

<sup>1</sup>Аспірант кафедри інформаційно-телекомунікаційних мереж Національного технічного університету України «Київський політехнічний інститут», Київ, Україна

<sup>2</sup>Канд. техн. наук, доцент, доцент кафедри інформаційно-телекомунікаційних мереж Національного технічного університету України «Київський політехнічний інститут», Київ, Україна

## ГІБРИДНА СИСТЕМА УПРАВЛІННЯ РЕСУРСАМИ ДЛЯ ВІРТУАЛІЗОВАНИХ МЕРЕЖЕВИХ ФУНКЦІЙ

**Актуальність.** Проблема зростання мобільного трафіку даних та збільшення кількості сервісів набуває глобального рівня, крім того, збільшуються об'єми та частота надходження службового трафіку, що передається мережею, а тому виникає необхідність у його ефективному управлінні з метою забезпечення потрібної якості обслуговування користувачів та оптимального використання ресурсів мережі оператора зв'язку. В таких умовах навантаження на сервер, що створюється в процесі організації з'єднання та його обслуговування, має свої особливості. Динамічне управління ресурсами є корисним методом для роботи в умовах коливань, які спостерігаються у навантаженні систем зв'язку. Технологія віртуалізації дозволяє реалізувати даний підхід. Аналітична модель системи буде привабливою, оскільки вона зможе оцінити характеристики системи в широкому діапазоні умов і бути обчисленою порівняно легко. Вона також може залучати методи чисельної оптимізації для проектування системи.

**Мета.** Підвищення ефективності роботи мобільної мережі за допомогою оптимального виділення ресурсів у телекомунікаційному середовищі.

**Метод.** Аналіз відомих публікацій, присвячених віртуалізації мережеских функцій мобільної мережі, дав змогу виявити підхід до моделювання виділення ресурсів, а також показав відсутність рішень щодо важливих питань цього процесу (продуктивності управління, системи прогнозування навантаження).

**Результати.** Пропонується підхід до моделювання і дослідження системи гібридного динамічного управління ресурсами мережеских функцій, де ресурси віртуальним вузлам виділяються динамічно і опортуністично в залежності від передбачуваних потреб.

**Висновки.** У статті вирішено завдання побудови системи управління ресурсами для віртуалізованих мережеских функцій. Запропоновано метод адаптації величини інтервалу управління розподілом ресурсів мережеских функцій, який забезпечує динамічне налаштування роботи системи, що дозволяє зменшити кількість надлишкової службової інформації, яка передається в мережі, та розвантажити мережескі вузли. Побудовано модель системи обслуговування навантаження, за допомогою якої представлено метод прогнозування навантаження, який враховує як накопичені за довгий строк дані статистики, так і останні тенденції, які спостерігаються у мережі, що дозволяє досягати раціонального співвідношення затрат на управління та підсумкового значення якості обслуговування.

**Ключові слова:** Network Functions Virtualization, управління ресурсами, моніторинг, прогнозування навантаження.

### НОМЕНКЛАТУРА

CPU – Central Processing Unit;

NFV – Network Functions Virtualization;

NIC – Network Interface Card;

SDN – Software-Defined Networking;

SLA – Service Level Agreement;

VNE – Virtual Network Embedding;

$C_i$  – кількість ресурсів мережескої функції  $i$ ;

$d_i$  – цільовий час відповіді мережескої функції  $i$ ;

$I_{base}$  – базове значення інтервалу управління;

$K$  – константа нормалізації;

$q_i^0$  – довжина черги на початку вікна адаптації функції  $i$ ;

$q_i$  – середня довжина черги;

$s_i$  – середній час обслуговування запиту однією одиницею ресурсу;

$T_i$  – середній час відповіді;

$W$  – інтервал управління

$\lambda_{obs(t)}$  – реальна інтенсивність надходження навантаження протягом інтервалу  $t$ ;

$\lambda_{pred(t)}$  – передбачена інтенсивність надходження навантаження під час інтервалу  $t$ ;

$\lambda_{basepred(t)}$  – базова передбачена інтенсивність надходження навантаження під час інтервалу  $t$ ;

$\mu_i$  – оцінка інтенсивності обслуговування.

### ВСТУП

Проблема зростання мобільного трафіку даних та збільшення кількості сервісів набуває глобального рівня,

крім того, збільшуються об'єми та частота надходження службового трафіку, що передається мережею, а тому виникає необхідність у його ефективному управлінні з метою забезпечення потрібної якості обслуговування користувачів та оптимального використання ресурсів мережі оператора зв'язку. В таких умовах навантаження на сервер, що створюється в процесі організації з'єднання та його обслуговування має свої особливості.

Завдяки новим віртуалізованим архітектурам, заснованим на SDN і NFV, можна забезпечити нові, диференційовані схеми функціонування, які будуть прибутковими для оператора, та основою для динамічного використання ресурсів. Визначення ресурсів, необхідних для належного функціонування віртуалізованих мереж, у значній мірі залежить від точної оцінки характеристик обслуговування навантаження. Правильне управління ресурсами є складним завданням через коливання навантаження. Як правило, навантаження дата центрів мають дуже пульсуючий характер і часто значно змінюються протягом дня. Тим не менш, багато навантажень в дата центрах зазвичай мають періодичні шаблони. Якщо визначити ці шаблони в навантаженні, тоді зможемо внести зміни відповідно до них в розподіл ресурсів і, отже, підвищити точність надання ресурсів і знизити енергоспоживання.

Таким чином, з огляду на труднощі в прогнозуванні пікових навантажень, у віртуалізованій мережі слід використовувати комбінацію прогностичного і реактивного управління.

У цьому напрямку в статті пропонується підхід до моделювання і дослідження системи гібридного динамічного управління ресурсами мережевих функцій у мережі телекомунікаційного оператора, де замість того, щоб виділяти фіксовану кількість ресурсів для даного віртуалізованого мережевого функціонального блоку протягом всього його життєвого циклу, ресурси віртуальним вузлам виділяються динамічно і опортуністично в залежності від передбачуваних потреб. З цією метою використовується гібридний підхід, який виділяє ресурси віртуальним вузлам з використанням методики прогнозування.

## 1 ПОСТАНОВКА ЗАДАЧІ

Метою запропонованого методу 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 надавати і коли.

Скільки надавати: для вирішення питання про те, скільки ресурсів виділити для кожної мережевої функції, будується аналітична модель. Представлена модель приймає в якості вхідних даних інтенсивність надходження вхідних запитів і вимоги обслуговування окремого запиту, і обчислює кількість ресурсів, необхідних кожному мережевому функціональному блоку, щоб впоратися з вимогами.

Коли надавати: рішення про те, коли надавати ресурси, залежить від динаміки навантажень. Телекомунікаційні навантаження зазнають довгострокових змін, таких як вплив години дня або сезонні ефекти, а також короткострокових коливань таких як спалахи навантаження. У той час як довгострокові коливання можуть бути передбачені заздалегідь, спостерігаючи за змінами в минулому, короткострокові коливання менш передбачувані, а в деяких випадках, не передбачувані. Запропонована методика використовує два різних методи для роботи в умовах змін, які спостерігаються в різних часових масштабах. Використовується прогностичне управління ресурсами для оцінки навантаження і відповідного управління, а також реактивне управління ресурсами для виправлення помилок у довгострокових прогнозах або для реагування на непередбачені спалахи навантаження.

Розглянемо мережу, в якій функціонує кілька мережевих функцій. Передбачається, що кожна така мережева функція вказує бажану вимогу до якості обслуговування (QoS); при цьому в даному випадку передбачаємо, що вимоги до QoS визначені в термінах цільового часу відповіді, аналогічно до [1]. Метою системи є забезпечення того, що середній час відповіді (або деякий процентиль часу відповіді), який спостерігається запитом мережевої функції не перевищує бажаний цільовий час відповіді. Загалом, кожен вхідний запит обслуговується декількома апаратними та програмними ресурсами на сервері, такими як CPU, NIC, диск і т. д. Припускаємо, що заданий цільовий час відповіді розділяється на кілька значень часу відповіді для конкретних ресурсів по одному для кожного такого ресурсу. Таким чином, якщо кожен запит на кожному ресурсі не витрачає часу більше, ніж призначене цільове значення, то загальний цільовий час відповіді для сервера буде задоволений [1]. Задача розді-

лення зазначеного значення часу відповіді сервера на значення часу відповіді для конкретного ресурсу виходить за рамки даного дослідження; у статті передбачається, що такі конкретні для ресурсу значення часу відповіді задані. Для простоти викладу припускаємо, що в системі наявний лише один тип ресурсу.

Формально,  $d_i$  позначає цільовий час відповіді мережевої функції  $i$  і  $T_i$  – спостережуваний середній час відповіді, тоді мережеві функції потрібно виділити таку кількість ресурсів, щоб  $T_i \leq d_i$ .

Використаємо таку постановку задачі щоб одержати механізм динамічного виділення ресурсів, який описано далі.

## 2 ОГЛЯД ЛІТЕРАТУРИ

Хоча NFV обіцяє істотну економію коштів, гнучкість і простоту розгортання, потенційні проблеми в реалізації віртуалізованих мережевих елементів, які можуть підтримувати вимоги до продуктивності реального світу, і досі залишаються відкритим питанням, й в даній часті NFV все ще перебуває на початкових етапах реалізації.

Кілька дослідницьких праць були зосереджені на розробці адаптивних систем, які можуть реагувати на зміни у навантаженні в контексті систем зберігання, загальних операційних систем, мережевих сервісів, веб-серверів і Інтернет дата центрів (наприклад, [2]). У даній статті розглядається абстрактна модель серверного ресурсу і представляються методи динамічного розподілу ресурсів; запропоновані модель і методи розподілу ресурсів застосовні до багатьох сценаріїв, де система або ресурс можуть абстрагуватись за допомогою GPS сервера.

Одним з ключових аспектів у області віртуалізації мережі є виділення фізичних ресурсів віртуальним функціям мережі. Вбудовування Віртуальної Мережі (VNE) є добре вивченою задачею. Тим не менше, більшість сучасних рішень пропонують статичну схему розподілу ресурсів, в якій коли віртуальна мережа відображається, перерозподіл ресурсів не відбувається протягом всього її життєвого циклу. Існує обмежена кількість децентралізованих і динамічних рішень VNE (як [3] або [4]). І навіть підходи, які пропонують рішення по динамічному вбудовуванню віртуальної мережі, все одно виділяють фіксовану кількість ресурсів для віртуальних вузлів і каналів на увесь період існування. Оскільки мережевий трафік не є статичним, це може привести до неефективного використання загальних мережевих ресурсів, особливо якщо фізична мережа відхиляє запити на вбудовування нових віртуальних мережевих функцій, при цьому резервуючи ресурси для віртуальних мережевих функцій, які знаходяться в умовах низької завантаженості [5].

Більшість існуючих робіт по динамічному управлінню ресурсами засновані на трьох підходах: теорії управління, моделюванні динаміки роботи і прогнозуванні навантаження [5]. Серед адаптивних систем, що використовують метод на основі теорії управління, можна виділити роботу [6]. Серед робіт, що засновані на динаміці роботи, можливо назвати працю [7]. Автори [8] використовують прогнозування навантаження.

Підводячи підсумок, різниця між запропонованим у даній статті підходом і згаданими вище є те, що в ньому ресурси, зарезервовані для використання віртуальними

мережевими функціями, не залишаються незмінними протягом усього життя віртуальної мережі. Здійснюється моніторинг віртуальних вузлів, і на основі їх реальних потреб у ресурсах, ресурси перерозподіляються, і в цьому випадку невикористані ресурси повертаються до фізичної мережі для використання іншими віртуальними мережами.

В аспекті того, яким чином отримується інформація про поточну ситуацію в мережі, управління ресурсами в NFV мережах подібне до управління прикладними програмами в дата центрах і хмарах. Існуючі рішення управління ресурсами серверів можуть бути класифіковані як прогностичні і реактивні рішення. Прогностичне виділення ресурсів передбачає наявність передбачуваного і стабільного шаблону у вимогах і розподіляє об'єми, як правило, в масштабі часу декількох годин або днів на основі шаблону. Однак, великі, непередбачувані сплески вимог можуть викликати серйозні порушення SLA. Реактивне виділення ресурсів, з іншого боку, виділяє ресурси в короткі проміжки часу (наприклад, кожні кілька хвилин) у відповідь на зміни навантаження. Чисто реактивні політики потенційно можуть швидко реагувати на зміни навантаження, але такі проблеми, як непередбачуваність, нестабільність і високі витрати управління обмежують їх застосування на практиці [9].

Отже, прогнозувати пік навантаження прикладної програми та виділяти ресурси на основі оцінок найгіршого випадку вкрай складно [10]. З огляду на труднощі в прогнозуванні пікових навантажень, прикладна програма має використовувати комбінацію прогностичного і реактивного управління. У той час як прогностичні методи добре працюють для онлайн прогнозування на великих часових інтервалах від декількох хвилин до декількох годин, реактивні методи дозволяють прогнозувати навантаження на короткі часові інтервали до декількох хвилин і швидко реагувати на нестационарні перевантаження [1].

Існує кілька підходів, які поєднують в собі прогностичне і реактивне управління [9, 10]. Хоча ці підходи мають спільні риси з гібридним підходом, який пропонується у даній статті, вони розрізняються за кількома аспектами. Запропонований підхід спрямований на оптимізацію продуктивності, енергоспоживання і вартості виділення ресурсів одночасно. На основі гібридної системи управління ресурсами розроблено метод динамічного моніторингу для ефективного управління ресурсами мережі та зменшення кількості службової інформації – інтервали управління ресурсами мають змінну довжину, тоді як в інших підходах до управління використовуються прості фіксовані інтервали. В [9] схожим чином використовується змінна довжина інтервалів, однак на протипагу цьому підходу, запропонований у статті підхід довжину інтервалів визначає динамічно в залежності від фактичної ситуації в мережі.

Основна мета полягає в тому, щоб розробити систему, яка може реагувати на перехідні зміни в навантаженні, в той час як теоретичний підхід на основі масового обслуговування намагається планувати запити, базуючись на стаціонарному навантаженні. Метод управління ресурсами на основі моделі було запропоновано в [1], що виконує розподіл ресурсів на основі моделювання ресурсу для співвіднесення метрик QoS і часток ресурсів,

що виділяються прикладній програмі. Робота [1] подібна до запропонованого у статті підходу, але у ньому модель масового обслуговування у часовій області адаптована до задачі віртуалізації мережевих функцій, а також доповнена механізмом вимірювання та передбачення.

### 3 МАТЕРІАЛИ І МЕТОДИ

Щоб виконати динамічне виділення ресурсів, засноване на вищевказаному формулюванні, на кожному сервері необхідно буде використовувати три компоненти, як вказано у [1]: (1) модуль моніторингу, який вимірює навантаження і показники продуктивності кожної мережевої функції (такі як інтенсивність надходження запитів, середній час відповіді тощо), (2) модуль прогнозування, який використовує вимірювання з модуля моніторингу для оцінки характеристик навантаження в найближчому майбутньому, і (3) модуль розподілення ресурсів, який використовує ці оцінки навантаження для визначення кількості ресурсів, яку необхідно виділити мережевими функціям. На рис. 1 показані ці три компоненти [1].

Використовуючи традиційні методи моніторингу стану ресурсів мережі, надлишкова службова інформація значно збільшується, що може негативно впливати на ефективність роботи мережі загалом через завантаженість каналів. Тому пропонується застосовувати механізм, суть якого полягає у динамічній зміні інтенсивності здійснення управління станом мережевого елемента залежно від різниці між передбаченим значенням навантаження і фактичним. Формула (1) описує принцип зміни частоти управління, де її значення розраховується з урахування певної базової величини  $I_{base}$  та історичних даних навантаження та його прогнозу:

$$W(t) = I_{base} - K \cdot \sum_{j=t-h}^{t-1} \frac{\max(0; \lambda_{obs}(j) - \lambda_{pred}(j))}{h} I_{base}. \quad (1)$$

На рис. 2 показано адаптацію частоти здійснення управління станом мережевого елемента до відхилення реального навантаження від передбаченого на мережевому елементі.

Модуль розподілу ресурсів викликається періодично (кожне вікно адаптації або при досягненні порогу) для динамічного розділення ресурсного об'єму між різними мережевими функціями, які працюють на загальних серверах в мережі. Як вже сказано, алгоритм адаптації запускається кожні  $W$  часові одиниці. Нехай  $q_i^0$  позначає довжину черги на початку вікна адаптації. Нехай  $\lambda_i$  позначає оцінку інтенсивності надходження заявок і  $\mu_i$  позначає оцінку інтенсивності обслуговування у наступному вікні адаптації (тобто, на наступні  $W$  часові одиниці). Тоді, припускаючи що значення  $\lambda_i$  і  $\mu_i$  є константними, довжина черги в будь-який момент часу  $t$  всередині наступного вікна адаптації задається формулою (2):

$$q_i(t) = \max(0; q_i^0 + (\lambda_i - \mu_i)t). \quad (2)$$

Оскільки ресурс моделюється як GPS сервер, інтенсивність обслуговування запиту мережевої функції дорівнює  $\mu_i = C_i / s_i$ . Середня довжина черги під час вікна адаптації визначається за формулою (3):

$$q_i = \frac{1}{W} \int_0^W q_i(t) dt. \quad (3)$$

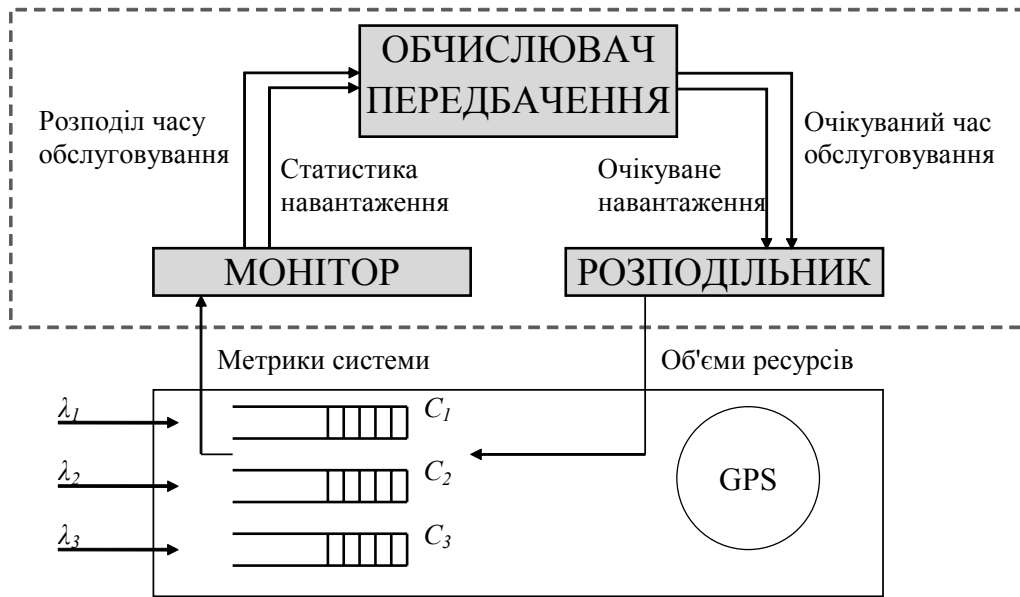
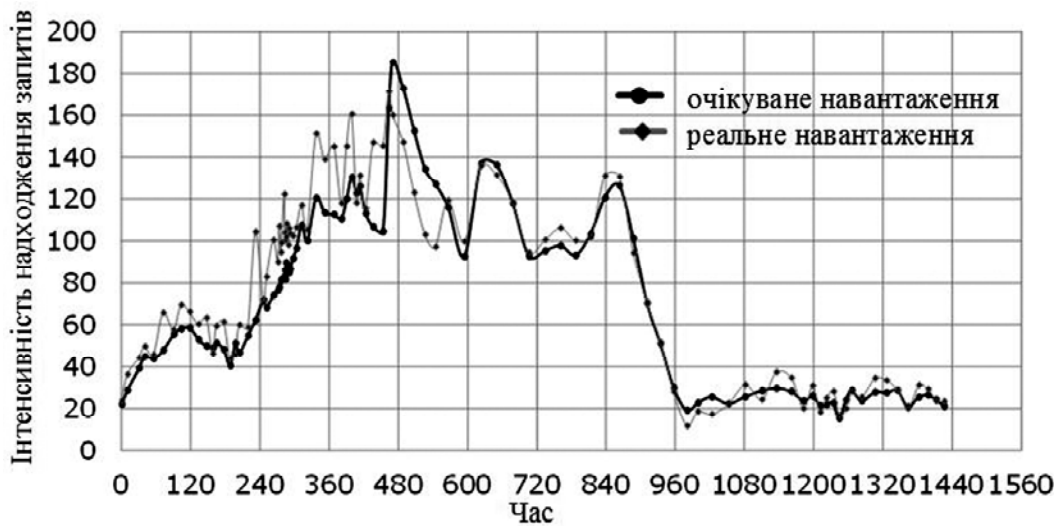
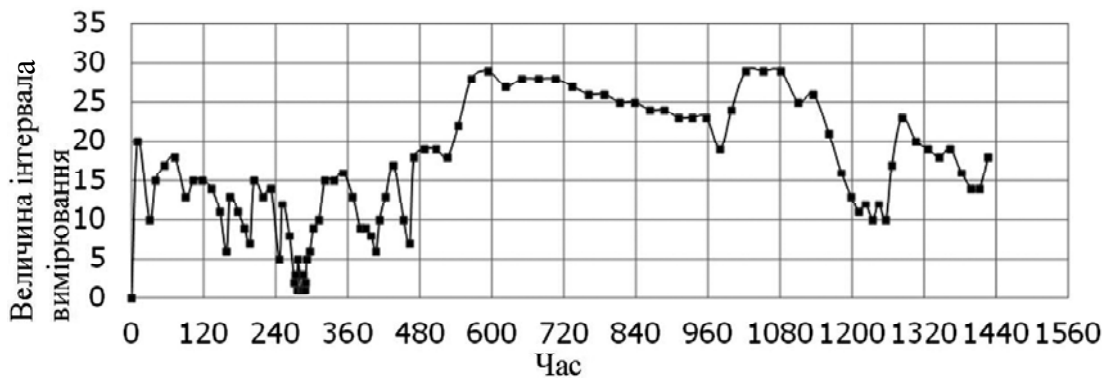


Рисунок 1 – Система динамічного виділення ресурсів [1]



а



б

Рисунок 2 – Динамічна зміна частоти адаптації:  
 а – навантаження мережевого елемента;  
 б – зміна інтенсивності адаптації відповідно до відхилення реального навантаження від очікуваного

Середній час відповіді  $T_i$  у той самий інтервал часу оцінюється за формулою (4):

$$T_i = \frac{q_i + 1}{\mu_i} \quad (4)$$

Параметри такої моделі залежать від її поточних характеристик, відповідно, ця модель застосовна в онлайн сценарії реагування на динамічні зміни в навантаженні.

Мережеві функції потрібно виділити кількість ресурсів, так що  $T_i \leq d_i$ , тоді кількість ресурсів, виділена мережеві функції  $C_i$  повинна задовольняти умову формули (5):

$$C_i \geq s_i \frac{q_i + 1}{d_i}. \quad (5)$$

Модифікований обчислювач передбачення інтенсивності надходження навантаження заснований на методиці, запропонованій в [10], використовує минулі спостереження за навантаженням, щоб передбачити пікову вимогу, яка буде зазнаватися протягом часу  $W$ .

Прогнозоване значення протягом наступного інтервалу  $\lambda_{pred}$  обчислюється відповідно до певного базового прогнозованого значення  $\lambda_{basepred}$  і коригується з використанням помилки, що спостерігається протягом певного історичного періоду  $h$ , відповідно до формули (6):

$$\lambda_{pred}(t) = \lambda_{basepred}(t) + \sum_{j=t-h}^{t-1} \frac{\lambda_{obs}(j) - \lambda_{pred}(j)}{h}. \quad (6)$$

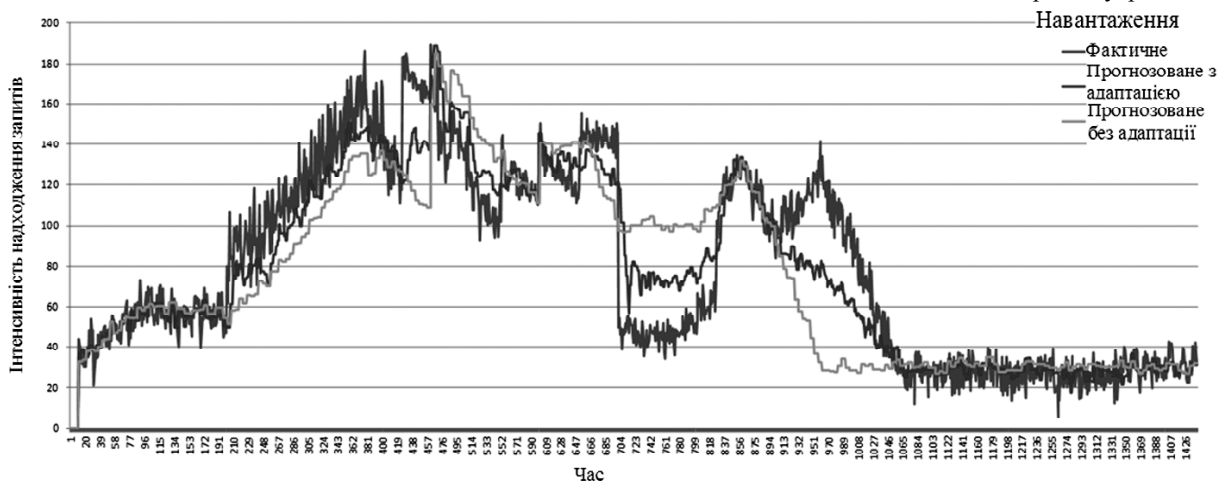


Рисунок 3 – Результати моделювання системи з динамічною адаптацією величини інтервала управління та прогнозом навантаження і системи без них



Рисунок 4 – Результати моделювання системи з 4 кроками управління

#### 4 ЕКСПЕРИМЕНТИ

Розглянемо задачу в системі Mathcad. Розглянемо роботу одного блоку протягом одного дня (1440 хвилин) і будемо вважати заданою базову компоненту прогнозованої інтенсивності надходження заявок протягом кожної хвилини  $\lambda_{pred}$ , а також нехай відомо середнє значення часу обслуговування заявки однією одиницею ресурсу  $s_i$ , і воно не змінюється, також припускаємо наявність ресурсу одного типу. Нехай вікно адаптації динамічно визначається на основі його базового значення і чотирьох останніх значень з історії моніторингу, де базове значення вікна адаптації складає 5 хвилин. Обмежень на розмір черги не вводимо, однак для розрахунків приймаємо, що у кінці інтервалу адаптації теоретичне значення розміру черги в поточних умовах складає 40 запитів.

Проведено збір статистики навантаження протягом одного дня тижня для місяця роботи серверу сайту, в результаті чого було визначено прогнозовану базову компоненту інтенсивності надходження заявок протягом кожної хвилини цього дня.

#### 5 РЕЗУЛЬТАТИ

Моделювання інтенсивності запитів на сервер сайту показало наступні результати.

На рис. 3 подано зміну величини інтервалу управління і зміни у прогнозованому значенні інтенсивності надходження навантаження у порівнянні з навантаженням, що фактично надійшло на сервер сайту протягом дня. З метою порівняння, на рис. 4 та рис. 5 подано результати моделювання для систем з іншими значеннями інтервалів управління.

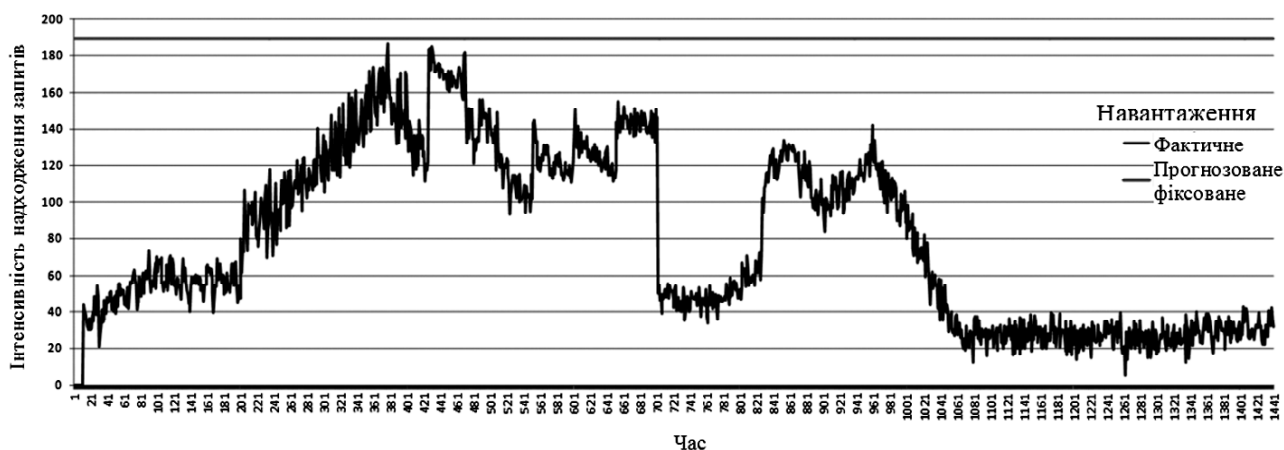


Рисунок 5 – Результати моделювання системи з фіксованим управлінням

## 6 ОБГОВОРЕННЯ

Для телекомунікаційної системи важливо забезпечити, з одного боку, бажану якість обслуговування користувачів, тобто виділити достатню кількість ресурсів для вузла, а з іншого боку – цінні ресурси потрібно використовувати доцільно, тобто не допускати зайвих їх витрат, що може впливати на витрати оператора та доступність ресурсів для інших віртуалізованих вузлів. З огляду на це, як «від’ємна», так і «додатня» помилка у прогнозуванні є небажаною.

Результати моделювання показали, що помилка у прогнозованому значенні у порівнянні з реальним може скласти 16%, при цьому «додатня» помилка складає 9% від реальної інтенсивності надходження заявок. Якщо не застосовувати систему динамічного регулювання величини вікна адаптації та систему врахування історичних даних при передбаченні, то помилка складатиме 26%, «додатня» помилка – 15%, з чого можна зробити ще і той висновок, що ресурси будуть застосовуватись вкрай неефективно. Для порівняння на рис. 4 також зображена система з усього лише 4 вікнами адаптації на день – для неї помилка у прогнозованому навантаженні складає вже аж 53%, додатково на рис. 5 оцінено систему без динамічного управління ресурсами, тобто де ресурси виділяються під найгірший випадок – тут помилка буде дорівнювати 235%.

## ВИСНОВКИ

Визначення характеристик застосування системи віртуалізації ресурсів мережі показало доцільність її використання для вирішення актуальних телекомунікаційних проблем, хоча такі фактори, як підвищена складність роботи і проблеми управління в порівнянні з перевагами даного підходу, повинні бути ретельно проаналізовані, перш ніж застосовувати його у системі зв’язку.

У статті обговорювалось те, що у присутності динамічно змінюваних навантажень, потрібні методи динамічного розподілу ресурсів, для того щоб забезпечити гарантії для віртуалізованих мережевих функцій, що працюють на загальних дата центрах. Для вирішення цієї проблеми представлено систему, яка поєднує в собі онлайн вимірювання з методами прогнозування і розподілу ресурсів. Зокрема, розроблено новий метод моніторингу та вимірювання, застосовуючи який можливо

зменшити кількість службової інформації, що циркулює у мережі, та уникнути негативних явищ перевантаження мережевих функцій. Для того, щоб захопити перехідну поведінку навантажень мережевих функцій, отримала подальший розвиток модель сервера GPS. Застосовано метод, який поєднує прогностичне і реактивне надання ресурсів: прогностичне надання ресурсів виділяє ресурси завчасно в очікуванні певного пікового навантаження, в той час як реактивне надання ресурсів вносить корективи після того, як спостерігалось аномальне збільшення навантаження.

Проведено оцінку методів з використанням моделювання в системі Mathcad. Результати показали, що ці методи можуть раціонально розподіляти ресурси системи, особливо при перехідних умовах перевантаження.

Система може застосовуватись для управління розгортанням віртуалізованих мережевих функцій на нижчорозташованій фізичній інфраструктурі для мінімізації витрат оператора зв’язку та покращення якості обслуговування абонентів.

У майбутньому планується оцінити компроміс між точністю та ефективністю використовуючи більш складні моделі аналізу часових рядів для прогнозування.

## ПОДЯКИ

Роботу проведено в рамках держбюджетної науково-дослідної роботи «Технологія обробки сервісів з інтеграцією інформаційних ресурсів в системах підтримки операційної діяльності підприємств зв’язку» 0115U000217.

## СПИСОК ЛІТЕРАТУРИ

1. Chandra A. Dynamic resource allocation for shared data centers using online measurements / A. Chandra, W. Gong, P. Shenoy // Quality of service : 11th International Workshop, Berkeley, CA 2–4 June 2003 : proceedings. – Springer, 2003. – P. 381–398.
2. Sampaio L. R. Towards practical auto scaling of user facing applications / L. R. Sampaio, R. V. Lopes // Cloud Computing and Communications (LATIN CLOUD) : 2012 IEEE Latin America Conference, Porto Alegre, Rio Grande do Sul, Brazil, 26–27 November 2012 : proceedings. – IEEE, 2012. – P. 60–65.
3. Virtual network embedding for evolving networks / [Z. Cai, F. Liu, N. Xiao et al.] // Global Telecommunications (GLOBECOM 2010) : conference, Miami, Florida, 6–10 December 2010 : proceedings. – IEEE, 2010. – P. 1–5.

4. A cost efficient framework and algorithm for embedding dynamic virtual network requests / [G. Sun, H. Yu, V. Anand, L. Li] // *Future Generation Computer Systems*. – 2013. – Vol. 29, № 5. – P. 1265–1277.
5. Design and Evaluation of Learning Algorithms for Dynamic Resource Management in Virtual Networks / [R. Mijumbi, J.-L. Gorricho, J. Serrat et al.] // *Network Operations and Management Symposium (NOMS), Krakow, 5–9 May 2014*. – IEEE, 2014. – P. 1–9.
6. Patikirikorala T. A multi-model framework to implement self-managing control systems for QoS management / [T. Patikirikorala, A. Colman, J. Han, and L. Wang] // *Software Engineering for Adaptive and Self-Managing Systems: symposium, Waikiki, Honolulu, 21–28 May 2011: proceedings*. – New York: ACM, 2011. – P. 218–227.
7. Game Theoretic Distributed Dynamic Resource Allocation with Interference Avoidance in Cognitive Femtocell Networks / [W. Lai, M. Chiang, S. Lee, T. Lee] // *Wireless Communications and Networking: conference, Shanghai, 7–10 April 2013: proceedings*. – IEEE, 2013. – P. 3364–3369.
8. Prediction-Based Dynamic Resource Allocation for Video Transcoding in Cloud Computing / [F. Jokhio, A. Ashraf, S. Lafond et al.] // *Parallel, Distributed, and Network-Based Processing: 21st Euromicro International Conference, Belfast, 27 February – 1 March 2013: proceedings*. – IEEE, 2012. – P. 254–261.
9. Gandhi A. Minimizing Data Center SLA Violations and Power Consumption via Hybrid Resource Provisioning / [A. Gandhi, Y. Chen, D. Gmach and other] // *Green Computing: International Conference and Workshops, Orlando, FL, 25–28 July 2011: proceedings*. – IEEE, 2011. – P. 1–8.
10. Agile dynamic provisioning of multi-tier Internet applications / [B. Urgaonkar, P. Shenoy, A. Chandra and other] // *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*. – 2008. – Vol. 3, № 1. – P. 1–39.

Стаття надійшла до редакції 27.10.2016.

Після доробки 03.12.2016.

Сулима С. В.<sup>1</sup>, Скулиш М. А.<sup>2</sup>

<sup>1</sup>Аспирант кафедри інформаційно-телекомунікаційних мереж Національного технічного університету України «Київський політехнічний інститут», Київ, Україна

<sup>2</sup>Канд. техн. наук, доцент, доцент кафедри інформаційно-телекомунікаційних мереж Національного технічного університету України «Київський політехнічний інститут», Київ, Україна

#### ГИБРИДНАЯ СИСТЕМА УПРАВЛЕНИЯ РЕСУРСАМИ ДЛЯ ВИРТУАЛИЗИРОВАННЫХ СЕТЕВЫХ ФУНКЦИЙ

**Актуальность.** Проблема роста мобильного трафика данных и увеличения количества сервисов приобретает глобальный уровень, кроме того, увеличиваются объемы и частота поступления служебного трафика, передаваемого сетью, а поэтому возникает необходимость в его эффективном управлении с целью обеспечения требуемого качества обслуживания пользователей и оптимального использования ресурсов сети оператора связи. В таких условиях нагрузка на сервер, создаваемая в процессе организации соединения и его обслуживания, имеет свои особенности. Динамическое управление ресурсами является полезным методом для работы в условиях колебаний, наблюдаемых в нагрузке систем связи. Технология виртуализации позволяет реализовать данный подход. Аналитическая модель системы будет привлекательной, поскольку она сможет оценить характеристики системы в широком диапазоне условий и быть рассчитанной сравнительно легко. Она также может привлекать методы численной оптимизации для проектирования системы.

**Цель.** Повышение эффективности работы мобильной сети с помощью оптимального выделения ресурсов в телекоммуникационной среде.

**Метод.** Анализ известных публикаций, посвященных виртуализации сетевых функций мобильной сети, позволил выявить подход к моделированию выделения ресурсов, а также показал отсутствие решений по важным вопросам этого процесса (производительности управления, системы прогнозирования нагрузки).

**Результаты.** Предлагается подход к моделированию и исследованию системы гибридного динамического управления ресурсами сетевых функций, где ресурсы виртуальным узлам выделяются динамически и оппортунистически в зависимости от прогнозируемых потребностей.

**Выводы.** В статье решена задача построения системы управления ресурсами для виртуализированных сетевых функций. Предложен метод адаптации величины интервала управления распределением ресурсов сетевых функций, который обеспечивает динамическую настройку работы системы, что позволяет уменьшить количество избыточной служебной информации, передаваемой в сети, и разгрузить сетевые узлы. Построена модель системы обслуживания нагрузки, с помощью которой представлен метод прогнозирования нагрузки, учитывающий как накопленные за длительный срок данные статистики, так и последние тенденции, наблюдаемые в сети, что позволяет достигать рационального соотношения затрат на управление и итогового значения качества обслуживания.

**Ключевые слова:** Network Functions Virtualization, управление ресурсами, мониторинг, прогнозирование нагрузки.

Sulima S.<sup>1</sup>, Skulysh M.<sup>2</sup>

<sup>1</sup>Past-graduate student of Department of Information and Telecommunication Networks, National University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine

<sup>2</sup>PhD, Associate Professor, Associate Professor of Information and Telecommunication Networks, National University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine

#### HYBRID RESOURCE PROVISIONING SYSTEM FOR VIRTUAL NETWORK FUNCTIONS

**Context.** The problem of growth of the mobile data traffic and the number of services becomes global, moreover, volume and frequency of control traffic transmitted through the network are increasing, and therefore there is a need for its effective management to ensure the quality of service required by users and optimal use of mobile network resources. In such circumstances, the load on the server that is created in the process of establishing the connection and its serving has its considerations. Dynamic resource provisioning is a useful technique for handling the variations seen in communication systems workloads. Virtualization technology allows to implement this approach. An analytic model of a system would be attractive as it would be able to evaluate system characteristics under a wide range of conditions, and to be computed comparatively easily. It is also can incorporate numerical optimization techniques for system design.

**Objective.** To improve the efficiency of mobile network through optimal resource allocation in telecommunication environment.

**Method.** Analysis of the known publications devoted to virtualization of network functions of mobile network has shown the modeling approach to resource allocation and also has shown the absence of decisions on important issues of this process (performance of management, load prediction system).

**Results.** An approach to model and investigate hybrid dynamic system of network functions resource management, where resources are provided to virtual nodes dynamically and opportunistically based on predicted needs, is proposed.

**Conclusions.** In the paper the problem of provisioning system design for virtualized network functions is solved. A method for adapting the size of network function's resource allocation control interval is proposed that provides dynamic configuration of the system, reducing excessive service data transmitted in the network, and decrease the load of network nodes. The workload service system model is built which outlines a method of workload forecasting, which takes into account long-term accumulated statistics and recent trends observed in a network that allows a rational level of management costs and final values of quality of service.

**Keywords:** Network Functions Virtualization, resource provisioning, monitoring, workload forecasting.

## REFERENCES

1. Chandra A., Gong W., Shenoy P. Dynamic resource allocation for shared data centers using online measurements, *Quality of service : 11th International Workshop, Berkeley, CA 2-4 June 2003 : proceedings*. Springer, 2003, pp. 381-398.
2. Sampaio L. R., Lopes R. V. Towards practical auto scaling of user facing applications, *Cloud Computing and Communications (LATIN CLOUD) : 2012 IEEE Latin America Conference, Porto Alegre, Rio Grande do Sul, Brazil, 26-27 November 2012 : proceedings*, IEEE, 2012, pp. 60-65.
3. Cai Z., Liu F., Xiao N., Liu Q., and Wang Z. Virtual network embedding for evolving networks, *Global Telecommunications (GLOBECOM 2010) : conference, Miami, Florida, 6-10 December 2010 : proceedings*. IEEE, 2010, pp. 1-5.
4. Sun G., Yu H., Anand V., Li L. A cost efficient framework and algorithm for embedding dynamic virtual network requests, *Future Generation Computer Systems*, 2013, Vol. 29, No. 5, pp. 1265-1277.
5. Mijumbi R., Gorricho J.-L., Serrat J., Claeysy M., Turcky F. D., Latr S. Design and Evaluation of Learning Algorithms for Dynamic Resource Management in Virtual Networks, *Network Operations and Management Symposium (NOMS), Krakow, 5-9 May 2014*, IEEE, 2014, pp. 1-9.
6. Patikirikoralala T., Colman A., Han J., and Wang L. A multi-model framework to implement self-managing control systems for QoS management, *Software Engineering for Adaptive and Self-Managing Systems : symposium, Waikiki, Honolulu, 21-28 May 2011 : proceedings*. New York, ACM, 2011, pp. 218-227.
7. Lai W., Chiang M., Lee S., Lee T. Game Theoretic Distributed Dynamic Resource Allocation with Interference Avoidance in Cognitive Femtocell Networks, *Wireless Communications and Networking : conference, Shanghai, 7-10 April 2013 : proceedings*. IEEE, 2013, pp. 3364-3369.
8. Jokhio F. Ashraf A., Lafond S., Porres I., Lilius J. Prediction-Based Dynamic Resource Allocation for Video Transcoding in Cloud Computing, *Parallel, Distributed, and Network-Based Processing : 21st Euromicro International Conference, Belfast, 27 February - 1 March 2013 : proceedings*. IEEE, 2012, pp. 254-261.
9. Gandhi A., Chen Y., Gmach D., Arlitt M., and Marwah M. Minimizing Data Center SLA Violations and Power Consumption via Hybrid Resource Provisioning, *Green Computing : International Conference and Workshops, Orlando, FL, 25-28 July 2011 : proceedings*. IEEE, 2011, pp. 1-8.
10. Urgaonkar B., Shenoy P., Chandra A., Goyal P., Wood T. Agile dynamic provisioning of multi-tier Internet applications, *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 2008, Vol. 3, No. 1, pp. 1-39.