів зв'язку з достатньою пропускнуою здатністю, процесу резервування ресурсів за допомогою протоколу RSVP та формування агрегованого потоку даних уздовж віртуальних каналів зв'язку.

Голова комісії:  
 Члени комісії:



О.В. Балаба  
 Л.Д. Бакалов  
 А.О. Артеменко

