

## ПРИНЦИП ПРОГНОЗУВАННЯ НЕОБХІДНОГО ВІРТУАЛЬНОГО РЕСУРСУ ХМАРНОЇ СИСТЕМИ ДЛЯ ОПЕРАТОРА МОБІЛЬНОГО ЗВ'ЯЗКУ

*Задача планування ресурсів системи обслуговування службових потоків оператора мобільного зв'язку є важливим завданням при організації гетерогенного телекомунікаційного середовища. У статті запропоновано принцип прогнозування необхідного віртуального ресурсу у хмарі для забезпечення ефективної роботи гетерогенного телекомунікаційного середовища, який дозволить, використовуючи дані системи моніторингу, розрахувати конфігурацію системи обслуговування та розклад використання ресурсів з метою забезпечення показників ефективності.*

*Скулиш М.А., Романов О.І., Нестеренко Н.Н. Принцип прогнозирования необходимого виртуального ресурса облачной системы для оператора мобильной связи. Планирование ресурсов системы обслуживания служебных потоков оператора мобильной связи является важной задачей при организации гетерогенной телекоммуникационной среды. В статье предложен принцип прогнозирования необходимого виртуального ресурса в облаке для обеспечения эффективной работы гетерогенной телекоммуникационной среды, который позволит, используя данные системы мониторинга, рассчитать конфигурацию системы обслуживания и расписание использования ресурсов с целью обеспечения показателей эффективности.*

*M. Skulysh, O. Romanov, M. Nesterenko The principle of predicting the necessary virtual resource in Cloud for the mobile operator. Planning the resources for managing the mobile operator system service flows is an important task in the organization of a heterogeneous telecommunications environment. The article proposes the principle of predicting necessary virtual resource in the Cloud to ensure efficient operation of a heterogeneous telecommunication environment. It will allow to calculate the configuration of service system and to create the resource usage schedule using the monitoring system data in order to provide performance indicator at a given level.*

*Ключові слова: гетерогенна телекомунікаційна мережа, прогнозування навантаження, планування ресурсів, віртуалізація, розклад залучення ресурсів.*

### **Вступ та постановка завдання**

Наслідком стрімкого розвитку хмарних технологій стала поява на ринку великої кількості провайдерів, які займаються наданням телекомунікаційних сервісів та послуг. Для обслуговування більшої кількості абонентів оператори залучають додаткові (орендовані) ресурси серверів, що також дозволяє покращити показники *QoS* [1-2]. Основна проблема полягає у вирішенні задачі керування орендованими віртуальними ресурсами, в гетерогенній телекомунікаційній системі оператора мобільного зв'язку, із дотриманням необхідних показників якості обслуговування абонентів. А саме, задача зводиться до визначення додаткової кількості віртуальних ресурсів або їх комбінацій (конфігурацій), які потрібно залучити в поточний момент часу.

В умовах гетерогенного телекомунікаційного середовища блок керування обслуговуванням системи мобільного зв'язку вирішує два завдання: якому вузлу необхідно збільшити продуктивність та яку кількість ресурсів додати. При чому, у випадку використання хмарної системи, під ресурсами слід розуміти віртуальні: контролери базових станцій (*Base station controllers*); комутатори (або центри комутації) мобільного зв'язку (*Mobile Switching Centers*), сервери баз даних профілів абонентів (*Home Location Registers/Home Subscriber Servers*) та інше обладнання або серверні платформи.

При укладенні договору з орендодавцем обговорюється кількість та порядок використання додаткового віртуального ресурсу. Занадто велика кількість доданих (орендованих) ресурсів буде неефективною і призведе до значних затрат, а занадто мала негативно відобразиться на якості обслуговування абонентів.

В свою чергу, задача блоку керування – знайти оптимальну кількість додаткових віртуальних ресурсів, яка дозволить забезпечити необхідне співвідношення між витратами та якістю обслуговування, зазначене в договорі [3].

Проблема визначення додаткової кількості ресурсів особливо актуальна для „безшовних технологій”, які використовуються в місцях великого скупчення людей. Таким чином, постає задача територіального розподілу ресурсів для задоволення інформаційних потреб абонентів, які швидко змінюються та залежать від типу необхідного сервісу.

*Аналіз останніх публікацій* вказує на той факт [4 – 6], що сьогодні оператор використовує лише статистичну інформацію про інтенсивність надходження запитів, територіальний розподіл абонентів та не враховує нестационарність абонентської активності, що в свою чергу не дозволяє корегувати та забезпечити необхідну якість обслуговування при пікових коливаннях навантаження. Альтернативою являється динамічна зміна кількості замовлених віртуальних ресурсів, в залежності від активності абонентів, що дозволить забезпечити необхідну якість обслуговування та оптимізувати витрати оператора.

*Тому метою статті* є вирішення задачі прогнозування кількості додаткових ресурсів на основі даних про поведінку абонентів протягом проміжку часу  $\Delta T$  та заданої якості обслуговування. На основі отриманого результату можна зробити висновок про необхідність додавання або згортання ресурсів обслуговування для уникнення надлишковості і забезпечення заданих параметрів  $QoS$ .

### **Викладення основного матеріалу**

Один із методів оптимізації гетерогенної мережі передбачає прогнозування необхідного віртуального ресурсу у хмарі для забезпечення ефективної роботи мережі мобільного оператора зв'язку. Основний принцип полягає у використанні даних системи моніторингу для визначення необхідних конфігурацій системи обслуговування оператора мобільного зв'язку та розкладу залучення додаткових віртуальних ресурсів [7] із метою забезпечення заданих показників  $QoS$ . При цьому потрібно періодично контролювати достатність обраної конфігурації. Тобто, на основі розрахованих конфігурацій та статистичних даних системи моніторингу розробити розклад використання ресурсів та періодично перевіряти достатність ресурсів, як фізичних так і віртуальних (рис. 1). Під конфігурацією системи обслуговування розуміється комбінація залучених віртуальних ресурсів в залежності від їх функціональності та кількості для забезпечення необхідних показників  $QoS$  в поточний момент часу.

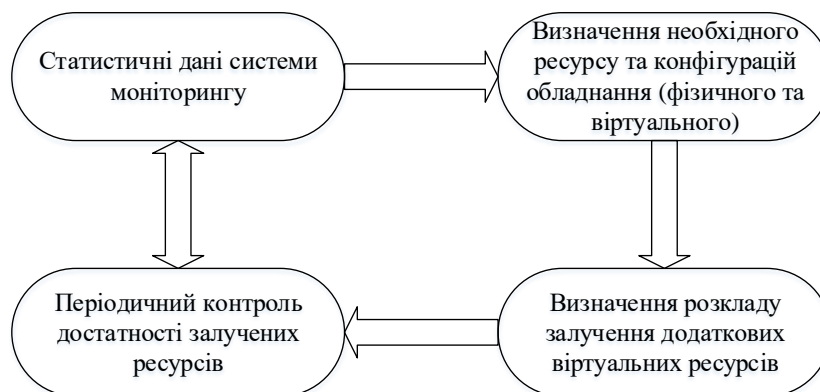


Рис. 1. Схема процесу прогнозування та контролю необхідного ресурсу системи обслуговування

Основними факторами, які впливають на прогноз, є нестационарність абонентської активності протягом доби та можливості гнучкої конфігурації ресурсів при використанні хмарного обчислювального середовища.

Розглянемо два основних методи побудови розкладу та вибору конфігурацій системи обслуговування оператора мобільного зв'язку в залежності від активності абонентів.

### *Метод побудови розкладу вибору конфігурацій на основі довгострокової статистики*

Метод побудови розкладу вибору конфігурацій на основі довгострокової статистики базується на великій кількості статистичних даних, які збираються протягом тривалого часу (наприклад, протягом року). Ці дані групуються по різних днях тижня. Аналізується

довгострокова статистика навантаження, що створюється групою вузлів доступу, якщо даний метод застосовується для планування конфігурацій груп вузлів обслуговування, або навантаження на вузол обслуговування, якщо метод застосовується для планування конфігурацій розподілу ресурсів вузлів [8 – 9].

Метод вирішує завдання визначення порядку застосування конфігурацій ресурсів обслуговування, а саме моментів їх переключення (рис. 2.). Основна наукова ідея методу полягає у ітераційному визначенні конфігурацій для різних інтервалів часу та перевірки нерівності „Чебишива” для забезпечення вимог щодо достатності ресурсів обслуговування протягом всієї часової прямої [10].

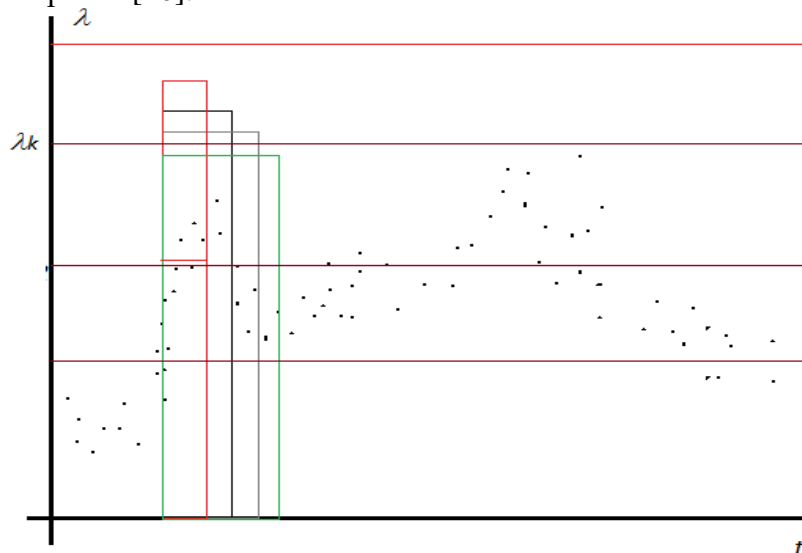


Рис. 2. Метод побудови розкладу вибору конфігурацій на основі довгострокової статистики

*Вхідні дані:*

–  $\lambda_k$  – максимальна інтенсивність обслуговування (вузла або групи вузлів), яка відповідає  $k$ -й конфігурації;

–  $\vec{\lambda}_t$  – вектори статистичних значень кількості заявок,  $t \in [0, T]$ ;

–  $\vec{\lambda}_t = (\lambda_t^1, \lambda_t^2, \dots, \lambda_t^{50})$  – це набір значень кількості заявок, які надійшли у  $i$ -й момент часу, за останні 50 тижнів.

*Вихідні дані:*

–  $(k, t)$ , де  $t \in [0, T]$  – рекомендовані моменти зміни інтенсивності обслуговування,  $k$  – рекомендована конфігурація системи, на яку слід перейти в момент часу  $t$ .

*Алгоритм методу:*

*Етап 1.* Розбити часову вісь на інтервали часу тривалістю  $\Delta t$ ;

*Етап 2.* Для кожного інтервалу  $\forall i = \overline{1, \frac{24}{\Delta t}}$  знайти середнє значення  $M\lambda$  та дисперсію  $\sigma^2$ , для аналізу береться множина даних зібраних за тривалий час обслуговування;

*Етап 3.* для кожного інтервалу знайти мінімальну конфігурацію, для якої виконується нерівність:

$$M\lambda + 3\sigma^2 \leq \lambda_k.$$

Створити часову множину  $Kt = \{Kt_i\}$ ,  $Kt_i = (k, i)$ ,  $i = \overline{1, T}$ ;

*Етап 4.* Створити постійну множину  $Kt'$ , якщо  $Kt' = \emptyset$ , то  $Kt' := Kt$ , перехід до Етапу 5; інакше попарно порівняти елементи множин  $Kt'$  і  $Kt$ , якщо  $\forall Kt'_i$  виконується  $Kt'_i = Kt_i$ , то перехід до Етапу 6, інакше  $Kt'_i = \max(Kt'_i, Kt_i)$ , перехід до Етапу 5;

*Етап 5.* Змінити значення  $\Delta t$ , перейти на Етап 2;

*Етап 6.* На основі  $Kt'_i$  визначити  $(k, t)$ , де  $t \in [0, T]$  – рекомендовані моменти зміни інтенсивності обслуговування,  $k$  – рекомендована конфігурація системи, на яку слід перейти в момент часу  $t$ .

### Метод короткострокового прогнозування навантаження

Метод короткострокового планування навантаження (КПН) є вдосконаленим методом прогнозування *ARIMA* (*Autoregressive integrated moving average*) – авторегресійний метод із ковзковим математичним очікуванням. Однак на відміну від відомого методу, пропонується розв'язати задачу пошуку мінімального ковзкового інтервалу, використання якого буде задовольняти вимогам, що дозволить мінімізувати кількість операцій із плаваючою точкою для виконання передбачення, що забезпечить оптимальну швидкість виконання передбачення.

Запропонований метод складається з двох етапів – розрахунок інтервалу прогнозування на основі статистики роботи даного вузлу обслуговування і безпосередньо періодичне прогнозування навантаження та контроль достатності ресурсів.

*Вхідні дані:*

- $T_{\Pi}$  – інтервал часу для якого потрібен прогноз;
- $\lambda_i$  – кількість заявок за 1мс, ( $i \in 0, \dots, N$ );
- $N = T_{\text{інф}}/1\text{мс}$ ,  $\lambda_i \in \Lambda$ ,  $\Lambda$  – множина значень статистики кількості заявок, що надходили протягом часу  $T_{\text{інф}}$  (спочатку задається, потім корегується на 2-му етапі методу) до початку здійснення прогнозу,  $|\Lambda| = N$ ;
- $T_{\text{прог}}$  – період прогнозування, час по завершенню якого, запускається алгоритм прогнозування;
- $M$  – гранична кількість заявок, яка може бути обслужена при заданій конфігурації обслуговуючого пристрою;
- $P$  – ймовірність помилки прогнозування.

Короткострокова статистика збирається локально на обслуговуючому пристрої та зберігається не довше  $T_{\text{інф}} + T_{\text{прог}}$ , інтервал дискретизації часу 1мс.

*Вихідні дані:*

- $z \in \{0, 1\}$  –  $z=0$  не змінювати конфігурацію;  $z=1$  змінити конфігурацію.

*Алгоритм методу:*

*Підготовчий етап.* Навчання системи на основі статистичних даних. Пошук мінімального  $T_{\text{інф}}$  (інтервал часу збирання інформації)

$$T_{\text{інф}} \rightarrow \min$$

для якого виконується обмеження:  $\bar{\lambda}_{T_{\text{прог}}} + 3\sigma > M$ ,

де  $\bar{\lambda}_{T_{\text{прог}}}$  і  $\sigma$  – розраховуються відповідно до основного етапу.

Обмеження виконується для  $p=P \cdot 100\%$  статистичних вибірок отриманих у різні інтервали часу.

Спосіб розв'язку: перевірка значень для послідовності утвореної за принципом  $T_{\text{інф}}^{k+1} = T_{\text{інф}}^k + \Delta$ ;  $T_{\text{інф}}^0 = T_{\text{прог}}$ .

*Основний етап динамічного контролю.*

1. Аналіз статистичних даних  $\lambda_i$  за інтервал  $T_{\text{інф}}$  часу що передує моменту розрахунку. Побудова за методом найменших квадратів оцінок коефіцієнтів  $\hat{a}$  та  $\hat{b}$

$$\lambda = \hat{a}i + \hat{b}.$$

2. Розрахувати  $\bar{\lambda}_{T_{\text{прог}}} = \hat{a}T_{\text{прог}} + \hat{b}$ ;

3. Якщо  $\bar{\lambda}_{T_{\text{прог}}} + 3\sigma \leq M$ , тоді  $z=0$ , інакше  $z=1$ .

Принцип планування ресурсів гетерогенного телекомунікаційного середовища базується на аналізі великої кількості статистичних даних. Він дозволяє розрахувати конфігурації обслуговування та план переключення між ними у процесі експлуатації системи керування.

Основними етапами при плануванні ресурсів гетерогенного телекомунікаційного середовища є:

- аналіз даних;
- розрахунок конфігурацій;
- розрахунок плану застосування конфігурацій, які забезпечують ефективність використання ресурсів *NFV (Network Functions Virtualization)* [11 – 12];
- динамічний контроль достатності ресурсів, відповідно до короткострокової статистики.

Аналіз даних передбачає побудову моделей потоків даних для кожної точки доступу, моделі статистики навантаження на кожний вузол обслуговування (як віртуальний так і фізичний), поточний контроль потоків в мережі (рис. 3.).

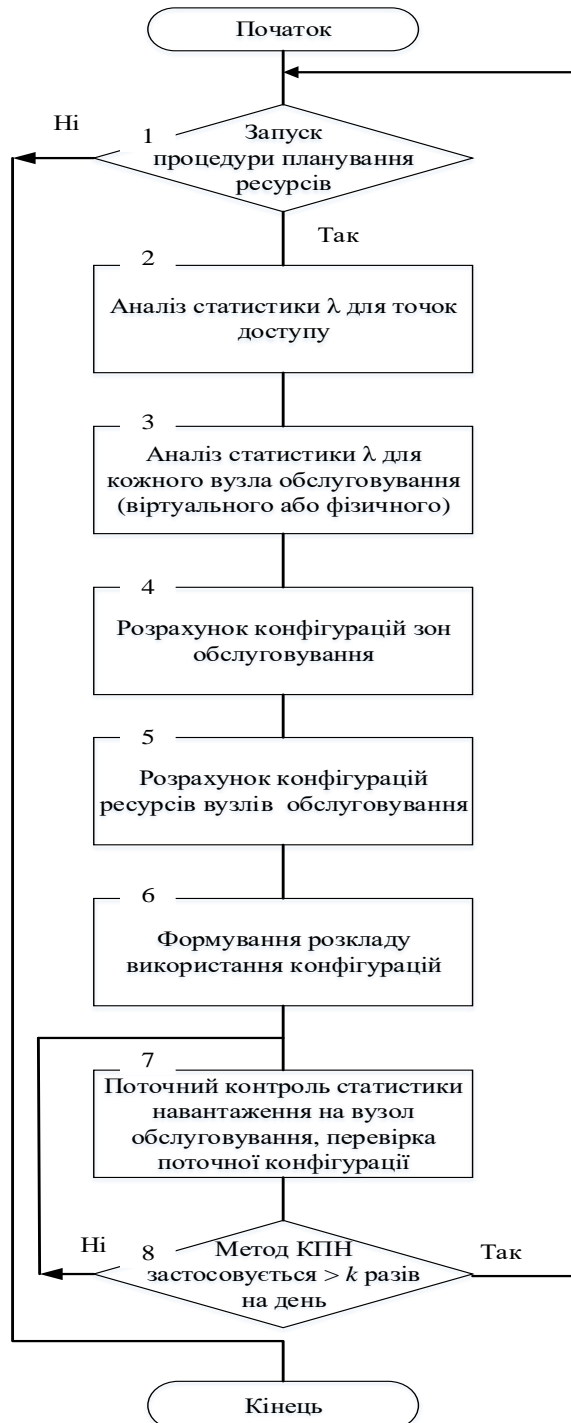


Рис. 3. Алгоритм планування ресурсів гетерогенного телекомунікаційного середовища на базі аналізу статистичних даних

Розрахунок конфігурацій складається з розрахунку зон обслуговування та розрахунку ресурсів обслуговування, відповідно до статистики навантажень на вузли.

Розрахунок конфігурацій зон обслуговування передбачає визначення груп точок доступу по яким є статистика навантаження, та визначення необхідної конфігурації вузлів обслуговування віртуалізованої (або фізичної) системи ядра оператора зв'язку.

Далі за методом побудови розкладу на основі довгострокової статистики визначається графік використання розрахованих конфігурацій системи.

Для забезпечення контролю за перевантаженням у вузлах обслуговування здійснюється моніторинг достатності ресурсів на наступний період часу на основі короткострокової статистики.

На рис. 4. показані основні процедури принципу планування ресурсів. На схемі можна бачити, які системи задіяні для реалізації контролю та планування ресурсів.

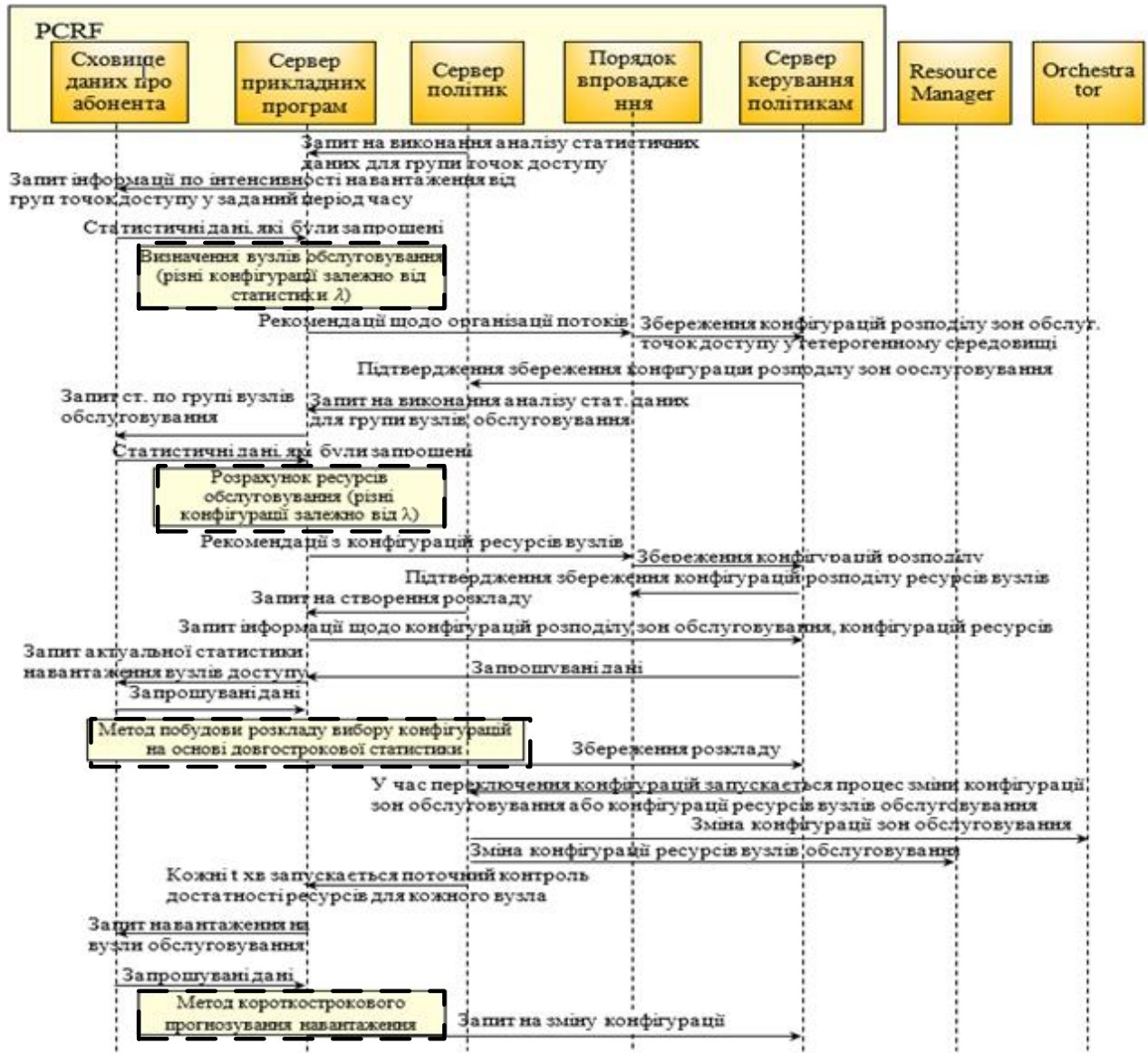


Рис.4. Процедура планування обсягу ресурсів гетерогенного телекомунікаційного середовища

Основні операції виконуються у системі *PCRF* (*Policy and Charging Rules Function*) під час взаємодії між її підсистемами, а саме між підсистемами „Сховище даних про абонента”, „Сервер прикладних програм”, „Сервер політик”, „Порядок впровадження політик”, а також системою керування гетерогенним середовищем – „Ресурс менеджер” (*Resource Manager*) та „Оркестратор потоків” (*Orchestrator*). Для збору даних про віртуальні вузли також

використовується „Моніторинг віртуальних сутностей” – система, яка також відноситься до системи керування гетерогенним середовищем, але не включається в опис процедури. Передбачається, що довгострокова статистика віртуальних вузлів зберігається у „Сховищі даних про абонента”.

#### **Висновок**

Запропоновано принцип прогнозування необхідного віртуального ресурсу у хмарній системі для забезпечення ефективної роботи гетерогенного телекомунікаційного середовища оператора мобільного зв'язку. Він полягає у використанні статистичних даних системи моніторингу для розрахунку мінімально-необхідних конфігурацій із залученням віртуальних ресурсів системи обслуговування та розкладу їх використання із метою забезпечення показників якості обслуговування абонентів, періодичного контролю достатності обраної конфігурації. Застосування даного принципу дозволить забезпечити гнучкість та масштабованість системи керування процесом обслуговування гібридних телекомунікаційних сервісів у гетерогенному телекомунікаційному середовищі, а також ефективно використовувати наявні фізичні та орендовані віртуальні ресурси.

В подальшому планується адаптація даного принципу для розширення надання мережевих сервісів та мережевих функцій для операторів фіксованих мереж.

#### ЛІТЕРАТУРА

1. Романов О.І., Нестеренко М.М., Гордашник Є.С. Аналіз функціональних особливостей побудови IP-мереж на базі Softswitch // Збірник наукових праць ВІТІ № 1 – 2015, С.69 – 80.
2. Романов О.І., Діденко І.В. Особливості взаємодії контролера і мережевих пристроїв в мережах SDN з використанням протоколу OPENFLOW // Матеріали одинадцятої МНТК „Проблеми телекомунікацій”, Київ, 2017 р. С. 142 – 145.
3. ITU-T M.3371 (10/2016) Требования к управлению услугами в системе управления электросвязью, совместимой с облаком, available: <https://www.itu.int/ITU-T/recommendations/rec>.
4. ITU-T Y.2012 (04/2010) Функциональные требования и архитектура сетей последующих поколений, available: <https://www.itu.int/ITU-T/recommendations/rec.aspx>
5. ITU-T Y.3502 Information technology – Cloud computing – Reference architecture, available: <https://www.itu.int/rec/T-REC-Y.3502/en>
6. ETSI GS NFV 003: „Network Functions Virtualization (NFV); Terminology for Main Concepts in NFV” 5 Cloud computing framework for end to end resource management Recommendation ITU-T Y.3520, available: <https://www.itu.int/rec/T-RECY.3520/en>
7. M.A. Skulysh, O.I. Romanov, “The structure of a mobile provider network with network functions virtualization”, Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), 2018 14th International Conference on, IEEE, Pages: 1032-1034, 2018.
8. Y.3500 : Information technology – Cloud computing- Overviewandvocabulary, available: <https://www.itu.int/rec/T-REC-Y.3500-201408-I/en>
9. M.A. Skulysh, “The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters”, Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), IEEE Conference Publications, Pages: 1–4, 2017.
10. Медведєв М.Г., Пашенко І.О. Теорія ймовірностей та математична статистика. Підручник – К.: Видавництво „Ліра-К”. 2017. – 536 с.
11. L.S. Globa, M.A. Skulysh, „Planning the loading of datacenters resources based on down load statistics”, Вісник Національного технічного університету України Київський політехнічний інститут. Серія: Радіотехніка. Радіоапаратобудування, №. 65, 2016.
12. ITU-T Y.2011: General principles and general reference model for Next Generation Networks, available: <https://www.itu.int/rec/T-REC-Y.2011-200410-I/en>