

МЕТОД ВИЗНАЧЕННЯ ПРИВ'ЯЗКИ ДО ВУЗЛІВ АГРЕГАЦІЇ ТРАФІКУ ВІРТУАЛІЗОВАНИХ СЕРВІСІВ

Суліма С.В.

Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна

E-mail: itssulima@gmail.com

Method of assignment for traffic aggregation points of virtualized services

The basic trends of mobile services providing are summarized. The method of determining the assignment of base stations to traffic aggregation points and their number in virtualized networks is proposed, which allows solving the problem of calculating the topology of EPC as a service, while taking into account the cost of organizing the service environment with minimizing delay and rational use of channels throughput capacity.

Зі зростанням об'єму трафіку мобільних даних та числа сервісів, які надаються мобільною мережею, зростають і потоки службового навантаження, що викликає необхідність у зміні принципів, моделей та методів організації обслуговування телекомунікаційних потоків з метою забезпечення заданої якості надання широкого кола послуг гнучким та економічно ефективним способом. Наявність потужних дата центрів значно розширює можливості організації процесу надання послуг на сучасному телекомунікаційному ринку. Переваг масштабованості та еластичності з підтримкою високого показника коефіцієнту використання технічних ресурсів мережі можна досягти з використанням технологій віртуалізації.

Одним з ключових аспектів у області віртуалізації мережі є розподіл фізичних ресурсів для віртуальних функцій мережі, що передбачає відображення віртуальних мереж на фізичні мережі. Оптимальність і гнучкість розподілу ресурсів є основними факторами для успішної віртуалізації.

Нехай визначення місця розміщення та необхідної ємності віртуальних зарезервованих обчислювальних ресурсів у разі виникнення перевантаження фізичної мережі здійснюється відповідно до [1]. Доповнимо цей метод.

Кількість сервісних ланцюгів віртуальних мереж, що будуть відображатися на хмарну інфраструктуру потрібно визначити заздалегідь. Крайнім випадком був б розгляд одного сервісного ланцюга на стільник/eNodeB. Оскільки реалістичні сценарії для мобільних мереж складають 10000 eNodeB, результуюча оптимізаційна модель буде величезна, і для її вирішення потрібен досить тривалий час обчислення. Тому приймаємо обґрунтовано великі кластери eNodeBs і припустимо, що кожен з цих кластерів eNodeB звертається до одного сервісного ланцюга базової мережі.

Більшість сучасних досліджень не розглядає задачу визначення кількості мереж LTE EPC як сервісів, а передбачається їх наперед задана кількість. Найбільш близькою виступає задача з дослідження [2], яка, проте, направлена на визначення розміщення та прив'язки RRH сайтів до BBU серверів у мережах

з застосуваннях технології NFV і не враховує обмежень затримки та пропускної здатності.

Далі описано алгоритм вибору вузлів агрегації трафіку. Розглядаємо випадок, коли провайдер телекомунікаційних послуг вже має існуючу топологію базових станцій. Потрібно визначити підмножину мережевих вузлів, де будуть розміщені блоки агрегації навантаження, які будуть формувати запити до одного віртуалізованого сервісу ЕРС. Після цього для кожного сайту базових станцій призначаємо вузол агрегації (Traffic Aggregation Point – ТАР).

Нехай x_i – це бінарна змінна рішення, яка приймає значення 1, якщо в точці i необхідно розмістити ТАР, та 0 – в іншому випадку. Крім того, визначаємо y_{ji} як бінарну змінну, яка приймає значення 1, якщо базова станція j направляє навантаження в i -у ТАР, і 0 – в іншому випадку. Необхідно визначити значення x_i та y_{ji} , так щоб знайти оптимальне значення цільової функції.

Цільова функція (1) прагне мінімізувати мережеві затримки. Цільова функція (2) представляє загальну умовну вартість встановлення вузлів агрегації трафіку та вартість встановлення каналів між базовими станціями та відповідними ТАР, що їх обслуговують. Цільова функція (3) має на меті залишати більше вільної пропускної здатності на кожному фізичному каналі. Максимізується залишкова пропускна здатність по всіх каналах, оскільки значно завантажені канали можуть призвести до перевантажень у мережі, тому бажано отримати рішення, де залишається більше потужностей каналів.

Ці цілі оптимізації можуть бути корисними для мережевих операторів для планування найкращої стратегії розгортання.

$$\min_{x_i, y_{ji}} \left(\sum_i \sum_j y_{ji} \cdot L_{ji} \right), \quad (1)$$

$$\min_{x_i, y_{ji}} \left(\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot costl_{ji} \right), \quad (2)$$

$$\max_{x_i, y_{ji}} \left(\sum_i y_{ji} \cdot \left(c_{ji} - B_{ji} \right) \right), \quad (3)$$

де L_{ji} – затримка каналу зв'язку між сайтом j і ТАР i ;

$cost_i$ – вартість, яка складається з двох частин: фіксованої початкової вартості f_i , яка відповідає за фіксовані інвестиції, такі як простір та встановлення обладнання, а також додаткові витрати $costN_i$ на одиницю обробної потужності, встановленої на обчислювальному вузлі, де d_i – обсяг задіяних обчислювальних ресурсів обробки: $cost_i = f_i + costN_i d_i$;

$costl_{ji}$ – вартість встановлення зв'язку між сайтом j і ТАР i , визначається як лінійна комбінація початкової фіксованої вартості fl_{ji} та змінної частини залежної від смуги пропускання B_{ji} , необхідної каналу, та вартості одиниці пропускної здатності $costL_j$: $costl_{ji} = fl_{ji} + costL_j B_{ji}$;

c_{ji} – доступна пропускна здатність.

Можливо використовувати лінійну комбінацію (4) виразів (1)-(3) з ваговими коефіцієнтами a , b , c , які можуть використовуватись не лише щоб задати більшу вагомість тої чи іншої складової, але також щоб масштабувати значення виразів з метою зведення до порівнюваних значень та мати значуще складення:

$$\min_{x_i, y_{ji}} \left(a \cdot \sum_i \sum_j y_{ji} \cdot L_{ji} + b \cdot \left(\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot cost_{jl_{ji}} \right) - c \cdot \left(\sum_i y_{ji} \cdot (c_{ji} - B_{ji}) \right) \right) \quad (4)$$

Обмеження:

$$\sum_i y_{ji} = 1 \quad \forall j, \quad (5)$$

$$y_{ji} \leq x_i \quad \forall j \forall i, \quad (6)$$

$$\sum_i x_i \leq p, \quad (7)$$

$$\sum_j y_{ji} \cdot d_j \leq p_i \quad \forall i, \quad (8)$$

$$\sum_i y_{ji} \cdot (c_{ji} - B_{ji}) \geq 0 \quad \forall j, \quad (9)$$

$$\sum_i y_{ji} \cdot L_{ji} \leq T_j \quad \forall j. \quad (10)$$

Обмеження (5) гарантує, що кожна базова станція буде приєднана до одного ТАР. Обмеження (6) гарантує, що між сайтом базової станції j та ТАР i створюється канал, тільки якщо i було розміщено.

Обмеження (7) гарантує, що максимальна кількість ТАР не перевищує бюджет p , тоді як (8) є обмеженням потужності, що гарантує, що загальні вимоги до обробки всіх базових станцій, призначених для конкретного ТАР, не перевищують фактичні встановлені фізичні ресурси p_i . Обмеження (9) гарантує достатність каналних ресурсів для встановлення каналів, а (10) – допустимість значення затримки, тобто не перевищення порогу T_j .

Окреслені перспективні напрямки досліджень у сфері надання послуг мобільного зв'язку. Запропоновано метод визначення прив'язки базових станцій до вузлів агрегації трафіку та їх кількості у віртуалізованих мережах, який дозволяє вирішити задачу розрахунку топології ЕРС як сервісів, при цьому враховуючи вартість організації обслуговуючого середовища з мінімізацією затримки та раціональним використанням пропускної здатності.

Література

1. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6–9 June 2016 : conference proceedings. — Varna, Bulgaria, 2016. — P. 1–5.
2. Mijumbi R. Server placement and assignment in virtualized radio access networks / R. Mijumbi, J. Serrat, J.-L. Gorricho, J. Rubio-Loyola, S. Davy // 2015 11th International Conference on Network and Service Management (CNSM). – Barcelona, 2015. – P. 398-401.