МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ
НАЦІОНАЛЬНИЙ ЦЕНТР «МАЛА АКАДЕМІЯ НАУК УКРАЇНИ»

НАУКОЄМНІ ТЕХНОЛОГІЇ ОПТИМІЗАЦІЇ ТА КЕРУВАННЯ В ІНФОКОМУНІКАЦІЙНИХ МЕРЕЖАХ

Колективна монографія

Під загальною редакцією
В.М. Безрука, Л.С. Глоби, О.Є. Стрижака

Київ

2019

УДК 621.391+004.7
Н34

Рекомендовано до друку Вченою радою Національного центру «Мала академія наук України» (протокол № 1 від 15 січня 2019року), науково-технічною радою Харківського національного університету радіоелектроніки (протокол № 2 від 25.06.2018 р.)

Рецензенти:

Бараннік В.В., д.т.н., проф., начальник кафедри бойового застосування та експлуатації АСУ Харківського Національного університету Повітряних Сил імені Івана Кожедуба; Климаш М.М., д.т.н., проф., завідувач кафедри телекомунікацій Національного університету «Львівська політехніка».

Н34 **Наукоємні технології оптимізації та керування в інфокомунікаційних мережах** : монографія / Під загальною редакцією В.М. Безрука, Л.С. Глоби, О.Є Стрижака. – К.: Інститут обдарованої дитини НАПН України, 2019. – 194 с.
ISBN 978-617-7734-02-3

В монографії викладені сучасні результати досліджень ряду авторів з наукоємних технологій оптимізації та управління в інфокомунікаційних мережах. Розглядаються особливості методів багатокритеріальної оптимізації систем з урахуванням сукупності показників якості, а також особливості та результати вирішення задач управління при передачі та обробці інформації для різних типів інфокомунікаційних мереж.

УДК 621.391+004.7

© Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», 2019
© Харківський національний університет радіоелектроніки, 2019
© Національний центр «Мала академія наук України», 2019
© Інститут обдарованої дитини НАПН України, 2019
ISBN 978-617-7734-02-3

ЗМІСТ

	Вступ	4
1	МЕТОДИ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ МЕРЕЖ ЗВ’ЯЗКУ *Безрук В.М.*	7
2	БАГАТОКРИТЕРІАЛЬНИЙ ВИБІР ПРОЕКТНИХ ВАРІАНТІВ ПРИ ПЛАНУВАННІ СИСТЕМ МОБІЛЬНОГО ЗВ’ЯЗКУ *Безрук В.М., Чеботарьова Д.В.*	27
3	ВИБІР ПЕРЕВАЖНИХ ВАРІАНТІВ ЗАСОБІВ ЗВ'ЯЗКУ НА ОСНОВІ МЕТОДУ АНАЛІЗУ ІЄРАРХІЙ *Безрук В.М., Власова В.О., Скорик Ю.В.*	47
4	БАГАТОКРИТЕРІАЛЬНИЙ ПІДХІД ДО ОПТИМАЛЬНОЇ МАРШРУТИЗАЦІЇ У МЕРЕЖАХ ЗВ’ЯЗКУ *Безрук В.М., Гальченко К.Р.*	83
5	КЕРУВАННЯ РЕСУРСАМИ ДЛЯ ВІРТУАЛІЗОВАНИХ МЕРЕЖЕВИХ ФУНКЦІЙ *Скулиш М.А., Суліма С.В.*	97
6	ОПТИМІЗАЦІЯ ОБЧИСЛЕНЬ У ДАТА–ЦЕНТРІ ЗА КРИТЕРІЄМ ЕНЕРГОЕФЕКТИВНОСТІ *Гвоздецька Н.А., Глоба Л.С., Прокопець В.А., Степурін О.В.*	127
7	ДИНАМІЧНИЙ РОЗПОДІЛ РЕСУРСІВ У ВІРТУАЛІЗОВАНІЙ МЕРЕЖІ МОБІЛЬНОГО ОПЕРАТОРА УКРАЇНИ З ПІДВИЩЕННЯМ ЕНЕРГОЕФЕКТИВНОСТІ ОБРОБКИ ДАНИХ *Гвоздецька Н.А., Глоба Л.С., Прокопець В.А.*	151
8	МОДИФІКАЦІЯ МЕТОДУ ВЕРТИКАЛЬНОГО ХЕНДОВЕРУ ДЛЯ ГЕТЕРОГЕННОЇ МЕРЕЖІ *Глоба Л.С., Кірюшкін Р.О., Курдеча В.В.*	177

ВСТУП

Інфокомунікації – сучасна концепція конвергентної взаємодії телекомунікацій та інформаційних технологій з метою забезпечення своєчасної якісної передачі та обробки інформації.

Із впровадженням у світі технологій 4G LTE та розвитком тенденції переходу до технологій 5G постає питання збільшення швидкості та ефективності передачі та обробки даних. Все більшого поширення набувають сервіси Internet of Things (Інтернет речей) та сервіси міжмашинної взаємодії (M2M). Ці сервіси продукують великі масиви даних у режимі реального часу, що тягне за собою появу концепції BigData. Поряд із цим у світі гостро стоїть проблема економії електроенергії, що наразі споживається центрами обробки даних у величезних обсягах.

Такі виклики сучасного світу інфокомунікацій спричиняють необхідність пошуку нових наукоємних технологій оптимізації та управління при вирішенні задач планування та проектування інфокомунікаційних мереж. В монографії викладені деякі сучасні результати досліджень ряду авторів з питань оптимізації та управління в інфокомунікаційних мережах з урахуванням сукупності техніко-економічних вимог, включаючи швидкість передачі та обробки даних, енергоефективність, оптимальне використання ресурсів та інше. Це визначає актуальність опублікування даної колективної монографії, в якій викладені деякі результати досліджень авторів по вирішенню задач оптимізації та управління в інфокомунікаційних мережах з використанням сучасного математичного апарату та високоефективних обчислювальних засобів. Кожен розділ монографії являє собою завершене викладення результатів досліджень, отриманих авторами цього розділу.

Матеріали 1 розділу підготував Безрук В.М. У цьому розділі розглядаються деякі особливості методів багатокритеріальної оптимізації, що використовуються при розв’язанні різних типів задач оптимізації проектних рішень у мережах зв’язку з урахуванням сукупності показників якості. Наведено методи вирішення багатокритеріальних задач оптимізації при застосуванні безумовного критерію переваги, що приводить до отримання підмножини Парето-оптимальних рішень на множині допустимих варіантів. Для звуження підмножини Парето до єдиного переважного проектного варіанту застосовується умовний критерій переваги. Розглянуто різні методи формування умовного критерію переваги з урахуванням додаткової інформацію від експертів, що дають можливість звуження підмножини Парето до єдиного переважного рішення. У наступних розділах 2 – 4 наводяться деякі приклади, що ілюструють практичні особливості використання розглянутих методів при багатокритеріальному аналізі і виборі оптимальних проектних варіантів для різних засобів зв’язку.

Матеріали 2 розділу підготували Чеботарьова Д.В., Безрук В.М. У цьому розділі розглянуто особливості застосування методів багатокритеріальної

оптимізації на етапі номінального планування радіомережі системи стільникового мобільного зв’язку (СМЗ) 2-го та 3-го покоління. Вирішена задача оптимізації радіомережі СМЗ як задача дискретного вибору Парето-оптимальних проектних варіантів при врахуванні сукупності показників якості. Розглянуто також особливості застосування методів багатокритеріальної оптимизації при номінальному плануванні транспортної мережі СМЗ. Наведені результати вибору оптимальної топології транспортної мережі СМЗ з урахуванням сукупності показників якості.

Матеріали 3 розділу підготували Безрук В.М., Скорик Ю.В., Власова В.О. У даному розділі розглянуто особливості вибору переважного проектного варіанту з урахуванням сукупності показників якості при застосуванні умовного критерія переваги, що заснований на використанні методу аналізу ієрархій (МАІ). Цей метод полягає в отриманні суджень експертів по парним порівнянь різних елементів проблеми вибору. В результаті обробки цих даних за строго формалізованими процедурами обчислюється вектор глобальних пріоритетів, компоненти якого визначають пріоритетність вибору переважного варіанту системи. Досліджено практичні особливості застосування МАІ для вибору переважного варіанту на прикладі різних типів засобів зв'язку, зокрема, мовних кодеків для IP телефонії, системи масового обслуговування заявок у вузлах комутації мережі, проектних варіантів систем мобільного зв'язку 3-го покоління, технологій мобільного зв'язку 4-го покоління, типів мобільних телефонів, протоколів маршрутизації в ad-hoc мережах, модуляції в цифрових системах зв'язку, систем телевізійного мовлення стандартів DVB-T і ATSC, протоколів маршрутизації в бездротовій сенсорно-актуаторній мережі.

Матеріали 4 розділу підготували Безрук В.М., Гальченко К.Р. У даному розділі розглядаються особливості математичного формулювання та вирішення задач оптимальної маршрутизації у мережах зв’язку. Оскільки, практичні умови роботи мереж зв’язку вимагають врахування декількох показників якості (метрик) при виборі оптимальних маршрутів, тому розглянуто багатокритеріальний підхід до вирішення задач оптимальної маршрутизації у мережах зв’язку. Наведено метод формування підмножини Парето-оптимальних варіантів маршрутизації, що дають можливість організації багатошляхової маршрутизації у мережі зв’язку. Наводиться приклад вирішення задачі оптимальної маршрутизації у мультисервісній мережі зв’язку з урахуванням сукупності показників якості.

Матеріали 5 розділу підготували Скулиш М.А., Суліма С.В. У даному розділі запропоновано алгоритм оптимального розподілу ресурсів у віртуалізованих мережах. Метою алгоритму надання ресурсів є виділення достатньої їх кількості для мережевих функцій, так що їх SLA можна задовольнити навіть у присутності пікового навантаження. В основі будь-якого алгоритму надання ресурсів лежать два питання: скільки ресурсів надавати і коли. Ці питання глибоко досліджуються авторами розділу.

Матеріали 6 розділу підготували Глоба Л.С., Степурін О.В., Гвоздецька Н.А., Прокопець В.А. У даному розділі запропоновано підхід до підвищення енергоефективності обчислень для серверного кластера як інформаційно-телекомунікаційної одиниці інфраструктури ЦОД, який відрізняється одночасним врахуванням параметрів енергоефективності та продуктивності при розподілі задач. Розроблено алгоритм підвищення енергоефективності обчислень на основі запропонованого підходу, який складається з етапу попередньої індивідуальної атестації серверного кластера та етапу енергоефективного розподілу задач. Врахування індивідуальних залежностей енергоспоживання серверів від їх завантаженості є основною відмінною рисою запропонованого підходу.

Матеріали 7 розділу підготували Глоба Л.С., Прокопець В.А., Гвоздецька Н.А. У даному розділі проаналізовано ефективність методу динамічного розподілу ресурсів у віртуалізованій мережі мобільного оператора України з використанням механізмів підвищення енергоефективності обробки даних. Основною метою методу динамічного розподілу ресурсів є підтримка обсягу ресурсів, достатнього для виконання вимог SLA (Service Level Agreement – Угода про рівень обслуговування) при уникненні надмірності. Це дозволяє більш ефективно використовувати апаратні ресурси для економії електроенергії та зменшення викидів CO_2. Ефективність методу перевірено для віртуалізованого ядра EPC (Evolved Packet Core) мережі мобільного зв'язку України. Для додаткового підвищення енергоефективності обробки даних у дослідженні застосовано підхід PCPB (Power Consumption and Performance Balance), який запропоновано у розділі 6.

Матеріали 8 розділу підготували Кірюшкін Р.О., Курдеча В.В., Глоба Л.С. У даному розділі проведено огляд основних методів вертикального хендовера та зроблений висновок про неефективність існуючих рішень у бездротових гетерогенних мережах з багатокласовими груповими викликами мобільних вузлів. Описано модифікацію методу вертикального хендовера за рахунок зміни алгоритму прийняття рішень для різних типів викликів мобільного вузла що дозволило підвищити ефективність зв’язку у гетерогенних безпроводових мережах. Розглянуто створену математичну модель алгоритму прийняття рішення, який застосовує методи TOPSIS і MULTIMOORA. Проведено імітаційне моделювання методу прийняття рішення, який базується на запропонованій математичній моделі та доведено його ефективність у гетерогенних бездротових мережах.

7. ДИНАМІЧНИЙ РОЗПОДІЛ РЕСУРСІВ У ВІРТУАЛІЗОВАНІЙ МЕРЕЖІ МОБІЛЬНОГО ОПЕРАТОРА УКРАЇНИ З ПІДВИЩЕННЯМ ЕНЕРГОЕФЕКТИВНОСТІ ОБРОБКИ ДАНИХ

Гвоздецька Н.А., Глоба Л.С., Прокопець В.А.

З огляду на значне зростання мобільного трафіку даних, з одного боку, а також кількості та складності послуг, з іншого боку, постачальники послуг, зокрема оператори мобільного зв'язку, все більше віртуалізують частини своєї мережі за допомогою технологій NFV, хмарних обчислень для побудови економічно ефективних та еластичних мобільних мереж та їх застосування у якості хмарних сервісів.

Динамічно змінювана архітектура мережі вимагає застосування контролю показників функціонування мережі в так званому «end-to-end» середовищі. Технології віртуалізації та динамічної реконфігурації мережі дозволяють знизити витрати, підвищити ефективність, мають значний вплив на забезпечення продуктивності всієї мережі та додають функціональність в її програмне забезпечення.

Віртуалізація мобільних мереж в Україні має значні перспективи розвитку. У даному дослідженні проаналізовано ефективність методу динамічного розподілу ресурсів у віртуалізованій мережі мобільного оператора України з використанням механізмів підвищення енергоефективності обробки даних. Основною метою методу динамічного розподілу ресурсів є підтримка обсягу ресурсів, якого достатньо для виконання вимог SLA (Service Level Agreement – Угода про рівень обслуговування) при уникненні надмірності. Це дозволяє більш ефективно використовувати апаратні ресурси для економії електроенергії та зменшення викидів CO_2. Ефективність методу перевірено для віртуалізованого ядра EPC (Evolved Packet Core) мережі мобільного зв'язку України. Для додаткового підвищення енергоефективності обробки даних у дослідженні застосовано підхід PCPB (Power Consumption and Performance Balance) [12]. Дослідження проведено з використанням імітаційного моделювання у середовищі Matlab. Результати моделювання довели ефективність методу динамічного розподілу ресурсів у комбінації із підходом PCPB та показали, що економія ресурсів при цьому зростає на 41,78% при дотриманні вимог SLA.

7.1. Аналіз проблеми визначення оптимального обсягу ресурсів у віртуалізованих системах

Сучасні телекомунікаційні системи складаються з великої кількості вузькоспеціалізованого фірмового обладнання, яке, найчастіше, може виконувати невеликий обсяг конкретних операцій, наприклад, забезпечення NAT (Network Address Translation), функцій обмеження швидкості доступу, здійснення «батьківського» контролю та фільтрацію вмісту, firewall, тощо. Такий розподіл функцій мережевої апаратури спричинений декількома факторами:

1. Кожен постачальник мережевого обладнання спеціалізується на деяких функціях, обмежених визначеною ділянкою мережевої топології

(наприклад, виступає виробником обладнання для мережі доступу, або ядра мережі). Зазвичай, більшість вендорів не має задовільних технічних рішень для інших частин мережевої топології.

2. Практично будь-яке вузькоспеціалізоване обладнання, задля якомога кращого виконання своїх функцій має складну і специфічну структуру, яка не достатньо уніфікована з іншим обладнанням, навіть того ж самого виробника.

Ці фактори спричинюють необхідність підтримки гетерогенних мультивендорних мереж операторами зв'язку, які експлуатують подібні мережі. При цьому, запуск будь-якої нової послуги вимагає значних капітальних затрат (CAPEX – Capital Expenditures) для закупки нового комплекту вузькоспеціалізованого обладнання, вартість якого, як правило, значно перевищує вартість серверів загального призначення. Крім того, додатково необхідно придбати високовартісне програмне забезпечення (ПЗ) та вкладати кошти в переналаштування існуючих мережевих елементів у відповідності до нової топології. Разом з тим потрібно збільшувати операційні витрати (OPEX – Operational Expenditures) у вигляді плати за оренду додаткових площ для обладнання, додаткове енергоживлення, логістику, монтаж та пусконалагодження.

Існуюча модель розвитку, яка передбачає, що мережа складається з вузькоспеціалізованих пристроїв, та в якій впровадження нових послуг неможливе без введення в експлуатацію нового обладнання, не є оптимальною. Високі капітальні та операційні витрати спричинені затратним за часом проектуванням архітектури мережі. Оператор в майбутньому не матиме можливості вводити в експлуатацію нові послуги так швидко, як того вимагатиме ринок. Мережа має бути динамічною, сприяти як впровадженню нових сервісів та послуг, так і швидкій деактивації неактуальних послуг без втрат як для оператора, так і для користувача.

Функції, які виконують мережеві пристрої, значно різняться своїми характеристиками і можливостями їх віртуалізації. Частина з них, наприклад, переміщення мережевих пакетів з одної точки в іншу, застосовує фізичні мережеві пристрої – комутатори або маршрутизатори, їх неможливо відокремити від фізичної апаратури. Деякі інші функції, з погляду на сучасний розвиток інформаційних та комунікаційних технологій, навіть доцільно перенести в хмару.

Віртуалізація мережевих функцій (англ. Network Functions Virtualization, NFV) – це концепція мережевої архітектури, що пропонує використовувати технології віртуалізації для цілих класів функцій мережевих вузлів у вигляді складових елементів, які можуть бути з'єднані разом або пов'язані в ланку для створення телекомунікаційних послуг (сервісів).

Будь-яка віртуалізована мережева функція складається з однієї або декількох віртуальних машин, що використовують відповідне програмне забезпечення. Для реалізації безперебійної роботи сервісу, навіть в умовах значного навантаження, використовують високопродуктивні сервери, комутатори та сховища великих об'ємів. Проте, все одно є необхідність купівлі дорогих серверів з великою продуктивністю, об'ємом сховища, організовувати серверні кімнати, що збільшує як капітальні (вартість обладнання) так і

операційні (електроенергія, оренда площ) витрати. Для зменшення витрат наразі широко використовується така модель забезпечення доступу за вимогою до обчислювальних ресурсів, як хмарні обчислення.

Згідно з визначенням Національного інституту стандартів і технології (NIST) США, «Хмарні обчислення» (від англ. Cloud Computing) — це модель забезпечення повсюдного та зручного доступу за вимогою, через мережу до спільного пулу обчислювальних ресурсів, що підлягають налаштуванню, і які можуть бути оперативно надані та вивільнені з мінімальними затратами на керування інфраструктурою та зверненнями до провайдера.

Хмарні обчислення є концепцією надання IT-ресурсів у вигляді послуг:

1. Інфраструктури
2. Платформи
3. Програмного забезпечення
4. Сховища
5. Робочого місця
6. Резервного копіювання
7. Аварійного відновлення
8. Інше

Підхід використання хмарних обчислень в бізнесі є наслідком переходу від продукто-орієнтованого виробництва до сервіс-орієнтованого надання послуг. Головна ідея – оренда послуг замість купівлі обладнання. Провайдери хмарних рішень надають можливість орендувати свої обчислювальні потужності для виконання будь-яких дій, що вимагають значних обчислювальних ресурсів, дисковий простір для збереження великого об’єму даних із гарантіями безпеки та вільного цілодобового доступу, користування ліцензійним високовартісним програмним забезпеченням та інше.

Такий підхід має очевидні переваги – швидке та ефективне надання послуг зі збереженням можливості гнучкого масштабування об’єму орендованих ресурсів. В цьому випадку між бізнесом та IT впроваджується підхід, в якому бізнес набуває можливості оперативно реалізовувати свої проекти, щоб відповідати актуальним вимогам реалій ведення бізнесу. Ефективне досягнення цілей бізнесу відбуватиметься з допомогою адаптивної конфігурації IT інфраструктури згідно поставленим вимогам.

Основні характеристики хмарних обчислень:

1. Поняття самообслуговування за вимогою має на увазі, що споживач має можливість вільно визначати і змінювати обсяг запитуваних обчислювальних ресурсів, а саме: швидкість доступу та кількість наданих обчислювальних ресурсів, обсяг збережених даних в хмарних ресурсах згідно власних вимог без звернення до постачальників хмарних послуг;
2. Надання універсального доступу. Користувачі мають можливість отримувати доступ до хмарних послуг, використовуючи будь-який термінальний пристрій, що з’єднаний з будь-якою дротовою або бездротовою мережею передачі даних.
3. Об'єднання ресурсів. Провайдер хмарних послуг концентрує ресурси в спільний пул для адаптивного перерозподілу наданих обчислювальних потужностей між всіма споживачами, згідно їхнім потребам. При цьому

розподіл ресурсів виконується постачальником, а споживач може лише контролювати основні параметри отриманої послуги, а саме швидкість доступу та ручний вибір центру обробки даних з власних міркувань, наприклад, географічного місцерозташування.

4. Еластичність. Об’єм отримуваних споживачем послуг може бути автоматично змінений ним в довільний момент часу, без необхідності попереднього контакту з постачальником послуг.
5. Облік споживання. Білінгова система провайдера автоматично розраховує обсяг спожитих кожним користувачем ресурсів (наприклад, обчислювальні потужності, зайняті канальні ресурси, обсяг даних, що зберігаються в хмарі і т.д.), і оцінює обсяг наданих кожному споживачеві послуг, що дозволить обчислити орендну виплату за отримані ресурси.

З цього випливає основний принцип хмарних обчислень – користувач платить лише за ті ресурси, які йому потрібні, і отримує їх за запитом у короткий термін. У зв’язку з цим, постає проблема динамічного визначення необхідної кількості орендованих ресурсів у постачальників хмарних сервісів для зменшення вартості обчислень за рахунок відсутності надлишкових ресурсів, що не використовуються із забезпеченням необхідних показників QoS. У зв’язку з цим потрібно визначитись, наскільки часто переглядати обсяг виділених ресурсів та змінювати його для недопущення порушення вимог до якості обслуговування та надлишкового використання ресурсів.

Для вирішення даної проблеми у розділі 5 запропоновано метод динамічного розподілу ресурсів для визначення оптимального обсягу ресурсів для віртуалізованих вузлів мережі мобільного зв’язку. Метод дозволяє з’ясувати обсяги потрібних ресурсів для забезпечення високого рівня обслуговування користувачів, уникнути надлишковості ресурсів та, як наслідок, простою. Крім того, метод дозволяє з’ясувати моменти часу, що є оптимальними для перерахунку обсягів ресурсів.

У даному дослідженні ефективність методу перевіряється шляхом його застосування у віртуалізованому ядрі мобільної мережі України.

7.2. Постановка задачі

Нехай технологія віртуалізації NFV застосовується до ядра мобільної мережі LTE (Long Term Evolution) України.

LTE – назва мобільного протоколу передачі даних, що відноситься до четвертого покоління рухомого (мобільного) радіозв'язку. Використовуючи сучасні технології обробки сигналу, такі як OFDM, Massive MIMO, Carrier Aggregation, спектрально ефективну модуляцію QAM-256, ця технологія забезпечує швидкість завантаження до 1 Гбіт/с. Наразі LTE є основним мобільним протоколом в більшості країн світу. Структура LTE мережі представлена на рис. 7.1.

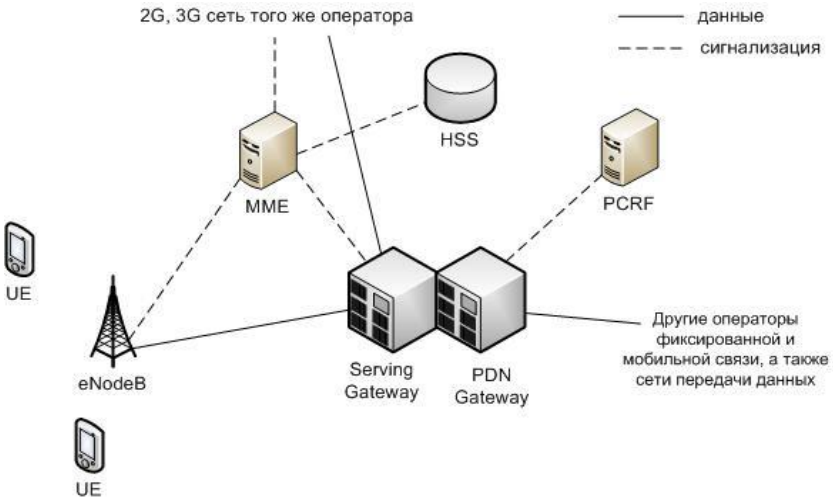

Рис. 7.1. Структура мережі LTE

Фізичні вузли мережі можуть бути віртуалізовані та мають таке призначення:

HSS – є великою базою даних і призначений для зберігання даних про абонентів. Окрім того, HSS генерує дані, необхідні для виконання процедур шифрування, аутентифікації і т.д. Мережа LTE може включати один або декілька HSS в залежності від географічної структури мережі та кількості абонентів.

MME – призначений для обробки сигналізації, зв'язаної з керуванням мобільністю абонентів у мережі.

SGW – призначений для обробки і маршрутизації пакетних даний, що поступають з/в підсистему базових станцій. Має пряме з'єднання з мережами 2 та 3 поколінь того ж оператора.

PGW – шлюз до мереж передачі даних інших операторів для мережі LTE. Основна задача PGW – маршрутизація трафіку мережі LTE іншим мережам передачі даних, таких як Інтернет-трафік, а також мережам GSM та UMTS.

PCRF – призначений для виконання білінгових послуг, таких як нарахування плати за надані послуги зв'язку, а також за забезпечення якості з'єднання у відповідності до заданих конкретним абонентом характеристик.

Віртуалізовані можуть бути всі фізичні вузли, або деякі з них. Нехай в загальному випадку віртуалізована мережа складається з n мережевих функцій, що відповідають фізичним вузлам мережі.

Термін «одиниця ресурсів» (resource unit – *RU*) в цьому контексті позначає обсяг обчислювальних ресурсів, ресурсів зберігання та ресурсів мережі, необхідний для обробки цифрового пакетного трафіку однією віртуалізованою мережевою функцією з інтенсивністю, що дорівнює пропускній здатності Fast Ethernet (тобто 100 Мбіт/с), за час d_i, де d_i є цільовим часом обробки трафіку відповідно до вимог SLA.

Завдання, які необхідно вирішити в рамках процесу розподілу ресурсів у віртуалізованій NFV мережі:

1. Який обсяг ресурсів (яка кількість одиниць ресурсів) має бути виділено кожній віртуалізованій мережевій функції?
2. Як часто необхідно змінювати обсяг ресурсів, щоб мінімізувати витрати і забезпечити достатній рівень QoS (якість обслуговування)?

Для цього запропоновано метод динамічного розподілу ресурсів у віртуалізованій мережі.

7.3. Опис досліджуваного методу динамічного розподілу ресурсів

Основна ідея методу динамічного розподілу ресурсів у віртуалізованій мережі мобільного оператора полягає у використанні довгострокової статистики, скоригованої на основі короткочасного прогнозу трафіку. Цей процес схематично показано на рис. 7.2.

Припустимо, що оператор зв'язку володіє довгостроковою статистикою зміни навантаження з часом (сіра крива) з деяким допустимим відхиленням (межі допустимого відхилення показані штриховими лініями на графіку). Обсяг ресурсів, необхідний для обробки статистичного навантаження показано блакитною лінією. Зелена крива показує реальний розподіл трафіку.

У момент τ крива навантаження перевищує рівень допустимого відхилення. У такому випадку діє механізм прогнозування: змінюється часовий інтервал адаптації ресурсу (часовий інтервал відображено пунктирними лініями); прогнозується зростання навантаження на цей інтервал (помаранчева крива), а кількість ресурсів перераховується відповідно до прогнозованого навантаження та перерозподіляється (червона крива на графіку).

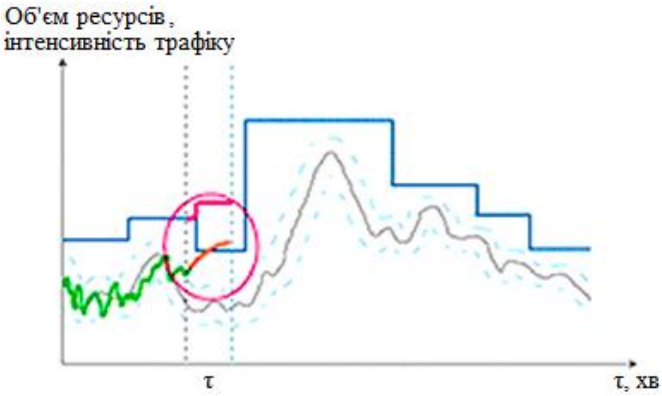

Рис.7.2. Прогнозування часового інтервалу адаптації та необхідного обсягу ресурсів у віртуалізованій мережі мобільного оператора

Тобто, заздалегідь визначені обсяги ресурсів для статистичного навантаження коригуються динамічно з огляду на поточний розподіл трафіку.

Таким чином, для здійснення вищеописаного підходу необхідно мати в системі розподілу ресурсів наступні блоки:

1. Блок моніторингу;
2. Блок прогнозування;
3. Блок розподілу ресурсів.

Архітектура системи показана на рис. 7.3. Дані, з якими оперують кожен блок, описані докладніше нижче.

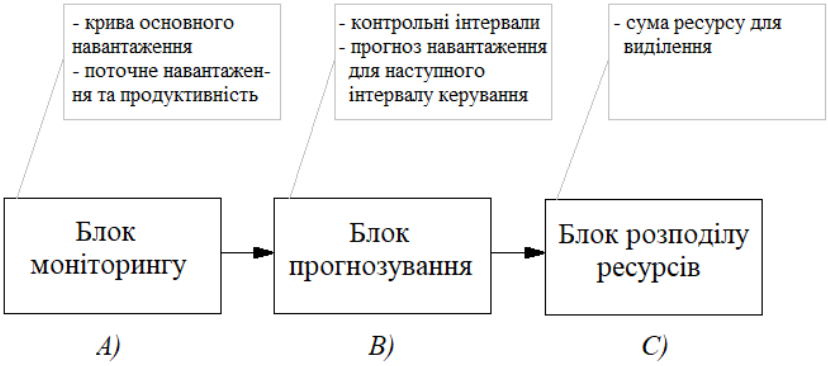

Рис.7.3. Архітектура системи розподілу ресурсів

7.4. Функції блоку моніторингу

У телекомунікаційній мережі може відбуватися довгострокове змінення трафіку (наприклад, за рахунок появи нових телекомунікаційних сервісів), а також короткочасне коливання трафіку (викликане, наприклад, масовими заходами або просто випадкові флуктуації). Статистичні дані про трафік за порівняно тривалі періоди утворюють основну криву трафіку (за день, за тиждень, за сезон тощо). Короткочасні коливання важко передбачити. Для врахування можливих змін основної кривої трафіку використовується блок моніторингу.

Блок моніторингу періодично перевіряє навантаження та продуктивність кожної віртуалізованої мережевої функції, порівнює дані з еталонними статистичними кривими навантаження та відстежує випадки, коли відхилення є більшим за статистично припустиме.

7.5. Функції блоку прогнозування

У разі зміни навантаження, що перевищує рівень статистичного відхилення, запускається механізм прогнозування. Змінюємо інтервали часу відповідно до виразу (5.1).

Після визначення інтервалу часу адаптації, коли переглядається обсяг виділених ресурсів, необхідно спрогнозувати навантаження на прийнятий інтервал. Для цього використовується змінений метод прогнозування

навантаження. Здійснюємо прогнозування навантаження на наступний інтервал адаптації *W*. Цей метод дозволяє прогнозувати максимальне навантаження, яке буде мати місце через інтервал часу *W*, на основі даних моніторингу попередніх періодів. Відповідно до цього методу прогнозоване навантаження визначається виразом (7.1):

$$\lambda(t) = \lambda_{pred}(j) + \sum_{j=t-h}^{t-1} \frac{\lambda_{obs}(j) - \lambda_{pred}(j)}{h}, \quad (7.1)$$

де $\lambda_{pred}(t)$ – прогнозована інтенсивність навантаження протягом певного інтервалу *t*, що виходить з аналізу історичних даних за останні дні;

$\lambda_{obs}(t)$ – фактична інтенсивність надходження запитів, яка спостерігається під час інтервалу часу *W* до наступної перевірки обсягу ресурсів;

$(\lambda_{obs}(t) - \lambda_{pred}(t))$ – помилка прогнозування.

Таким чином, блок прогнозування дозволяє визначити оптимальний інтервал часу до наступної зміни обсягів ресурсів та передбачити трафік на цей інтервал.

7.6. Функції блоку розподілу ресурсів

Метод динамічного розподілу ресурсів у віртуалізованій мережі дозволяє передбачити час до наступної зміни обсягів виділених ресурсів та наперед визначати їх достатній обсяг на прогнозований інтервал часу. Докладний опис наведено у розділі 5.

Дане дослідження має на меті експериментальне доведення ефективності методу при його застосуванні у віртуалізованому ядрі мобільної мережі України.

Для додаткового підвищення енергоефективності обробки даних у віртуалізованому ядрі мережі застосовано підхід РСРВ, що докладно описаний у розділі 6. Підхід реалізовано у вигляді алгоритму планувальника завдань. Він відрізняється одночасним врахуванням параметрів енергоефективності та продуктивності при розподілі завдань та складається з етапу попередньої індивідуальної атестації серверного кластера та етапу енергоефективного розподілу завдань.

Таким чином, у рамках даного дослідження проводиться аналіз ефективності методу динамічного розподілу ресурсів з додатковим використанням підходу РСРВ для підвищення енергоефективності обчислень на прикладі віртуалізованого ядра мережі LTE в масштабах України.

7.7. Експериментальне дослідження ефективності методу динамічного розподілу ресурсів у віртуалізованому ядрі мобільної мережі України з використанням підходу до підвищення енергоефективності обробки даних РСРВ

Експериментальне дослідження проводиться засобами імітаційного моделювання середовища Matlab за таким планом:

Підготовчі кроки:

Крок 1. Визначення розподілу трафіку в мережі.

Крок 2. Побудова топології мережі.

Крок 3. Моделювання добової кривої навантаження.

Крок 4. Визначення реального навантаження на мережах LTE кожного міста.

Крок 5. Визначення навантаження на елементах мережі.

Основні кроки:

Крок 6. Обчислення інтервалів адаптації за формулою (7.1) для проведення зміни кількості орендованих ресурсів в дата-центрі згідно формули (7.5).

Крок 7. Визначення енергетичної кривої для ЦОД.

Крок 8. Оцінка енергоспоживання для усієї мережі з використанням підходу РСРВ.

Крок 9. Загальна оцінка кількості спожитих ресурсів.

В моделі для спрощення топології допущено, що мережа LTE розгорнута в 5 найбільших містах в різних частинах України.

Крок 1. Визначення розподілу трафіку в мережі

Для урахування об’єму навантаження на вузли в залежності від географічного розташування останніх необхідно ввести корегуючі географічні коефіцієнти навантаження. Для цього розглянемо таблицю 7.1, в якій наведена кількість населення в містах Київ, Харків, Одеса, Дніпро, Львів станом на 2014 рік за статистикою з відкритих джерел. Вважатимемо, що послугами зв'язку користується 80% населення, з них, згідно прогнозу Vodafone-Україна, 20% матиме LTE-сумісні телефонні апарати.

Таблиця 7.1. Кількість населення в найбільших містах України станом на 2014 рік

Місто	**Область, АРК, міськрада**	**Чисельність населення, 2014**	**Абонентів LTE**
Київ	м. Київ	2 868 702	458 992
Харків	Харківська	1 451 132	232 181
Одеса	Одеська	1 017 022	162 724
Дніпро	Дніпропетровська	993 094	158 895

Місто	Область, АРК, міськрада	Чисельність населення, 2014	Абонентів LTE
Львів	Львівська	729 038	116 646

Додатково необхідно врахувати особливості профілю користувача в кожному регіоні. Молоді люди у віковій категорії 15-24 років споживають набагато більше трафіку, ніж представники інших вікових категорій. У табл. 7.2 показана доля LTE трафіку згідно статистики оператора «Мегафон» у 15 найбільших ВНЗ Новосибірська [15].

Таблиця 7.2. Порівняльний аналіз структури спожитого трафіку серед студентів ВНЗ із загальномережевими показниками

Середня доля LTE трафіку в мережі «Мегафон»	Середня доля LTE трафіку в мережі "Мегафон" серед студентів ВНЗ	Середній об'єм спожитого трафіку у мережі «Мегафон»	Середній об'єм спожитого трафіку у мережі «Мегафон» серед студентів ВНЗ	Середній об'єм спожитого трафіку у мережі LTE «Мегафон»	Середній об'єм спожитого трафіку у мережі LTE «Мегафон» серед студентів ВНЗ	Відношення спожитого LTE трафіку серед студентів ВНЗ до загальномережевого
43,00%	57,26%	4286 МБ	6852 МБ	1843 МБ	3923 МБ	2,13

З табл. 7.2 видно, що у середньому студент споживає у 2,13 рази більше LTE трафіку, ніж медіанне значення по мережі.

Використовуючи дані Держстату [18], визначимо кількість студентів у кожному місті, та кількість абонентів LTE мережі серед них та занесемо отримані дані в табл. 7.3.

Таблиця 7.3. Порівняльний аналіз кількості LTE абонентів серед найактивніших споживачів трафіку

Місто	**Чисельність населення, 2014**	**Відсоток студентів**	**Кількість студентів**	**Кількість LTE абонентів серед студентів**	**Відсоток LTE-суміних пристроїв**
Київ	2 868 702	13,40%	384 406	172 983	45%
Харків	1 451 132	11,90%	172 685	77 708	45%
Одеса	1 017 022	4,30%	43 732	19 679	45%
Дніпро	993 094	3,80%	37 738	16 982	45%
Львів	729 038	4,50%	32 807	14 763	45%

Враховуючи, що середній об’єм спожитого трафіку згідно дослідженням GlobalData складатиме 1Гб/міс, а для абонентів-студентів ВНЗ, відповідно – 2,13 Гб, та приймаючи географічний коефіцієнт Львова таким, що дорівнює одиниці, розрахуємо значення географічних коефіцієнтів інших міст. Отримані дані занесемо в табл. 7.4.

Таблиця 7.4. Географічний розподіл навантаження

Місто	**Абонентів LTE**	**Кількість LTE абонентів серед студентів**	**Кількість інших абонентів**	**Спожитий трафік абонентами – студентами,** *ГБ/міс*	**Спожитий іншими абонентами трафік,** *ГБ/міс*	**Сумарний спожитий трафік,** *ГБ/міс*	**Значення географічного коефіцієнта (Vtraff.city -Vtraff.Lviv)**
Київ	458 992	172 983	286 010	368453,2	286009,6	654462,8	4,91
Харків	232 181	77 708	154 473	165518,3	154473	319991,3	2,40
Одеса	162 724	19 679	143 044	41917,07	143044,1	184961,2	1,39
Дніпро	158 895	16 982	141 913	36171,46	141913,1	178084,6	1,34
Львів	116 646	14 763	101 883	31445,23	101883,1	133328,3	1,00

Оскільки мережа LTE складається з декількох вузлів, то в моделі прийняті коефіцієнти навантаження кожного вузла:

HSS – 0,05. HSS є великою базою даних і призначений для зберігання даних про абонентів. Окрім того, HSS генерує дані, необхідні для виконання процедур шифрування, аутентифікації і т.д. Мережа LTE може включати один

або декілька HSS в залежності від географічної структури мережі і кількості абонентів. HSS облуговує службовий трафік, тому об’єм трафіку, що обробляється HSS складає 5% від загального.

MME – 0,05. Призначений для обробки сигналізації, зв’язаної з керуванням мобільністю абонентів у мережі. Як і HSS обслуговує службовий трафік, тому його об’єм складає 5% від загального.

SGW – 0,85. Призначений для обробки і маршрутизації пакетних даних, що поступають з/в підсистему базових станцій. Має пряме з’єднання з мережами 2 та 3 поколінь того ж оператора. SGW обслуговує приблизно 85% всього трафіку, оскільки існує певна кількість абонентів, що не мають сумісного з 4G обладнання.

PGW – 0,7. Шлюз до мереж передачі даних інших операторів для мережі LTE. Основне завдання PGW – маршрутизація трафіку мережі LTE іншим мережам передачі даних, таких як Інтернет-трафік, а також мережам GSM та UMTS. Оскільки існує незначний у загальних масштабах об’єм даних, які оброблюються SGW, призначений для з’єднання абонентів однієї мережі, тому PGW оброблятиме 70% від загального об’єму трафіку.

PCRF – 0,05. Призначений для виконання білінгових послуг, таких як нарахування плати за надані послуги зв’язку, а також за забезпечення якості з’єднання у відповідності до заданих конкретному абоненту характеристиками. PCRF обслуговує службовий трафік в об’ємі 5% від загального.

В табл. 7.5 представлено розподіл навантаження між вузлами мережі одного міста згідно їхнього опису в пункті застосування NFV-технології для віртуалізації ядра мережі LTE.

Таблиця 7.5. Розподіл навантаження вузлами мережі

Назва вузла мережі	**Об'єм трафіка, що оброблюється вузлом (% від загального)**
HSS	5%
MME	5%
SGW	85%
PGW	70%
PCRF	5%

Крок 2. Побудова топології мережі

Серверні конфігурації визначено таким чином, щоб забезпечити безперебійне функціонування мережі міста при максимально можливих для цього регіону навантаженнях із запасом в 10%. LTE мережа кожного міста складається з одного комплекту усіх функціональних елементів, віртуалізованих в дата-центрі.

м. Київ – 1 SGW, 1 PGW, 1 MME, 1 HSS, 1 PCRF
м. Львів – 1 SGW, 1 PGW, 1 MME, 1 HSS, 1 PCRF
м. Харків – 1 SGW, 1 PGW, 1 MME, 1 HSS, 1 PCRF
м. Одеса – 1 SGW, 1 PGW, 1 MME, 1 HSS, 1 PCRF
м. Дніпро – 1 SGW, 1 PGW, 1 MME, 1 HSS, 1 PCRF

Така топологія дозволить підтримувати в робочому стані LTE мережу міста у випадку зникнення зв'язку з елементами мережі оператора, розташованих в інших містах. Зазначений підхід дозволить виконувати реконфігурацію мережі у випадках, коли функціональні вузли мережі одного міста перевантажені, коли з'являються черги на обслуговування та погіршуються показники QoS.

Таким чином, топологія мережі матиме вигляд, представлений на рис. 7.4.

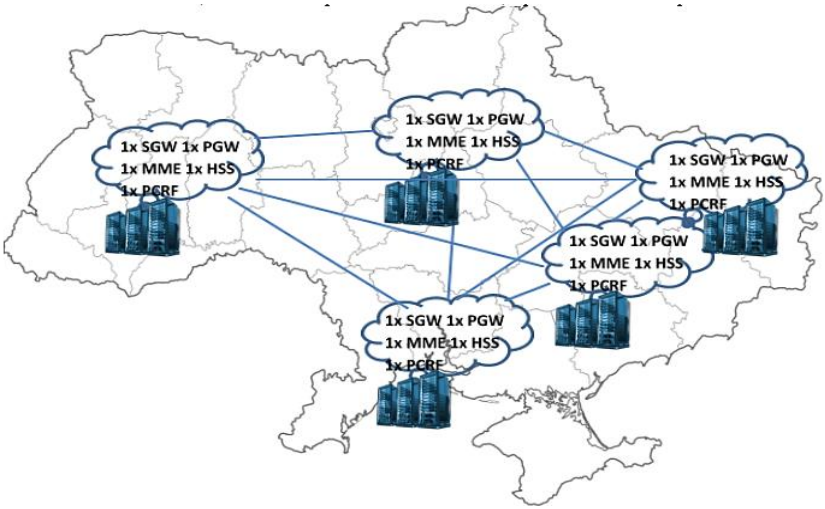

Рис. 7.4. Топологія мережі, яка моделюється

Крок 3. Моделювання добової кривої навантаження

Приклад розподілу навантаження на реальний кластер протягом доби зображено на рис. 7.5 [12]. Даний розподіл було використано при моделюванні.

Рис. 7.5. Приклад розподілу навантаження на реальний серверний кластер протягом доби [12]

З рис. 7.5 можна зробити висновок, що добова крива навантаження на серверний кластер має декілька проміжків зростання та спадання:

— від початку доби до першого мінімуму добового навантаження, позначеного цифрою 1, навантаження на серверний кластер має спадаючий характер, досягається мінімальне добове значення;
— від першого мінімуму до першого піку навантаження, позначеного цифрою 2, навантаження має тенденцію зростання;
— від першого піку до другого мінімуму, позначеного цифрою 3, навантаження спадає;
— від другого мінімуму до другого піку, позначеного цифрою 4, навантаження значно зростає, досягається максимальне добове значення;
— від другого піку навантаження до кінця доби навантаження спадає.

З урахуванням описаних вище закономірностей часового розподілу добового навантаження на серверний кластер, визначимо середні значення навантаження для моментів часу, що дорівнюють часу доби в 1,2,3,…,24 год. Сумарне добове навантаження при цьому дорівнює сумі споживання трафіку по містах за місяць (колонка «Сумарний спожитий трафік, ГБ/міс» табл. 7.4) поділене на кількість днів в місяці (30).

Всі отримані значення заносяться в табл. 7.6.

Таблиця 7.6. Добовий розподіл навантаження

Час доби, *год*	**Навантаження,** *Гб/год*	**Час доби,** *год*	**Середнє погодинне навантаження,** *Гб/год*
1	1500	13	2125
2	1250	14	2437,5
3	875	15	2625
4	750	16	2812,5
5	625	17	2843,75
6	875	18	2937,5
7	1062,5	19	3000
8	1687,5	20	3250
9	1875	21	2875
10	2125	22	2500
11	2625	23	2187,5
12	2250	24	1687,5

Для кращого сприйняття змін середнього погодинного навантаження телекомунікаційної мережі України на рис. 7.6 візуалізовано дані, наведені у табл. 7.5, із зазначенням характерних точок.

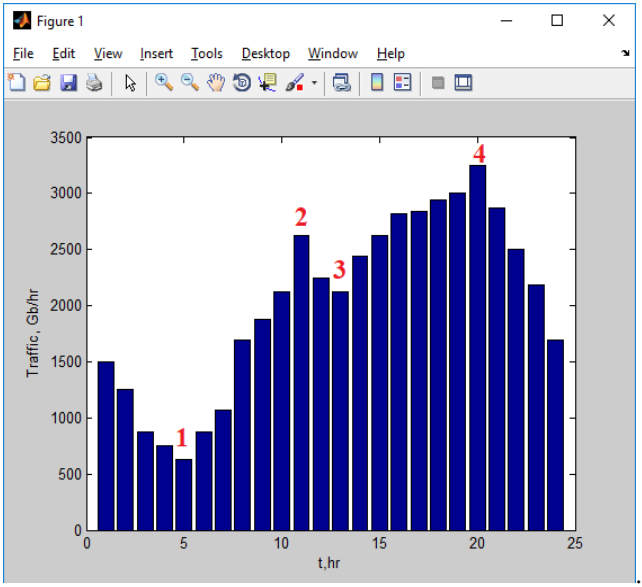

Рис. 7.6. Гістограма погодинного споживання трафіку LTE мережі

З рис. 7.6. видно, що погодинна характеристика споживання трафіку має декілька характерних точок, які відповідають локальним мінімумам та максимумам споживання.

Оскільки часовий крок в 1 годину для умов експерименту є занадто великим, то його необхідно зменшити до розмірності 1 хвилини. Використання однієї інтерполяційної формули для великої кількості вузлів, як у випадку інтерполяційних формул Ньютона або Лагранжа, є недоцільним. Такий інтерполяційний поліном (многочлен) характеризується значною інтенсивністю коливань, і його значення між вузлами сильно відрізнятимуться від значень інтерпольованої функції. Таким чином, при невисоких значеннях порядку точність інтерполяції є недостатньою. При використанні більш високих порядків починає проявляти себе феномен Рунге, який полягає в збільшенні до неприпустимих значень похибки між функцією і її інтерполюючим поліномом в точках, які не є точками інтерполяції,.

На рис. 7.7 показано ефект феномену Рунге при проведенні інтерполяції поліномом вищого порядку, де червона крива – це функція Рунге.

Блакитна крива – інтерполяція поліномом 5-го порядку (використовуючи п'ять рівновіддалених точок інтерполяції).

Зелена крива – інтерполяція поліномом 9-го порядку (використовуючи дев'ять рівновіддалених точок інтерполяції).

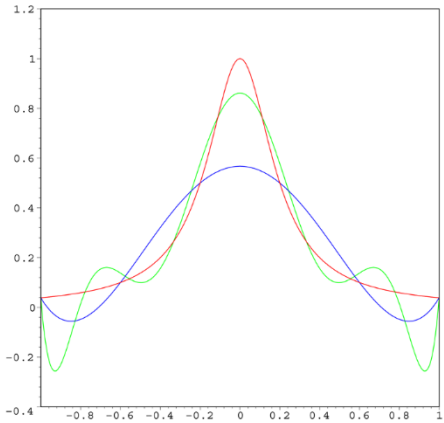

Рис. 7.7. Феномен Рунге

На рис. 7.7 видно, що в точках інтерполяції, помилка між функцією та інтерполючим поліномом (за визначенням) нульова. Між точками інтерполяції (особливо в регіоні близькому до крайніх точок *1, -1*), похибка між функцією і інтерполючим поліномом для поліномів більш високого порядку стає надто великою.

Однією з можливостей обійти такий недолік є застосування сплайн-інтерполяції [20]. Ідея сплайн-інтерполяції полягає в побудові поліномів між парами сусідніх вузлів інтерполяції, причому для кожної пари вузлів будується свій поліном. В задачах інтерполяції, інтерполяція сплайном краща, ніж інтерполяція многочленом, оскільки дає схожі результати навіть при менших степенях поліномів, а також при її використанні не виникає феномена Рунге.

Після проведення інтерполяції кубічними сплайнами за допомогою Matlab отримаємо похвилинний графік добового навантаження, який приймемо за базове навантаження, який зображено на рис. 7.8.

На рис. 7.8 зображено результат інтерполяції сплайнами, який точно відображає залежність зміни споживання трафіку протягом дня з точністю до хвилини. При цьому кумулятивна похибка інтерполяції складає менше 0,1%.

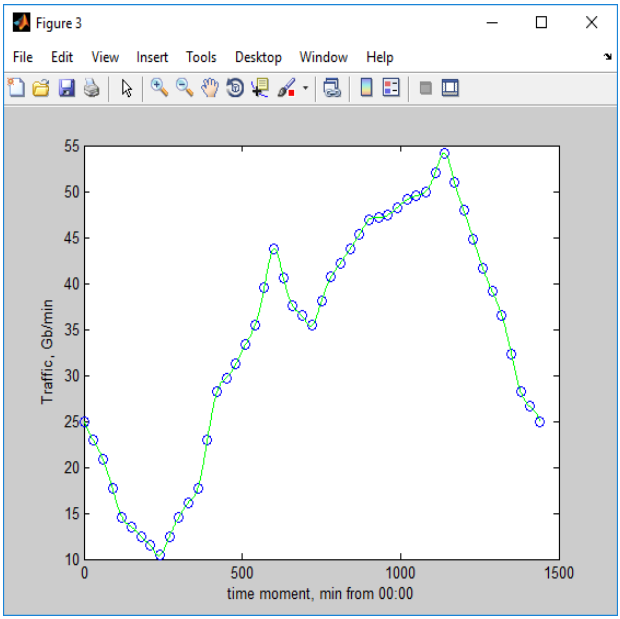

Рис. 7.8. Базова похвилинна крива навантаження

<u>Крок 4. Визначення реального навантаження на мережах LTE кожного міста</u>

Враховуючи те, що реальне навантаження не буде точно співпадати з базовим, оскільки останнє є усередненням статистичних даних, приймемо допущення, що реальне навантаження можна вирахувати таким чином:

$$Yreal = Ybasic + rand\ (-0.05 \ldots 0.05) * Ybasic. \qquad (7.2)$$

Іншими словами, до базового навантаження вноситься випадкове відхилення від базового значення, що відповідає реальній ситуації на будь-якій ділянці мережі.

На рис.7.9 показано розподілення сумарного навантаження по містам у відповідності до значень географічних коефіцієнтів та внесеного відхилення навантаження від базового.

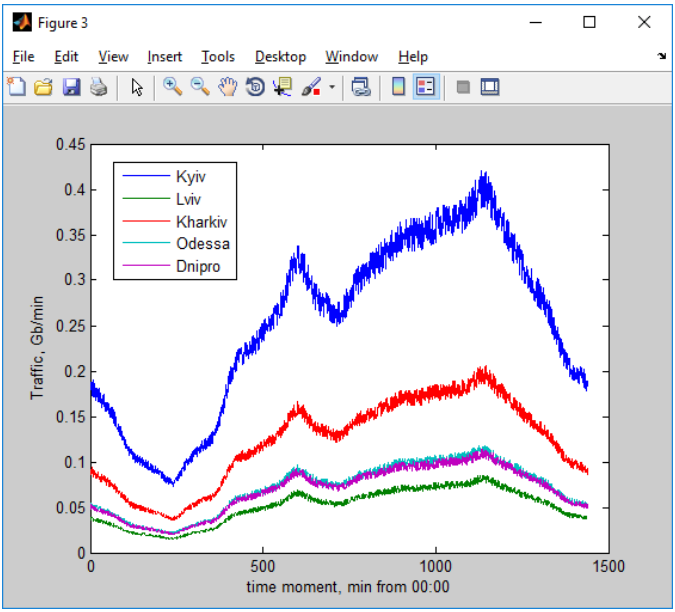

Рис. 7.9. Розподіл реального навантаження для визначених міст

На рис. 7.9 можна побачити, що абсолютний динамічний діапазон споживання трафіку для кожного міста різний та вимагатиме впровадження індивідуальних налаштувань роботи алгоритмів резервування обчислювальних ресурсів.

Крок 5. Визначення навантаження на елементах мережі

Навантаження, яке буде створене абонентами кожного міста, розподіляється згідно з наведеними у табл. 7.5 коефіцієнтами вузлового навантаження.

На рис. 7.10 показано навантаження, що надходить на функціональні вузли м. Києва протягом доби

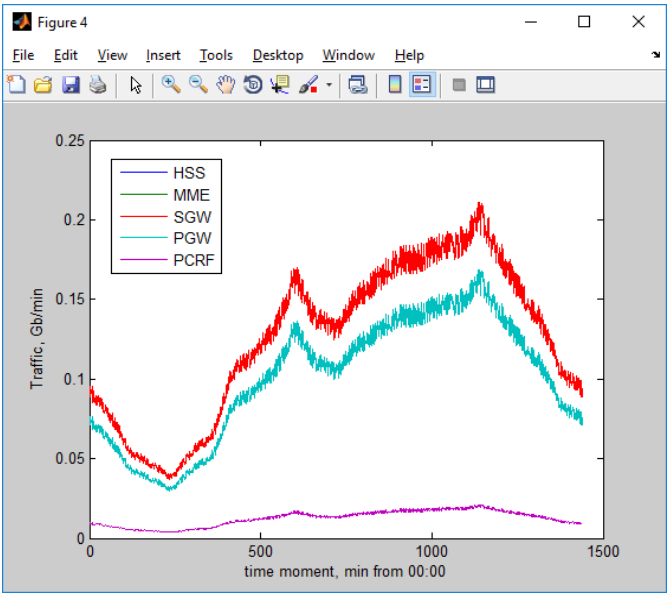

Рис. 7.10. Навантаження, що надходить на функціональні вузли м. Києва протягом доби

На рис.7.10 видно, що основна частина навантаження обслуговується двома вузлами: SGW і PGW, оскільки вони обробляють інформацію, а не сигнальний трафік.

Крок 6. Обчислення інтервалів адаптації для проведення зміни кількості орендованих ресурсів в дата-центрі

Центр обробки даних (ЦОД) кожного міста має певний пул ресурсів, що відповідає обладнанню, яке реалізовано. Відповідно до цього, експериментально обчислений базовий інтервал адаптації складає 10 хв. У цьому випадку досягається оптимальна кількість змін кількості виділенних ресурсів, при якій не відбувається перевантаження вузлів системи. Базовий інтервал був обчислений для базової кривої навантаження, що може не відповідати добовим флуктуаціям навантаження та різким пікам навантаження, які відбуваються на реальній мережі (масові заходи і т.п.) або значній зміні навантаження як в більшу, так і в меншу сторону (сезонні коливання, свята та ін.). Для цього реалізована адаптивна зміна інтервалу адаптації, яка опирається на базовий графік навантаження і попередні значення цього дня. Цим досягається більш оптимальна зміна кількості виділенних ресурсів, що відповідатиме реальній ситуації на мережі.

При цьому прийнята базова величина відхилення навантаження $\Delta\lambda$, що враховує доданий раніше фактор динамічної зміни навантаження. Зміна навантаження відбувається відповідно різниці реального відхилення навантаження та $\Delta\lambda$: якщо реальне відхилення більше за базове $\Delta\lambda$, то інтервал

зменшується, якщо менше – то збільшується. При цьому чим більша різниці між дельтами, тім більша зміна інтервалу адаптації.

Результати реалізації цього підходу виконано шляхом моделювання та показано на рис. 7.11.

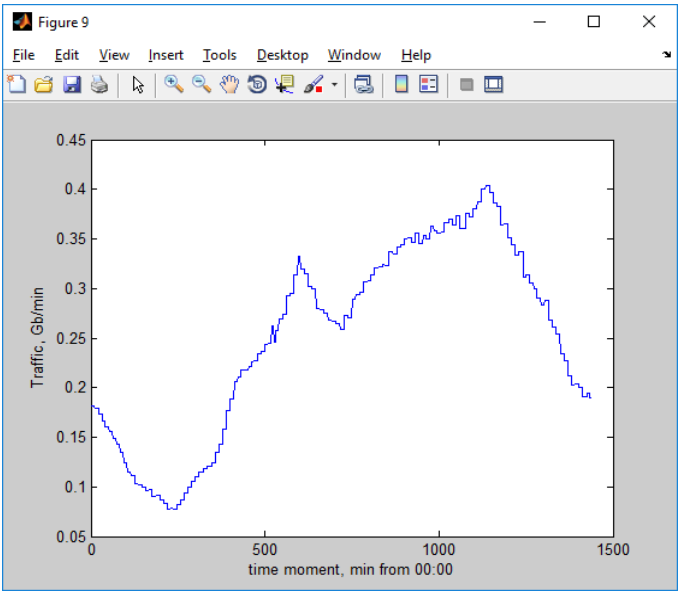

Рис. 7.11. Крива, що відображає визначені інтервали адаптації

На рис. 7.11 видно, що інтервали адаптації мають різні тривалості в залежності від оціненої похибки між прогнозованим і реальним значеннями навантаження. Для більшої наглядності на рис. 7.12 наведено скриншот з програми Matlab, на якому зображено масив значень тривалостей інтервалів у хвилинах.

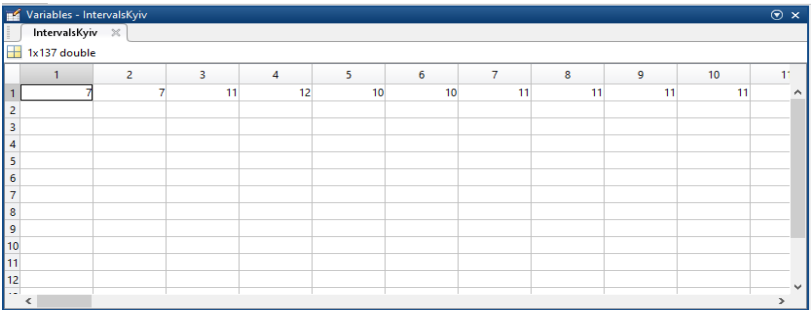

Рис. 7.12. Масив значень інтервалів адаптації у програмі Matlab

Як видно з рис. 7.12, перший та другий інтервали мають тривалість 7 хвилин, а третій інтервал – 11 хвилин і т.д.

Крок 7. Визначення енергетичної кривої для ЦОД

Наступним кроком у моделі задається енергетична крива для ЦОДу. За основу беруться експериментальні значення, отримані шляхом вимірювання енергоспоживання серверного кластеру ТУ Дрездена в залежності від завантаження його ресурсів. Для відповідності реальним значенням енергоспоживання ЦОД отримані значення були пропорційно збільшені.

Отримані значення представлені у таблиці 7.7.

Таблиця 7.7. Залежність споживання електроенергіі від завантаженості серверного кластеру

Завантаженість ЦП*, %	**Споживання електроенергії,** ***Вт/хв***
0	7633
25	9978
50	11322
75	12529
100	13608

*ЦП – центральний процесор

Отримані значення інтерполюються поліномом 5-ої ступені. Обране значення порядку поліному відповідає кількості точок інтерполяції та забезпечує достатню точність. Отриману криву енергоспоживання зображено на рис. 7.13.

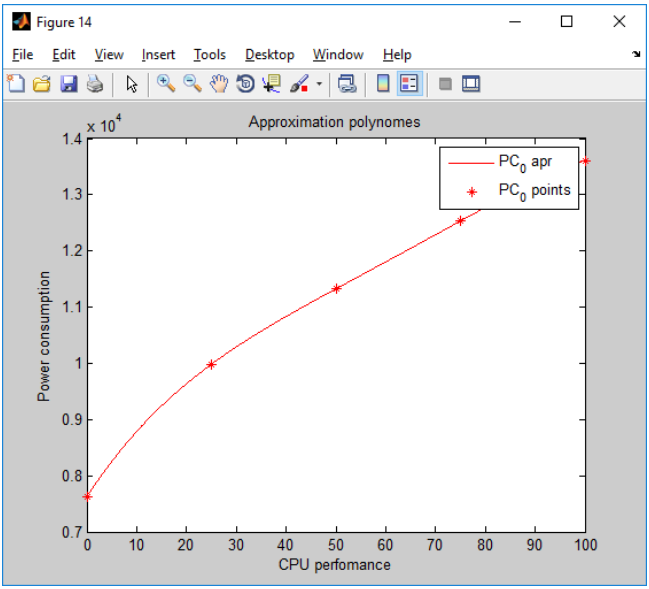

Рис. 7.13. Крива енергоспоживання мережі

На рис. 7.13 видно, що залежність енергоспоживання від навантаження центрального процесора (ЦП) є нелінійною: при невисоких значеннях навантаження ЦП (до 25%) кожний відсоток приросту завантаженості ЦП дає більший приріст енергоспоживання, ніж при високих значеннях завантаженості ЦП (75...100%).

Крок 8. Оцінка енергоспоживання для усієї мережі з використанням підходу РСРВ

Для оцінки можливості зменшення собівартості одиниці ресурсів проведено аналіз енергоспоживання для 2-х підходів:

1. Максимально можливе енергоспоживання запропонованої конфігурації кластера.
2. Зактивованим запропонованим у розділі 6 алгоритмом енергоефективного розподілу задач РСРВ.

Підхід 1 передбачає, що угода між провайдером хмарних послуг та споживачем (оператором) укладена щодо оренди певного обладнання (серверного кластера) та щомісячна вартість оренди є статичною. При цьому вважається, що вартість спожитої електроенергії дорівнює вартості максимально можливої кількості спожитої електроенергії для цього обладнання. Ресурси цього серверного кластеру не можуть бути віддані в оренду іншим споживачам незалежно від ступені використання. Необхідна конфігурація розраховується заздалегідь. Використання цього підходу прибирає можливість відмовлення в обслуговуванні через значну завантаженість серверних кластерів іншими споживачами хмарних послуг.

Підхід 2 передбачає, що угода між провайдером хмарних послуг та споживачем (оператором) укладена щодо оренди певної кількості ресурсів, яка може автоматично змінюватися згідно вимогам. Щомісячна вартість оренди змінюється в залежності від спожитої кількості ресурсів та, відповідно, спожитої електроенергії при цьому, яка розраховуватиметься згідно енергетичної моделі кроку 7. Ресурси цього серверного кластеру можуть бути віддані в оренду іншим споживачам, що у деяких випадках може привести до відмов в обслуговуванні операторами зв'язку з відповідним зменшенням показників QoS.

В рамках експерименту вважається, що взаємодія між провайдером хмарних послуг і оператором здійснюється згідно підходу 2. Результати моделювання наведені в табл. 7.8.

Таблиця 7.8. Оцінка енергоспоживання із використанням запропонованого алгоритму

Сумарне енергоспоживання в мережі, *МВт*	**Максимально можливе енергоспоживання,** *МВт*	**% від максимального**
1,3306	1,633	81,5%

Для зменшення енергоспоживання, а відповідно, і питомої вартості одиниці ресурсів можна використовувати ряд описаних у розділі 6 алгоритмів

балансування навантаження. Результати, показані в табл. 7.9, отримані унаслідок порівняння кожного з алгоритмів з алгоритмом балансування навантаження FIFO, принцип дії якого полягає в тому, що для обробки нової задачі обирається перший вільний зі списку вузол. У табл. 7.9 показані результати дослідження роботи популярних та запропонованих авторами алгоритмів балансування навантаження у порівнянні з FIFO.

Таблиця 7.9. Порівняльна характеристика енергоспоживання серверного кластеру при використанні описаних алгоритмів.

Алгоритм	**Сумарне енергоспоживання,** *Вт*	**Ефективність економії електроспоживання у порівнянні з алгоритмом FIFO**
FIFO	1324,836	-
RR	1396,948	-5,44% (Втрати)
RR Weighted	1359,887	-2,64% (Втрати)
Least Connections	1357,448	-2,46% (Втрати)
MES1	1309,055	1,19%
P-Algorithm	1381,166	-4,25% (Втрати)
PCPB	1281,167	3,30%
PCPB_v1	1284,339	3,06%
PCPB_v2	850,838	35,78%

З табл. 7.9 видно, що при використанні запропонованих алгоритмів PCPB можна додатково зекономити до 35,78% спожитої електроенергії. Оцінка роботи алгоритму PCPB_v2 значно залежить від конфігурації серверного кластеру та об’єму навантаження, що надходить на нього. Тому в даному дослідженні він не фігуруватиме як алгоритм для додаткового зменшення вартості одиниці споживаних ресурсів.

Крок 9. Загальна оцінка кількості споживаних ресурсів

Оскільки взаємодія між провайдером хмарних послуг і оператором здійснюється згідно підходу 2, то існує необхідність оцінити кількість ресурсів, які у випадку взаємодії згідно підходу 1 були б в стані простою.

Для оцінки ефективності запропонованого підходу середня кількість вільних ресурсів в день для кожного розглянутого міста визначалась як середня різниця між статично виділеними ресурсами та динамічно виділеними ресурсами за допомогою запропонованого підходу. Результати представлені в табл.. 7.10.

Таблиця 7.10. Результат динамічного розподілу ресурсів у ядрі EPC мобільної мережі

Місто	**Пул ресурсів**	**Середній обсяг звільнених ресурсів в день**	**Середній % звільнених ресурсів в день**
Київ	33	13,34	40,41%
Харків	30	11,52	38,40%
Дніпро	24	10,05	41,89%
Львів	14	6,35	46,43%
Одеса	20	9,29	45,39%
Загалом	120	50,55	41,78%

Як видно з табл. 7.10, сума ресурсів, що виділяються динамічно, в середньому на 41,78% менше, ніж статична сума в разі використання традиційного підходу розподілу.

На рис. 7.14 зображено результат роботи методу динамічного розподілу ресурсів у віртуалізованому ядрі мобільної мережі для м. Києва у графічному вигляді. Червона лінія ілюструє випадок статичного розподілу ресурсів. Зелена крива показує трафік протягом дня. Блакитна крива показує обсяг ресурсів, розподілених відповідно до запропонованого методу динамічно.

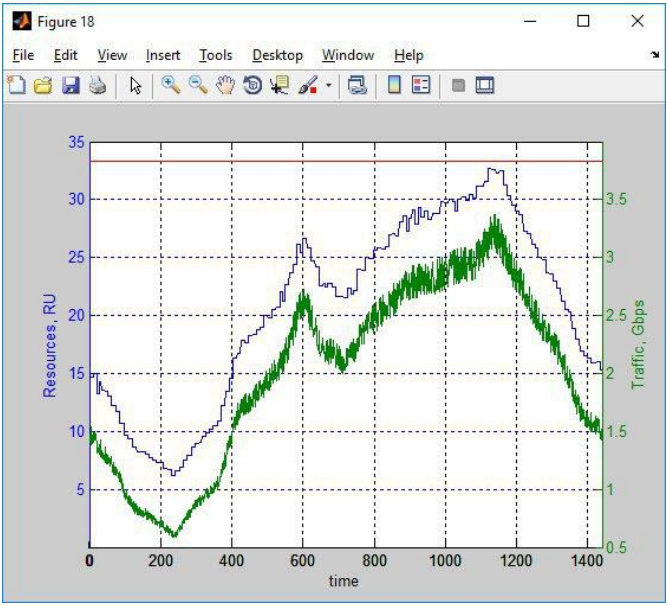

Рис. 7.14. Результат роботи методу динамічного розподілу ресурсів у віртуалізованому ядрі мобільної мережі для м. Києва.

З рис. 7.14 видно, що у випадку статичного розподілу ресурсів деяка кількість ресурсів буде знаходитись у стані простою.

Таким чином, застосування методу динамічного розподілу ресурсів дозволяє досягти економії ресурсів протягом доби, особливо в ранкові години, проте ресурси цього серверного кластеру можуть бути віддані в оренду іншим споживачам. Це означає, що існує ризик перевантаження провайдера запитами інших клієнтів і загальний доступний ресурсний пул може бути меншим за необхідне значення. Для зменшення цієї ймовірності провайдер може використовувати як апаратні (купівля додаткових обчислювальних потужностей), так і програмні засоби, наприклад, запропонований алгоритм PCPB покращує продуктивність (зменшує час обробки однієї задачі), що додатково зменшить час оренди ресурсів та, відповідно, грошових витрат на виплату оренди. В табл. 7.11 наведено порівняння продуктивності алгоритмів балансування навантаження у порівнянні з алгоритмом FIFO.

Таблиця 7.11. Порівняльна характеристика продуктивності серверного кластеру при використанні описаних алгоритмів за параметром часу обробки завдань.

Алгоритм	**Середній час обробки типової задачі,** *с*	**Ефективність,** *%*
FIFO	10,996	-
RR	16,221	-47,52%
RR Weighted	14,019	-27,49%
Least Connections	13,474	-22,54%
MES1	13,857	-26,02%
P-Algorithm	15,47	-40,69%
PCPB	10,88	1,05%
PCPB_v1	10,885	1,01%
PCPB_v2	10,893	0,94%

Як видно з табл. 7.11, використання алгоритму балансування навантаження PCPB не зменшує продуктивність обробки задач при одночасному підвищенні енергоефективності обчислень.

Висновки

1. У даному дослідженні проаналізовано ефективність методу динамічного розподілу ресурсів у віртуалізованій мережі мобільного оператора України з підвищенням енергоефективності обробки даних.
2. Аналіз проведено засобами імітаційного моделювання середовища Matlab. В якості об’єкта дослідження було обрано модель віртуалізованого ядра EPC мобільної мережі України. При моделюванні були використані реальні статистичні дані операторів мобільного зв’язку України.
3. Результати дослідження показали, що комбінація запропонованих підходів дозволяє досягти економії обчислювальних ресурсів на 41,78% завдяки застосуванню методу динамічного розподілу ресурсів та підвищити енергоефективність обчислень на 35,78% завдяки застосуванню підходу PCPB.

Література

1. http://www.sciencedirect.com/science/article/pii/S1389128615004752
2. http://www.sciencedirect.com/science/article/pii/S0140366416300858
3. Xiong G. A virtual service placement approach based on improved quantum genetic algorithm / G. Xiong et al. // Frontiers of Information Technology & Electronic Engineering. – 2016. – Vol. 17, No. 7. – P. 661–671.
4. Gember-Jacobson A. OpenNF: Enabling Innovation in Network Function Control / A. Gember-Jacobson et al. // Proceedings of the 2014 ACM Conference on SIGCOMM, ser. SIGCOMM ’14. – New York, NY, USA, 2014. – P. 163-174.
5. Rajagopalan S. Split/Merge: System Support for Elastic Execution in Virtual Middleboxes / S. Rajagopalan, D. Williams, H. Jamjoom, and A. Wareld // 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13). – Lombard, IL, USA, 2013. – P. 227-240.
6. Scholler M. Resilient deployment of virtual network functions / M. Scholler, M. Stiemerling, A. Ripke, R. Bless // Proc. 5th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT). – St. Petersburg, Russia, 2013. – P. 208-214.
7. Chowdhury N.M.M.K. Virtual Network Embedding with Coordinated Node and Link Mapping / N.M.M.K. Chowdhury, M. Rahman and R. Boutaba // INFCOM. – Rio de Janeiro, Brazil, 2009. – P. 1- 9.
8. Fajjari I. VNR Algorithm: A Greedy Approach For Virtual Networks Reconfigurations / I. Fajjari, N. Aitsaadi, G. Pujolle and H. Zimmermann. // IEEE Globel Communications Conference, Exhibition and Industry Forum. – Houston, United States, 2011. – P. 1-7.
9. Rahman M.R. Survivable Virtual Network Embedding / M. R. Rahman, I. Aib, and R. Boutaba // NETWORKING 2010. – 2010. – P. 40-52.
10. Abid H. A novel scheme for node failure recovery in virtualized networks / H. Abid; N. Samaan // 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013). – Ghent, Belgium, 2013. – P. 1154-1160.
11. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Information and Telecommunication Sciences. - 2016. № 1. P. 39–45.
12. Alexander Schill, Larysa Globa, Oleksandr Stepurin, Nataliia Gvozdetska, Volodymyr Prokopets “Power Consumption and Performance Balance (PCPB) scheduling algorithm for computer cluster”, // UkrMiCo’2017. – Odesa, Ukraine.– 11-15 September 2017. – P. 1- 8
13. http://www.osp.ru/os/2004/02/183912/
14. https://itc.ua/news/globaldata-srednee-potreblenie-trafika-na-odnu-sim-kartu-v-2017-godu-v-ukraine-sostavit-menee-1-gb-v-mesyats/
15. http://tayga.info/132324
16. https://corp.megafon.ru/press/news/federalnye_novosti/20170424-1510.html
17. http://database.ukrcensus.gov.ua/MULT/Dialog/statfile_c_files/stat_vik.html
18. http://uafrontier.com/vnz-v-regionah-ukrainy-na-pochatok-2014-roku/