

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут телекомунікаційних систем
Кафедра Телекомунікаційних систем**

«На правах рукопису»
УДК _____

«До захисту допущено»

Завідувач кафедри

_____ Л.О. Уривський

«__» _____ 20__ р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 172 Телекомунікації та радіотехніка

**на тему: «Дослідження принципів взаємодії віртуальних компонентів
мобільної мережі з існуючим телекомунікаційним обладнанням»**

Виконала:

студентка II курсу, групи ТС-71мп

Бабич Анастасія Олександрівна _____

Керівник:

доцент кафедри ТК, к.т.н., с.н.с.

Міночкін Д.А. _____

Рецензент:

професор кафедри ТК, д.т.н., проф.

Романов О.І. _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних
посилань.

Студентка _____

Київ – 2018

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут телекомунікаційних систем
Кафедра Телекомунікаційних систем

Рівень вищої освіти – другий (магістерський) за освітньо-професійною програмою

Спеціальність (спеціалізація) – 172 «Телекомунікації та радіотехніка»
 (172.3620.1 «Телекомунікаційні системи та мережі»)

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Л.О. Уривський

«__» _____ 20__ р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Бабич Анастасії Олександрівні

1. Тема дисертації «Дослідження принципів взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням», науковий керівник дисертації доцент кафедри ТК, к.т.н., с.н.с. Міночкін Дмитро Анатолійович, затверджені наказом по університету від «__» _____ 20__ р. № _____

2. Термін подання студентом дисертації _____

3. Об'єкт дослідження

4. Предмет дослідження

5. Перелік завдань, які потрібно розробити

6. Орієнтовний перелік графічного (ілюстративного) матеріалу

7. Орієнтовний перелік публікацій

8. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Пошук джерел інформації та їх	09.2017-12.2017	виконано

	систематизація		
2	Огляд технології віртуалізації	12.2017-03.2018	виконано
3	Аналіз принципів побудови віртуалізованої мережі	03.2018-05.2018	виконано
4	Дослідження ефективності використання віртуалізованих компонентів мобільної мережі з існуючим телекомунікаційним обладнанням	05.2018-06.2018	виконано
5	Розробка методики, здатної масштабуватися з кількома цілями і великими екземплярами	06.2018-09.2018	виконано
6	Підведення підсумків	10.2018-11.2018	виконано
7	Оформлення магістерської роботи	11.2018-12.2018	виконано

Студент

Бабич А.О.

Науковий керівник дисертації

Міночкін Д.А.

РЕФЕРАТ

Темою магістерської дисертації є дослідження принципів взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням.

Робота містить 107 сторінок, зокрема 32 ілюстрацій, 1 таблицю та 20 джерел інформації.

Тема магістерської дисертації є актуальною, адже створення високоякісних та високошвидкісних систем зв'язку є актуальним питанням сьогодення. Використання NFV є досить перспективним, бо по-перше, підвищується продуктивність площини користувача, а, по-друге, допомагає операторам створювати абсолютно нові мережеві архітектури та бізнес-моделі, відкриваючи нові можливості для ринку і досягнення успішних цифрових перетворень.

Мета дисертації полягає в підвищенні ефективності функціонування телекомунікаційного обладнання для надійного забезпечення високошвидкісного і високоякісного зв'язку шляхом використання віртуальних компонентів мобільної мережі.

Об'єктом дослідження є технології віртуалізації. Предметом дослідження є принципи взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням.

При виконанні роботи застосовувалося моделювання, що допомагає мінімізувати максимальне використання мережевих посилань та мінімізувати кількість VNF посилань.

У дисертації був запропонований математичний евристичний підхід, що дозволяє також запускати експерименти для великих випадків проблеми протягом прийняттого часу виконання. Було якісно і кількісно порівняно підхід VNF-PR до існуючого підходу Virtual Embedding (VNE).

Ключові слова: ВІРТУАЛІЗАЦІЯ, МЕРЕЖЕВА АРХІТЕКТУРА, NFV, XMAPA, 5G, LTE, 3GPP, LAA, NFVI, VNF

ABSTRACT

The theme of the master's thesis is to study the principles of interaction of virtual components of the mobile network with existing telecommunication equipment.

The work contains 107 pages, 32 figures and 1 table. 20 sources have been utilized.

The topic of the master's thesis is relevant, since the creation of high-quality and high-speed communication systems is a topical issue of the present. Using NFV is quite promising, because, firstly, the productivity of the user plane increases, and secondly, it helps operators create completely new network architectures and business models, opening up new opportunities for the market and achieving successful digital transformations.

The purpose of the thesis is to increase the efficiency of the functioning of telecommunication equipment for the reliable provision of high-speed and high-quality communication through the use of virtual components of the mobile network.

The object of research is virtualization technology. The subject of research is the principles of interaction of virtual components of the mobile network with existing telecommunication equipment.

During the work was used simulation, which helps to minimize the maximum use of network links and minimize the number of VNF links.

A mathematical heuristic approach was suggested in the dissertation, which also allows to run experiments for large instances of problem within an acceptable execution time. Was compare the VNF-PR approach to the legacy Virtual Network Embedding (VNE) approach, qualitatively and quantitatively.

Keywords: VIRTUALIZATION, NETWORK ARCHITECTURE, NFV, CLOUD, 5G, LTE, 3GPP, LAA, NFVI, VNF

ЗМІСТ

ЗМІСТ	1
ПЕРЕЛІК СКОРОЧЕНЬ.....	11
ВСТУП	11
РОЗДІЛ 1 ДОСЛІДЖЕННЯ РОЗВИТКУ СТАНДАРТУ МОБІЛЬНОГО ЗВ'ЯЗКУ LTE	13
1.1 Історія та розвиток LTE	13
1.2 Загальна характеристика 5G.....	14
1.3 Вплив FD-MIMO і неліцензійних операцій на підвищення швидкості передачі даних користувача і можливостей системи.....	17
1.3.1 FD-MIMO	17
1.3.2 Операції LTE в неліцензійному спектрі	19
1.4 Вплив розвитку LTE на скорочення затримки	22
1.5 Варіанти зменшення сигналізації у версіях Rel-13, Rel-14 та	24
Rel-15.....	24
1.6 Ряд удосконалень в LTE Rel-14 і Rel-15 для 5G.....	25
1.6.1 Масивні зв'язки типу машини	26
1.6.2 Критичне повідомлення.....	28
1.6.3 Інтелектуальні транспортні системи	29
1.7 Використання віртуалізації в системах 5G	31
1.8 Технології віртуалізації в 5G.....	32
1.9 Підняття питання QoS в системах 5G	33
1.10 Реакція постачальників мереж на використання 5G.....	34
1.11 Приклад використання мобільної хмарної мережі	35
1.12 Висновки з розділу 1	38
РОЗДІЛ 2 ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ.....	40
2.1 Віртуалізована архітектура та її види.....	40
2.2 Історія NFV. Довідкова архітектура ETSI та процес віртуалізації.....	41
2.3 Інфраструктура NFV (NFVI)	44
2.3.1 NFVI - Обладнання (дата-центр, edge і vCPE)	46
2.3.2 NFVI - рівень віртуалізації	51
2.3.3 NFVI - Віртуальна інфраструктура	52

2.3.4 NFVI - Майбутні тенденції.....	54
2.4 Висновки з розділу 2	60
РОЗДІЛ 3 АНАЛІЗ ВПРОВАДЖЕННЯ ТЕХНОЛОГІЙ ВІРТУАЛІЗАЦІЇ В ОПОРНІЙ МЕРЕЖІ ОПЕРАТОРІВ ЗВ'ЯЗКУ	62
3.1 Ericsson Virtual Evolved Packet Core	62
3.2 Huawei CloudCore	63
3.2.1 Service Oriented Core (SOC).....	64
3.2.2 SOC 2.0 будівництво 5G Core Networks.....	67
3.3 Впровадження 5G	68
3.4 Висновки з розділу 3	70
РОЗДІЛ 4 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВІРТУАЛЬНОЇ МЕРЕЖЕВОЇ ФУНКЦІЇ	72
4.1 Оптимальна організація віртуальних мережевих функцій	72
4.2 Сучасний стан оркестровки NFV	75
4.3 Мережева модель.....	78
4.3.1 Опис проблеми	78
4.3.2 Математичне формулювання.....	80
4.3.3 Багатоцільове математичне евристичне рішення	90
4.3.4 Подальші уточнення моделі.....	92
4.4 Результати математичного моделювання	97
4.4.1 Загальні результати	100
4.4.2 Підходи, засновані на VNF-PR і VNE.....	107
4.5 Висновки з розділу 4	110
ВИСНОВКИ.....	112
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	114

ПЕРЕЛІК СКОРОЧЕНЬ

- AS (access stratum) - прошарок доступу
- BS (base station) - базова станція
- CA (carrier aggregation) - об'єднання несучих
- CN (core network) - базова мережа
- CP (control plane) - площина управління
- CSI (channel state information) - інформація про стан каналу
- CSI-RS (CSI reference signal) - опорний сигнал CSI
- D2D (device-to-device) - технологія пристрій до пристрою
- DC (dual connectivity) - подвійне підключення
- DL (downlink) - низхідна лінія зв'язку
- DoNAS (data over non-access stratum) - дані по недоступному рівню
- DSRC (dedicated short range communications) - система передачі даних на короткі відстані (до 30м) при наявності прямої видимості
- eMBB (enhanced mobile broadband) - розширений мобільний широкопasmовий доступ
- eMTC (enhanced MTC) - покращений MTC
- eNB (evolved node B) - вдосконалений вузол B
- FD-MIMO (hybrid automatic repeat request MIMO) - повний розмір MIMO
- HARQ (hybrid automatic repeat request) - гібридний автоматичний запит повтору
- IoT (Internet of Things) - інтернет речей
- ITS (intelligent transportation system) - інтелектуальна транспортна система
- ITU (International Telecommunication Union) - Міжнародний союз електрозв'язку
- LAA (Licensed Assisted Access) - ліцензійний доступ до послуг
- MBMS (Multimedia Broadcast/Multicast Service) – мультимедійна широкопasmовна / багатоадресна служба
- MCL (maximum coupling loss) - максимальна втрата зчеплення

MIMO (multiple-input, multiple-output) - метод просторового кодування сигналу, що дозволяє збільшити смугу пропускання каналу, в якому передача даних і отримання даних здійснюються системами з декількох антен. Передавальні і приймальні антени розносять так, щоб кореляція між сусідніми антенами була слабкою

mMTC (massive machine type communications) - масова комунікація типу машини

mm-wave (millimeter wave) - міліметрова хвиля

MTC (machine type communications) - комунікація типу машини

MU-MIMO (multi-user MIMO) - розрахований на багато користувачів MIMO

NAS (non-access stratum) - прошарок без доступу

NB-IoT (Narrowband Internet of Things) - вузькосмуговий інтернет речей

NR (New Radio) - нове радіо

PCell (primary cell) - первинний елемент

RRC (Radio Resource Control) - управління радіоресурсами

RS (reference signal) - опорний сигнал

RTT (round-trip time) - час повернення

SCell (secondary cell) - вторинний елемент

SL (sidelink) - прямий інтерфейс D2D

SR (scheduling request) - запит на планування

TTI (transmission time interval) - часовий інтервал передачі

UL (uplink) - висхідна лінія зв'язку

UP (user plane) - площина користувача

URLLC (ultra-reliable low latency communications) - ультранадійні зв'язки з низьким часом очікування

V2I (vehicle-to-infrastructure) - "транспортний засіб до інфраструктури"

V2N (vehicle-to-network) - "транспортний засіб до мережі"

V2P (vehicle-to-pedestrian) - "транспортний засіб до пішоходу"

V2V (vehicle-to-vehicle) - "транспортний засіб-транспортний засіб"

V2X (vehicle-to-everything) - "транспорт до всього"

3GPP (3rd generation partnership project) - проект партнерства 3-го покоління

ВСТУП

Тема: Дослідження принципів взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням

Мета: Підвищення ефективності функціонування телекомунікаційного обладнання для надійного забезпечення високошвидкісного і високоякісного зв'язку шляхом використання віртуальних компонентів мобільної мережі.

Об'єкт дослідження: технології віртуалізації.

Предмет дослідження: принципи взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням.

Проблема, що вирішується – забезпечити користувачів мережею, що забезпечуватиме якісну та швидкісну передачу даних, обмін інформації.

Актуальність дослідження принципів взаємодії віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням є дуже важливим завданням, адже використання високоякісного та високошвидкісного зв'язку в сучасних умовах є актуальним питанням сьогодення. Використання NFV є досить перспективним. З одного боку підвищується продуктивність користувальницької площини, а з іншого боку-допомагає операторам створювати абсолютно нові мережеві архітектури та бізнес-моделі, відкриваючи нові можливості для ринку і досягнення успішних цифрових перетворень.

Наукові завдання:

1. Детальний аналіз ефективності роботи віртуальних компонентів.
2. Дослідження ефективності використання віртуальних компонентів мобільної мережі з існуючим телекомунікаційним обладнанням.
3. Забезпечення віртуалізованої інфраструктури, яка буде передавати голос та дані.
4. Розрахунок ефективності роботи віртуальних компонентів мобільної мережі.

Результат дипломної роботи: LTE цілком може відповідати всім

найбільш важливим вимогам 5G, включаючи підвищення швидкості передачі даних і підвищення пропускної спроможності системи за допомогою FD-MIMO, поліпшену підтримку неліцензованих операцій і скорочення затримки як в площині користувача, так і в сигналізації. Рекомендації щодо вибору технологій віртуалізації в опорній мережі операторів в'язку та висновки щодо доцільності використання цих технологій.

Практичним значенням результату є перехід до віртуалізованої мережі зв'язку та те, що результат може бути використаний для планування та моделювання телекомунікаційних безпроводових мереж.

РОЗДІЛ 1 ДОСЛІДЖЕННЯ РОЗВИТКУ СТАНДАРТУ МОБІЛЬНОГО ЗВ'ЯЗКУ LTE

1.1 Історія та розвиток LTE

З швидким розвитком досліджень 5G в 3GPP почався процес стандартизації. Будучи найпоширенішою технологією широкопasmового мобільного широкопasmового зв'язку в усьому світі, LTE являє собою невід'ємну частину головоломки 5G. Таким чином, його майбутні релізи (Rel-14 і Rel-15) призначені для задоволення якомога більшої кількості вимог 5G і розгляду відповідних варіантів використання, очікуваних в епоху 5G.

З моменту свого першого комерційного розгортання TeliaSonera в грудні 2009 року, LTE стала однією з найуспішніших технологій мобільного зв'язку в усьому світі. В даний час в 170 країнах розгорнуто 537 комерційних мереж LTE, 1,7 мільярда абонентів - число, яке, як очікується, зросте до приголомшуючого 4,6 мільярда до 2022 року.

За сім років, що минули з часу запуску LTE, значні успіхи були досягнуті з точки зору як продуктивності, так і універсальності. Наприклад, LTE Rel-8 представила смугу пропускання 20 МГц з максимальною швидкістю передачі даних по низхідній лінії зв'язку (DL) зі швидкістю передачі даних 300 Мбіт / с і швидкістю висхідній лінії зв'язку (UL) 75 Мбіт / с. Невеликі розширення були зроблені для Rel-9, таких як багатоадресні / ширококомовні послуги, послуги на основі розташування і дворівневе формування променя. LTE Rel-10, також відомий як LTE-Advanced, представив кілька нових функцій, таких як агрегування несучих (CA), щоб забезпечити пропускну здатність до 100 МГц, а також розширену підтримку мультиантени, гетерогенне розгортання і ретрансляцію. Ці функції дозволили забезпечити максимальну швидкість передачі даних понад 1 Гбіт / с в DL і 500 Мбіт / с в UL.

Rel-11 і Rel-12 включали такі удосконалення, як підтримка зв'язку типу машини (MTC), двухканальна взаємодія (DC), радіозв'язок з використанням

LTE-WLAN і послуги національної безпеки і громадської безпеки (NSPS), включаючи прямий зв'язок між пристроями (D2D). Подальші успіхи були досягнуті в Rel-13, включаючи вдосконалення спектральної ефективності за допомогою повнорозмірного множинного введення, множинного виведення (FD-MIMO), підтримка використання неліцензійного спектра з Licensed Assisted Access (LAA) і агрегацією LTE-WLAN, розширена підтримка для MTC через вузькосмугове інтернет-мовлення (NB-IoT) і розширений MTC (eMTC), розширений CA (до 32 несучих), поліпшені внутрішні позиціонування і Single Cell Point To Multipoint (SC-PTM) для ширококомовних /багатоадресних сервісів.

З жовтня 2015 року 3GPP використовує термін LTE-Advanced Pro для Rel-13 і далі, це означає, що LTE досяг рівня зрілості, який не тільки враховує розширені функціональні можливості / ефективність, але і підтримка нових варіантів використання.

1.2 Загальна характеристика 5G

Очікується, що в найближчі роки очікується зростання глобального мобільного трафіку даних зі змішаним річним показником в 45 відсотків, що в десять разів збільшиться в період між 2016 і 2022 роками. Це збільшення обумовлене головним чином масовим впровадженням потокової передачі відео. Крім того, IoT переходить від уявлення до реальності, а з 29 мільярдів підключених пристроїв, які передбачається включити до 2022 року, 18 мільярдів будуть пристроями IoT (або машинним). Майбутні мережі 5G повинні будуть підтримувати ці складні нові варіанти використання економічним і економічним способом [15].

Хоча вимоги до можливостей 5G все ще завершуються як в ITU, так і в 3GPP, є попередня угода щодо трьох основних варіантів використання, які повинна підтримувати технологія. Як показано на рис. 1.1, це: розширений мобільний широкосмуговий зв'язок (eMBB), наднадійний зв'язок з низькою

затримкою (URLLC) і масовою комунікація типу машини (mMTC). eMBB відноситься до розширеної підтримки звичайного MBW за допомогою поліпшених швидкостей передачі даних, пропускну здатності та покриття. URLLC - це вимога для нових критично важливих додатків, таких як промисловий інтернет, інтелектуальні мережі, захист інфраструктури, віддалена хірургія та інтелектуальні транспортні системи (ІТС). Нарешті, але не в останню чергу, mMTC необхідний для підтримки передбачуваного сценарію 5G IoT з десятками мільярдів підключених пристроїв і датчиків.

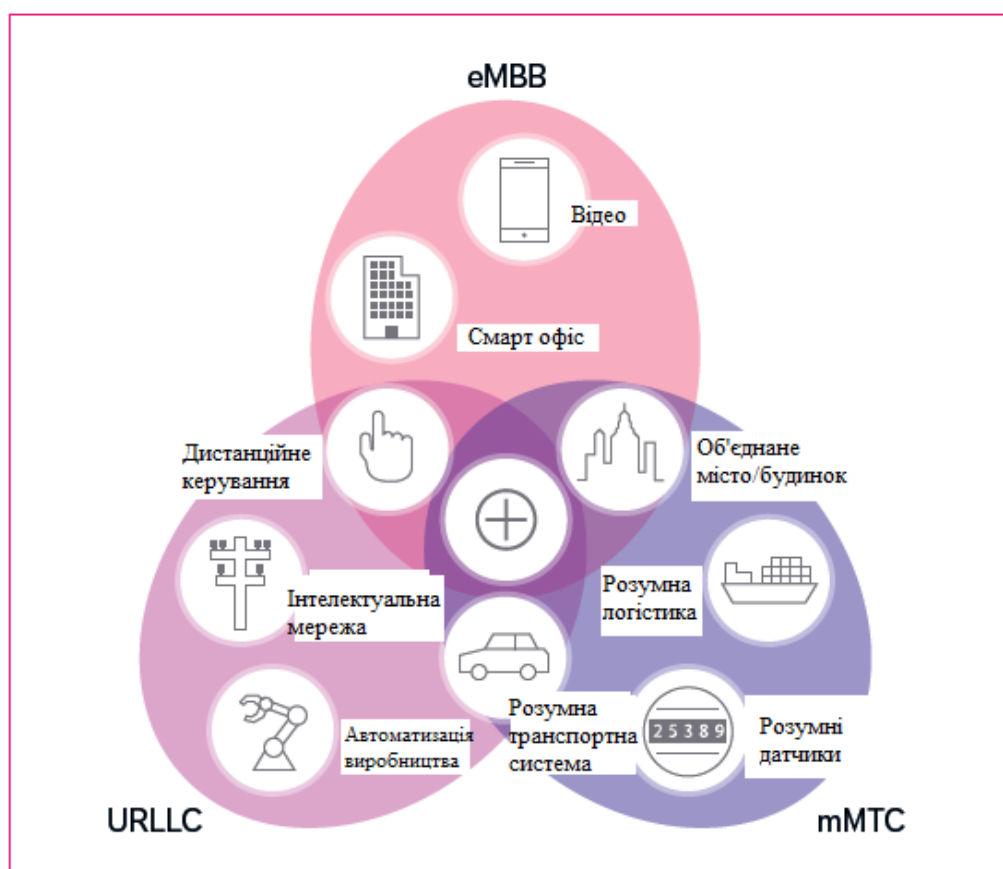


Рисунок 1.1 Три основні варіанти використання 5G і приклади пов'язаних додатків

В 3GPP є дві доріжки, які становлять дорожню карту радіодоступу 5G, як показано на рис. 1.2. Один з них заснований на еволюції LTE, а інший - на доступі New Radio (NR). На доріжці LTE-5G удосконалення будуть продовжувати дозволяти їй підтримувати якомога більше вимог 5G і

варіантів використання. На відміну від доріжки LTE-5G, доріжка NR-5G вільна від вимог сумісності і, таким чином, дозволяє вводити більш фундаментальні зміни, такі як спектр таргетингу на високих (мм-хвильових) частотах. Проте, NR розробляється масштабуючим чином, тому його можна в кінцевому підсумку перенести на частоти, які в даний час обслуговуються LTE [16].

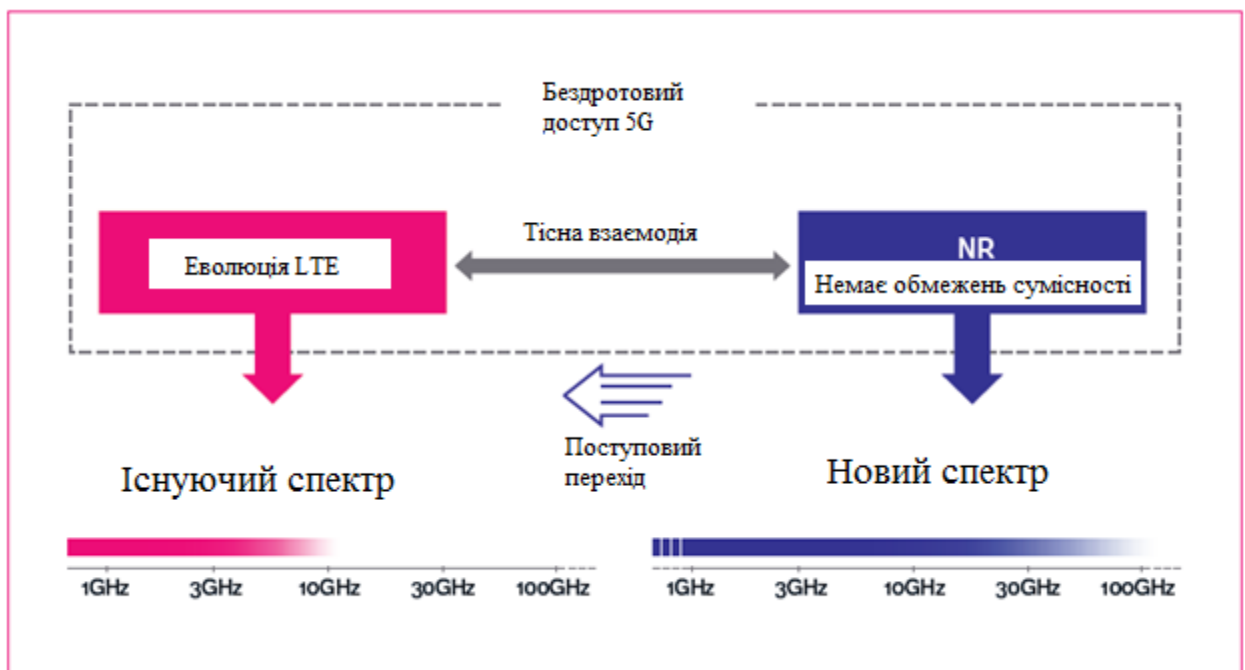


Рисунок.1.2 Дорожня карта радіодоступу 5G

Незважаючи на те, що перспективи NR-5G цікаві, операторам, які вже зробили значні інвестиції в LTE, не потрібно турбуватися - перехід від LTE до плагінів 5G через 5G є найбільш логічним шляхом. І очікування LTE Rel-14, які планується завершити в березні 2017 року, і сильні амбіції LTE Rel-15 показують, що плани розвитку для треку LTE-5G міцні.

Процес створення LTE 5G-ready включає в себе безліч поліпшень і нових функцій Rel-14 і Rel-15. Найбільш важливими з них є удосконалення швидкості передачі даних користувачів і пропускної здатності системи за допомогою FD-MIMO, поліпшена підтримка неліцензованих операцій і скорочення латентності як в плані управління, так і в призначених для

користувача площинах (UPs). Удосконалення в Rel-14 і Rel-15 також спрямовані на поліпшення підтримки таких випадків, як масивний MTC, критично важливі повідомлення і ITS.

1.3 Вплив FD-MIMO і неліцензійних операцій на підвищення швидкості передачі даних користувача і можливостей системи

FD-MIMO і неліцензовані операції є двома основними функціями майбутніх випусків LTE, які покликані підвищити швидкість призначених для користувача даних і пропускну здатність системи, що відповідають стандартам 5G.

1.3.1 FD-MIMO

Поліпшення MIMO в 3GPP дозволяє динамічно адаптувати передачу як по вертикалі, так і по горизонталі, використовуючи керовану двовимірну антенну решітку. Концепція FD-MIMO в майбутніх випусках LTE ґрунтується на механізмах зворотного зв'язку стану каналу (CSI), представлених в LTE Rel-13, в яких попередні матричні кодові книги з підтримкою двухрозмірних макетів портів с до 16 антенними портами. Сигнали опорного сигналу CSI без попереднього кодування (CSIRS) передаються від кожної антени і передаються в клітинку, а прекодер виводиться терміналом. LTE Rel-13 також представила інший тип зворотного зв'язку CSI з термінально-специфічним, заснованим на промені CSI-RS, таким же чином, як і фізичний спадний загальний канал (PDSCH). У цьому випадку напрямок формування променя для кожного терміналу визначається базовою станцією, а не виводиться з зворотного зв'язку терміналу.

Щоб поліпшити роботу як без попереднього кодування, так і з використанням діаграми спрямованості CSI-RS, Rel-14 представить кілька нових функцій:

1. Включаючи гібридний режим без кодованого / променевого кодування CSI з оптимізованим зворотним зв'язком;
2. Аперіодичний запуск вимірювань CSIRS;
3. Підтримка до 32 антенних портів;
4. Просторово багатий, розширений зворотній зв'язок CSI;
5. Схема передачі з розімкненим контуром.

Гібридний неконденсований і керований променем режим CSI з оптимізованим зворотним зв'язком дозволить періодично передавати початковий, чи некодований CSI-RS. Потім термінали можуть реагувати на бажаний напрямок для другого, більш частого, з формуванням променя CSI-RS.

Аперіодичний запуск вимірювань CSI-RS полегшує об'єднання ресурсів CSI-RS, що дозволяє ефективно використовувати вимірювальні ресурси і скоротити накладні витрати CSI-RS. В результаті більша кількість терміналів в комірці буде мати доступ до роботи з CSI-RS з формуванням променя.

Підтримка 32 антенних портів дозволяє використовувати роботу зі зворотним зв'язком з масивними настройками антен, що збільшує вигоду від розрахованого на багато користувачів MIMO (MU-MIMO).

Просторово багатий, розширений зворотній зв'язок CSI включатиме в себе інформацію про шляхи поширення декількох каналів, щоб можна було уникнути або придушити перешкоди між запланованими терміналами. Продуктивність тоді порівняна з масивними системами MU-MIMO на основі взаємності.

Схема передачі з розімкненим контуром об'єднує повнорозмірне формування променя і рознесення передачі, призначені для високошвидкісних терміналів, де відомо напрямок променя, але короткі CSI занадто швидко змінюються.

Очікуване поліпшення пропускної спроможності системи та пропускної здатності користувача за допомогою Rel-14 FD-MIMO показано

на рис. 1.3 - тривимірний міський сценарій 3GPP 3D з 8-кратним подвійним поляризованим масивом і неповним буфером трафіку. Продуктивність на кромці осередку збільшується приблизно в 2,5 рази з розширеним зворотним зв'язком CSI і підтримкою 32 антенних портів.

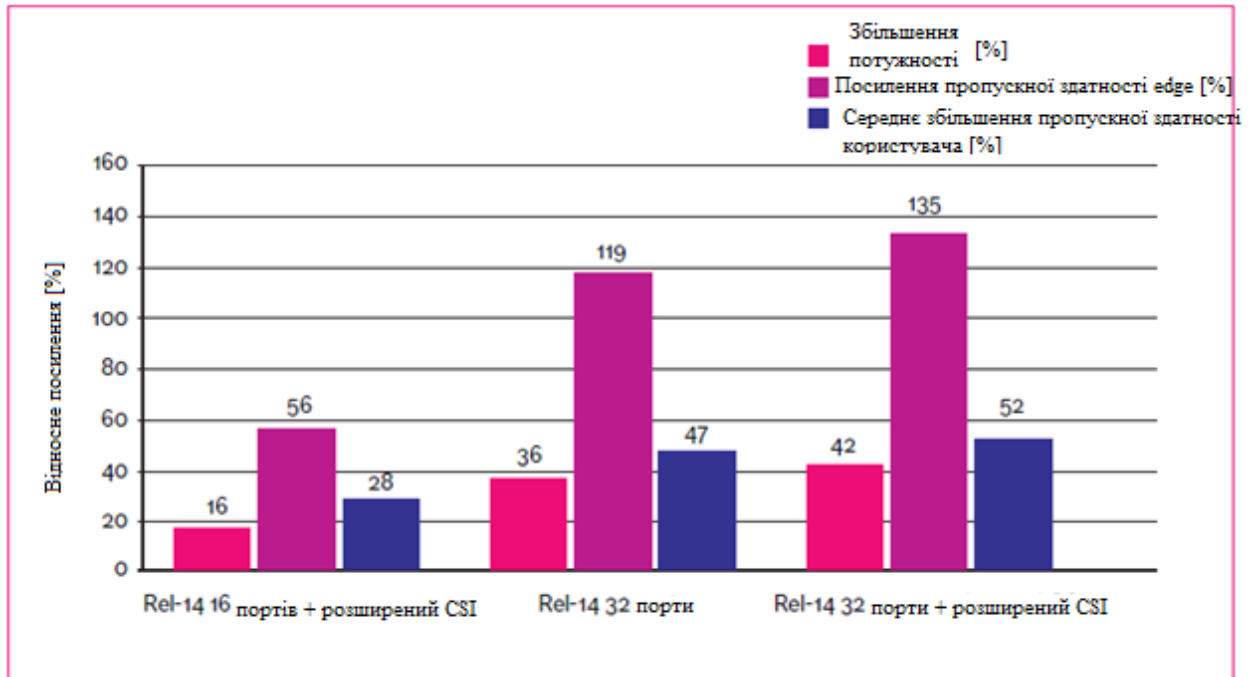


Рисунок. 1.3 Продуктивність Rel-14 FD-MIMO по 16-канальній базовій лінії Rel-13 (без удосконаленого CSI) при високому навантаженні на систему

1.3.2 Операції LTE в неліцензійному спектрі

Для задоволення постійно зростаючих вимог до трафіку багато мережевих операторів розглядають можливість додаткового використання неліцензійного спектра. LAA була введена в LTE Rel-13 для роботи з DL, і в Rel-14 вона покращена для підтримки UL. LAA використовує CA для об'єднання первинного осередку ліцензованої смуги (PCell) з неліцензованими вторинними осередками діапазону (SCells). Однак SCells зазвичай мають обмежену потужність передачі, що призводить до того, що зони покриття менші, ніж ті, які можуть надати PCells. У цій схемі PCell забезпечує надійне покриття керуючих повідомлень і високопріоритетний

трафіку, в той час як SCells забезпечують великий обсяг спектра і високі швидкості передачі даних, коли вони доступні. На рис. 1.4 показано, як LAA пропонує комбінацію основних переваг, що надаються як ліцензованим, так і неліцензійним спектром.

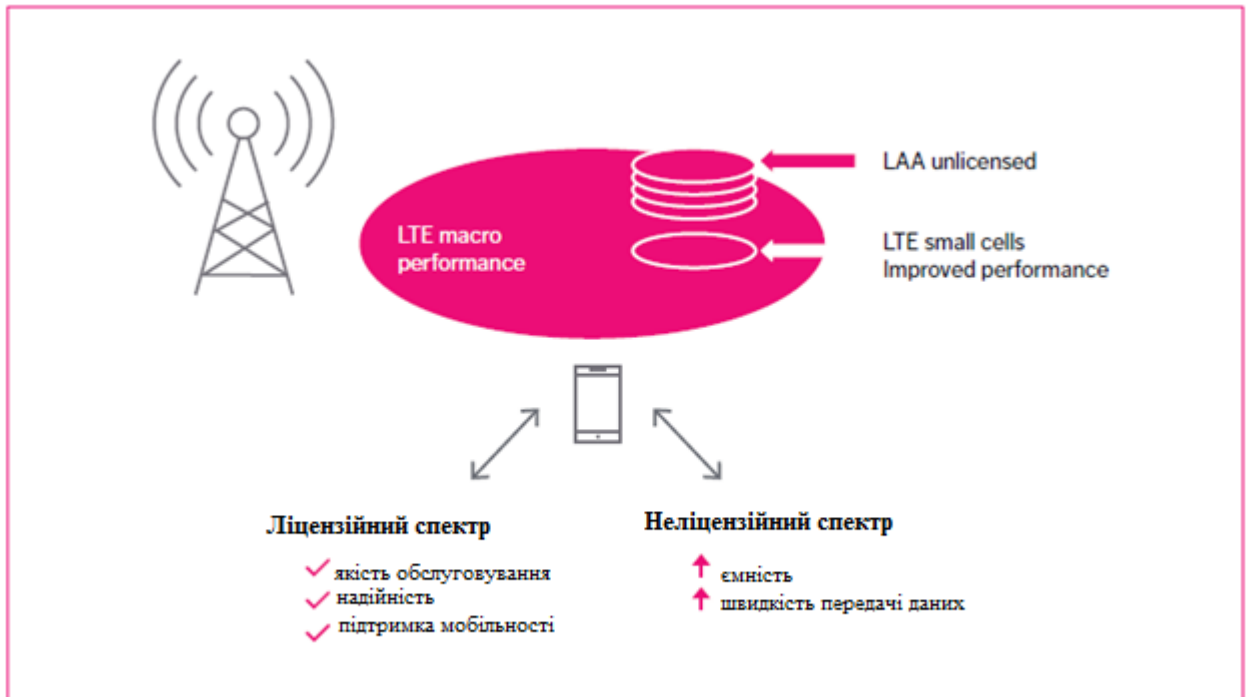


Рисунок. 1.4 Ілюстрація LAA

Кілька рішень були включені в 3GPP для досягнення співіснування з іншими технологіями, такими як WLAN, які працюють в тій же смузі, що і LAA. До них відносяться динамічний вимір / вибірка несучих, протокол Listen-Before-Talk і переривчаста передача з обмеженою максимальною тривалістю.

Інтелектуальне і адаптивне управління трафіком між ліцензованими і неліцензованими носіями і між неліцензованими носіями - також може ще більше поліпшити співіснування.

На рис. 1.5 показана пропускна здатність мережі в сценарії спільного співіснування LAA, де кожен з двох операторів розгортає чотири LAA або чотири вузли WLAN на точку доступу. Осередки LAA підтримують істотно

більш високу пропускну здатність на одному і тому ж каналі 20 МГц в порівнянні з вузлами WLAN. Це пов'язано з тим, що надійний дизайн фізичного шару LAA дозволяє надійно і ефективно використовувати частоту повторного використання. Фактично, більш ефективна мережа LAA залишає більше можливостей для спільної мережі WLAN.

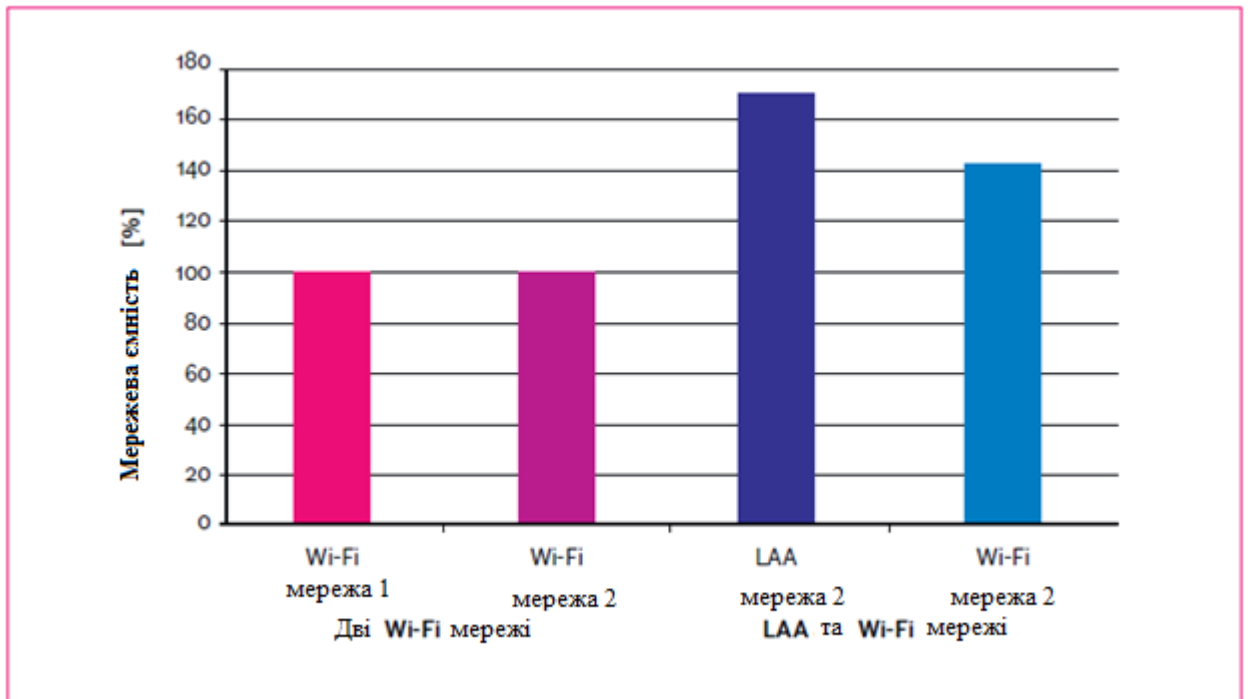


Рисунок. 1.5 LAA-WLAN відкрите співіснування (40 МГц розподілено на двох носіях, обидві мережі працюють на частоті 5 ГГц)

Подальші удосконалення LAA очікуються в LTE Rel-15, в першу чергу в UL-контрольній інформації і підтримці каналу довільного доступу на неліцензований смузі SCell. Це дозволило б вивантажити більше трафіку з ліцензійного діапазону PCells і дозволити подальше розгортання, а також дозволити використання таких випадків, як волоконно-оптичні видалені радіовиводи.

Ще одним потенційним удосконаленням LTE Rel-15 є подвійна взаємодія між основним еволюціонованим вузлом В (eNB) з ліцензованою смугою і неліцензованим вторинним eNB діапазону. Це ще більше

розширить можливості розгортання, дозволивши агрегацію між мережевими вузлами, які не підключені через шлюз з малою затримкою. Нарешті, Rel-15 може включати більше варіантів розгортання і сценаріїв, таких як автономні і mMTC-операції в неліцензійному спектрі.

1.4 Вплив розвитку LTE на скорочення затримки

Іншим важливим аспектом розширення LTE є впровадження методів скорочення затримки для призначених користувацьких і керуючих площин (UP і CP). Скорочення затримки не тільки сприяє підвищенню швидкості передачі даних, але також дозволяє використовувати нові варіанти використання, такі як критична зв'язок і ITS.

Зниження затримки користувацької площині. Впровадження швидкого доступу до UL - це перший крок до скорочення часу очікування UR. Як зазначено в Rel-14, швидкий доступ до UL дозволяє налаштувати термінал з підтримкою висхідної лінії зв'язку, доступним в кожній мілісекунді, для використання тільки тоді, коли є дані висхідної лінії зв'язку для передачі. Використовуючи поточний запит (SR) на основі запиту на планування, термінал повинен передати запит, дочекатися отримання гранту і потім чекати використання гранту. Порівняння швидкого доступу UL з доступом SR показано на доріжках а і б на рис. 1.6. Попередньо сконфігурований грант в швидкому доступі UL мінімізує час очікування, що зменшує середню затримку радіодоступу для даних висхідній лінії зв'язку більш ніж на половину.

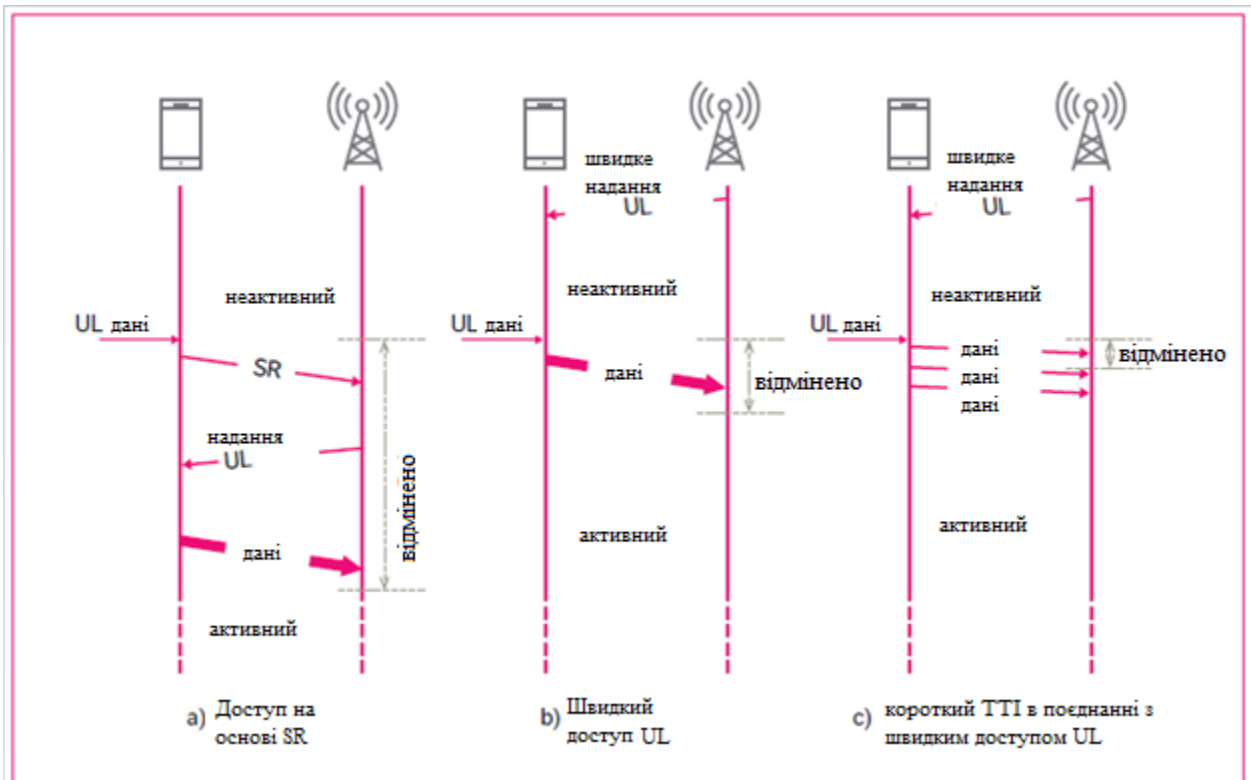


Рисунок.1.6 SR (a), швидкий доступ UL (b) і короткий TTI в поєднанні з швидким доступом UL (c)

Другий крок зниження затримки складається з двох поліпшень, які призначені для специфікації в Rel-15. Перше - це скорочення часу обробки: надання терміналу відповіді на дані низхідні лінії зв'язку і гранти висхідної лінії зв'язку за три мілісекунди замість чотирьох. По-друге, введення більш коротких інтервалів часу передачі (TTI): прискорення всього ланцюжка очікування можливості передачі, планування і підготовка до передачі, передача даних і, в кінцевому рахунку, обробка отриманих даних і відправка зворотного зв'язку.

При короткому TTI, як показано на доріжці c на рис. 6, передачі можуть виконуватися з меншою тривалістю (всього лише одна сьома від довжини звичайного LTE TTI). Кожна з цих коротких передач може бути запланована окремо з новим DL-каналом управління всередині каналу, зі зворотним зв'язком, відправленої в новий канал управління UL. Планування і зворотний зв'язок відправляються до сусідніх підкадрів для найкоротшого

часу передачі, в результаті чого загальна одностороння передача радіодоступу становить близько 0,5 мс, включаючи час обробки даних [13].

Рис. 1.7 ілюструє приріст часу в зворотному напрямку (RTT), виконаний з використанням короткого ТТІ і швидкого доступу UL. З симуляції також спостерігалися поліпшення в пропускній здатності для завантаження протоколу передачі файлів (FTP) на 70 відсотків: ефект, викликаний більш швидким збільшенням бітрейту TCP завдяки більш короткому RTT даних і відповідей.

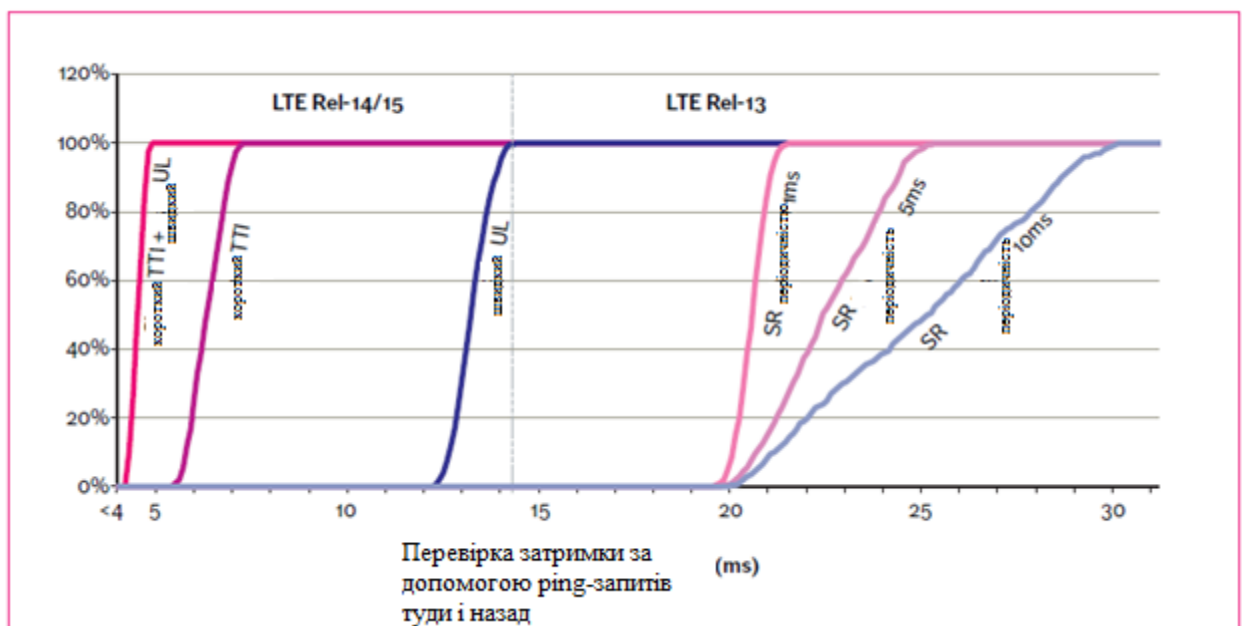


Рисунок. 1.7 Вплив короткого ТТІ і швидкого доступу UL на RTT

1.5 Варіанти зменшення сигналізації у версіях Rel-13, Rel-14 та Rel-15

Переходи стану LTE залучають сигналізацію: перехід від RRC_IDLE до RRC_CONNECTED включає 9 передач по повітряному інтерфейсу. У Rel-13 були введені два варіанти зниження сигналізації: підключення / відновлення з'єднання RRC для використання з передачею даних на основі UP на радіоканалах даних (DRB) і даних по недоступному рівню (DoNAS) для передачі даних на основі CP по сигнальному радіоканалу (SRB).

Функція `suspend / resume` дозволяє тимчасово призупинити підключення до даних і контекст, який буде зберігатися в RAN і базовій мережі (CN) під час `RRC_IDLE`. При наступному переході до `RRC_CONNECTED` з'єднання відновлюється зі збереженим контекстом, що значно знижує передачу сигналів до чотирьох або п'яти передач. Функція DoNAS забезпечує аналогічне скорочення сигналізації шляхом опускання безпеки доступу (AS) і шляхом передачі даних по CP замість створення традиційних радіоканалів UP. Для задоволення постійно зростаючого числа пристроїв, невеликих і / або нечастих обсягів даних і більш строгих вимог до затримки Rel-14 і Rel-15 націлені на подальше зниження сигналізації між терміналами і мережевими вузлами (RAN і CN).

У Rel-14 функція припинення / відновлення поліпшується за рахунок зменшення сигналізації між базовою станцією (BS) і CN. У Rel-13 з'єднання BS-CN було випущено разом із з'єднанням з повітряним інтерфейсом. У Rel-14 з'єднання BS-CN може підтримуватися, коли з'єднання BS-терміналу призупинено. Наприклад, RAN бере на себе відповідальність за пейджинг терміналу після прибуття DL-даних.






У Rel-14 або Rel-15 очікується ще два поліпшення зниження затримки контрольної площини. Перший - це удосконалення, яке дозволило б забезпечити більш ранню передачу даних, дозволяючи мультиплексувати дані радіоканалу UP з сигналізацією відновлення з'єднання. Друга, відома як індикація допомоги при звільненні, яка дозволяє терміналу вказувати, що у нього більше немає даних UL, і що він не очікує даних DL, тим самим забезпечуючи ранній перехід до `RRC_IDLE`.

1.6 Ряд удосконалень в LTE Rel-14 і Rel-15 для 5G

Ці удосконалення в покликані забезпечити поліпшену підтримку для використання таких випадків, як масивний МТС, критично важливі повідомлення і ITS.

1.6.1 Масивні зв'язки типу машини

LTE МТС і NB-IoT були розроблені для вирішення проблем використання mMTC. Вони пропонують аналогічні поліпшення відносно поліпшення покриття, часу автономної роботи, ефективності сигналізації і масштабованості, але вимагають трохи різних вимог з точки зору гнучкості і продуктивності. Як показано на рис. 1.8, LTE МТС краще здатна підтримувати більш високі швидкості передачі даних і як мобільність в режимі RAT, так і inter-RAT. З новою моделлю LTE МТС категорії M1 (Cat-M1) і NB-IoT, які були вказані в 3GPP Rel-13, очікується, що вартість модему може бути значно зменшена в порівнянні з пристроями Rel-8 Cat-1. Вартість буде варіюватися в залежності від особливостей, варіантів і реалізації. Очікується, що зниження витрат на модем становитиме 75-80 відсотків для Cat-M1 і навіть більше для NB-IoT з ще більш зменшеним набором функцій.

	 Пропускна здатність	 Покриття (MCL)	 Термін служби батареї	 Пропускна здатність (DL/UL)	 Мобільність
LTE МТС (Cat-M1)	1.4MHz	164dB	10+ років	300/375kbps ¹⁾ 0.8/1Mbps ²⁾	З'єднання і мобільність в режимі очікуванні
NB-IoT	200kHz	164dB+	10+ років	21/63kbps	Мобільність в режимі очікування

1) Half duplex, 2) Full duplex

Рисунок.1.8 Ключові показники ефективності NB-IoT і LTE МТС (Rel-13)

LTE Rel-14 націлений на подальше вдосконалення LTE MTC і NB-IoT за рахунок підвищення продуктивності і більш широкого використання. Більш високі швидкості передачі даних і ефективність будуть досягнуті в Rel-14, дозволяючи переносити великі шматки даних в кожній передачі і збільшуючи кількість процесів гібридного автоматичного повторного запиту (HARQ), щоб забезпечити паралельні видатні передачі під час очікування зворотного зв'язку. Збільшена пропускна здатність каналу для LTE MTC (до 5 МГц) покращує підтримку потокової передачі голосу і звуку, а також інших додатків і сценаріїв. Удосконалення NB-IoT для випадкового доступу і пейджінга збільшують універсальність несучих несучих.

Rel-14 також дозволить позиціонувати додатки (в яких знання свого місцезнаходження є критичним), підтримуючи розширені опорні сигнали, які враховують меншу пропускну здатність NB-IoT / LTE MTC. Покращення мобільності підключеного режиму поліпшить безперервність обслуговування. Багатоадресна передача зробить доставку одного і того ж контенту на кілька пристроїв більш ефективно, оптимізуючи такі варіанти використання, як оновлення прошивки і синхронний контроль над такими речами, як вуличні ліхтарі. Підтримка нижчого класу потужності NB-IoT 14 дБм дозволить використовувати менші батареї і пристрої підтримки з невеликим форм-фактором [14].

Голосове покриття для LTE MTC буде покращено в Rel-14 за рахунок збільшення охоплення VoLTE для напівдуплексного FDD / TDD за допомогою методів, які зменшують повторення DL, нові коефіцієнти повторення і скориговані затримки планування. Пристрої MTC і варіанти використання також отримують переваги від підвищення рівня сигналізації в LTE Rel-14. У випадках використання mMTC також будуть корисними декілька інших удосконалень LTE Rel-15, в тому числі:

1. Покращення затримки виникає в результаті мультиплексування призначених для користувача даних з сигналізацією відновлення з'єднання
2. Підвищення ефективності в результаті розширеного управління доступом / навантаженням в режимах очікування та підключення
3. Збільшення терміну служби батареї в результаті ослаблених вимог до моніторингу DL в режимі очікування
4. Покращена підтримка додаткових випадків використання, таких як портативні.

1.6.2 Критичне повідомлення

Такі випадки, як спостереження за електромережами, безпечно дистанційне керування і критичні виробничі операції, вимагають як низькою затримки, так і високої надійності вище поточного рівня HARQ (див. рис. 1.9). Для того щоб LTE відповідала цим вимогам 5G, існує мета зробити два поліпшення для Rel-15: надійний короткий TTI і надійна 1 мс операція. Грунтуючись на короткому TTI і швидких характеристиках UL, швидкість передачі пакетів може бути зменшена до рівня 10⁻⁵ завдяки об'єднанню надійного кодування повідомлень управління і даних, різноманітності і автоматичних повторень без зворотного зв'язку. Оскільки обробка зберігається в коротких тимчасових масштабах, весь ланцюжок передач може бути доставлений протягом 1 мс з об'єднаною надійністю кількох випробувань. (Мета - невеликі осередки, такі як фабрики і офіси). Крім того, широкосмугове покриття з ослабленою затримкою, але з підвищеною надійністю також може бути націлене на автоматичне повторення надійних кодованих 1 мс передач з поліпшеним зворотним зв'язком.

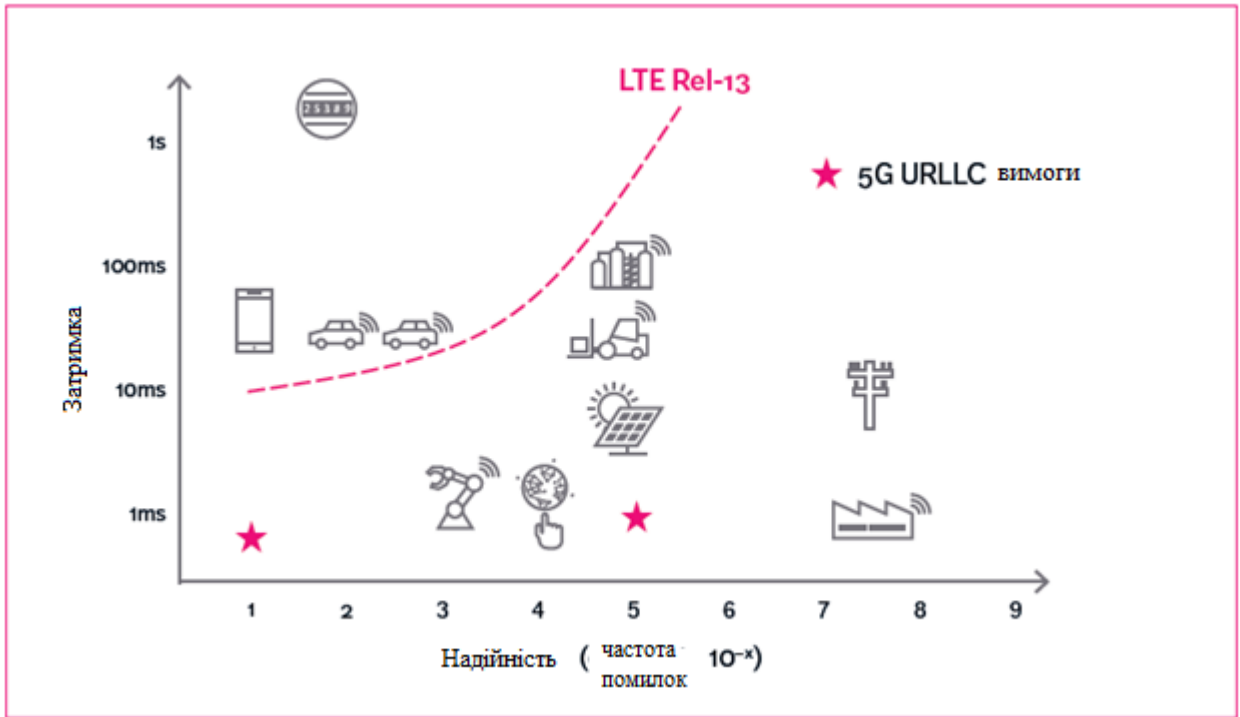


Рисунок 1.9 Критичні випадки використання зв'язку і вимоги

1.6.3 Інтелектуальні транспортні системи

Використання ICT для забезпечення більш безпечних та ефективних транспортних систем відомих як ITS. 3GPP розробляє рішення для обміну повідомленнями "транспорт до всього" (V2X) для Rel-14, звертаючись до з'єднання між транспортними засобами ("транспортний засіб-транспортний засіб" або V2V), "транспортний засіб до мережі" (V2N), "транспортний засіб до інфраструктури" (V2I) і до "транспортний засіб до пішоходу" (V2P), як показано на рис. 1.10.

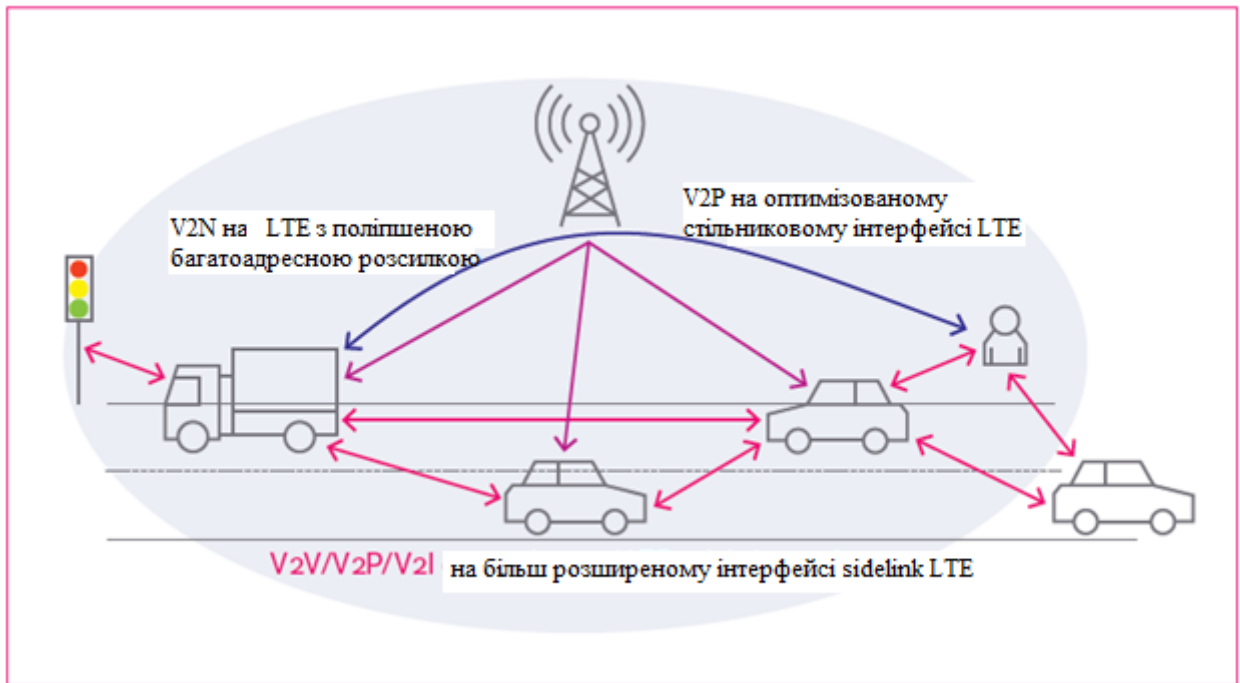


Рисунок.1.10 Ілюстрація різних ITS сценаріїв і інтерфейсів

LTE-based ITS вигоди від охоплення існуючих мереж і централізованої безпеки. Однак нові випадки використання ІТС потрібні з точки зору затримки і пропускну здатності системи. Тому прямий інтерфейс D2D, відомий як sidelink (SL) і інтерфейс стільникового зв'язку LTE, поліпшуються в Rel-14 для підтримки цих вимог.

Наприклад, збільшена щільність символів пілот-сигналу дозволить оптимізувати SL для швидкої зміни умов поширення і серйозних зрушень частоти на приймачі через високу відносної швидкості (до 500 км / ч) і більш високу несучу частоту (до 6 ГГц).

Поліпшене керування радіоресурсами - ще одне важливе вдосконалення для підтримки додатків ITS. Воно базується на протоколі вибору ресурсів на основі зондування, де кожен пристрій автономно вивчає, як інші пристрої використовують радіоресурс і передбачає їхню майбутню поведінку, використовуючи квазіперіодичний характер повідомлень ITS [3].

Rel-14 підтримує використання інформації географічного розташування для забезпечення централізованого розподілу ресурсів в eNB або для автономного вибору ресурсу в сконфігурованому полі радіоресурсів.

Він також підтримує протоколи мультимедійної ширококомовної / багатоадресної передачі (MBMS), які оптимізовані для забезпечення низької затримки і покриття, а також для ефективної доставки повідомлень V2X. Нарешті, очікувані поліпшення забезпечать справедливе та ефективне співіснування з технологіями ITS, відмінними від 3GPP, такими як спеціалізовані комунікації короткого діапазону (DSRC).

На рис. 1.11 показано чисельне порівняння можливостей різних технологій для трансляції повідомлень V2V. У типових сценаріях (міських і автомобільних) рішення, засновані на LTE (SL з централізованим розподілом ресурсів і стільникового многоадресної розсилкою), працюють значно краще, ніж на основі DSRC.

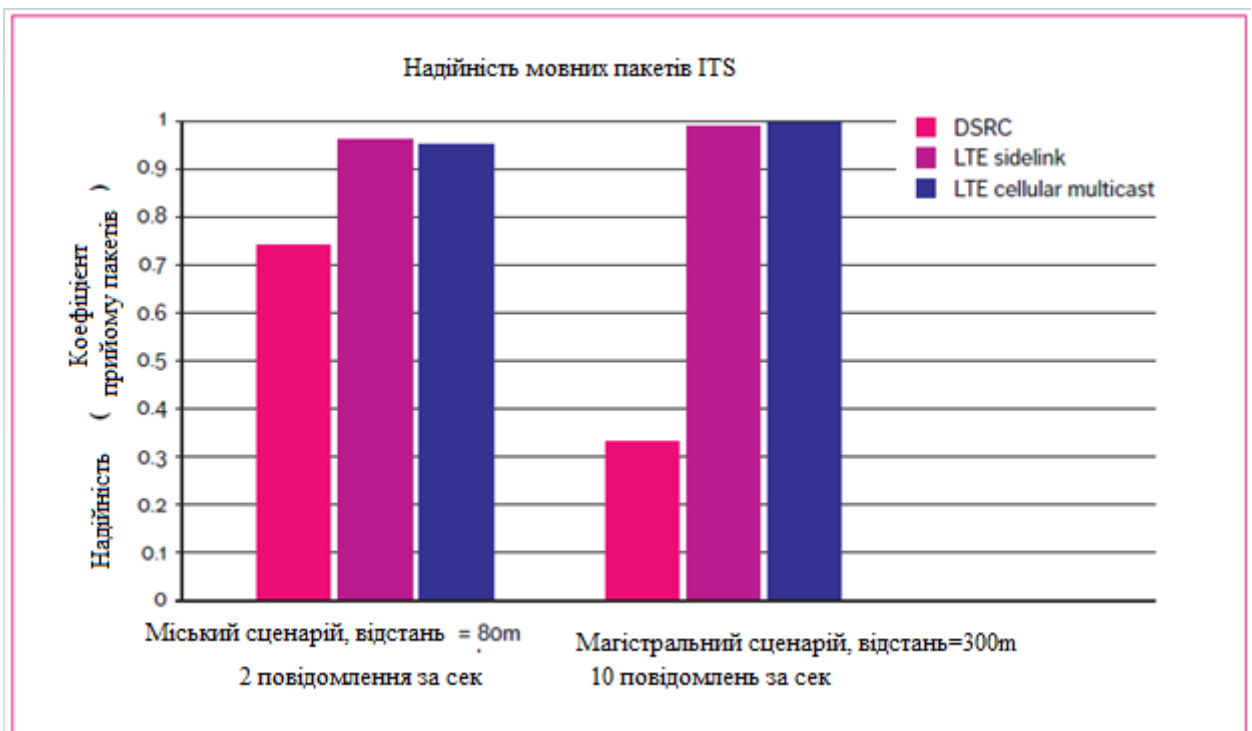


Рисунок.1.11 Порівняння різних технологій для мовлення ITS повідомлень

1.7 Використання віртуалізації в системах 5G

Нова технологія 5G прекрасно поєднується з такими технологічними областями, як мережева віртуалізація (NV), програмна мережу (SDN),

віртуалізація мережевих функцій (NFV) і IoT. Ринок 5G дуже очікуваний, оскільки він має можливість об'єднати багато з цих ринків з використанням технологій NV.

5G, який все ще перебуває на ранніх стадіях розробки стандартів, обіцяє велику пропускну здатність і більшу гнучкість для мобільних мереж. Існує кілька стандартів або визначень 5G, з яких вже стає ясно, що 5G буде керувати віртуалізацією через свою залежність від інтеграції з NV і SDN.

1.8 Технології віртуалізації в 5G

Існує еволюційний прогноз стандартів 5G, які до сих пір не визначені. Яку б остаточну форму 5G ми б не використали, ми знаємо, що вона забезпечить як велику пропускну здатність, так і гігабітну швидкість, і буде все більше орієнтуватися на якість обслуговування (QoS) або можливість «зрізати» додатки в віртуальній мережі відповідно до потреб служби. Це буде сприяти наскрізній віртуалізації у всіх мережах, включаючи необхідність віртуальних підключень до підприємств, використовуючи 5G як визначається програмним забезпеченням WAN [1].

Проте стандарти 5G не близькі до існуючих, і більшість людей не очікують чогось до 2018 року. Багато особливостей далекі від вирішення. 5G буде забезпечувати пропускну здатність до 100 разів у порівнянні з 4G (LTE) через мережу радіодоступу (RAN).

Питання про пропускну здатність буде включати в себе найбільш технічно складні частини мережі - як доставляти гігабітні швидкісні з'єднання по повітря, використовуючи мобільний спектр. 5G має потенціал для створення віртуалізованих наскрізних підключень, включаючи визначення рівнів обслуговування, що проходять через ядро мережі. Поки що, найбільші оператори мобільного зв'язку використовують концепцію 5G, і

перший початковий додаток, швидше за все, буде фіксованим мобільним доступом.

1.9 Підняття питання QoS в системах 5G

Традиційні послуги, в тому числі мобільні, були розділені на стратегічні точки мережі у вигляді цільових пристроїв. Здебільшого існує розрив між приватними і загальнодоступними мережами, включаючи хмарні сервіси, і мережею провайдерів електронних послуг. 5G збільшить необхідність створення безшовної віртуалізації в глобальній мережі постачальника послуг, щоб корпоративні додатки могли підтримувати QoS через хмару [10].

У той же час NFV отримує найбільшу вигоду в мобільній області, включаючи служби шарів 4-7 і мобільні додатки, такі як віртуальне розвинене ядро пакета (vEPC). Це означає, що світ мобільних мереж переходить до моделі віртуалізації, заснованої на центрах обробки даних, яка ближче до того, яка створюється корпоративною хмарою. Тепер уявіть собі мережу, яка може використовувати регульовану пропускну здатність з пропускнуою спроможністю в гігабітній мережі з нарізкою мережі QoS на основі пакетів. Ідея полягає в тому, щоб розрізнити людей, які скачують Netflix і тих, хто проводить віртуальну операцію, і ставитися до них належним чином. Це матиме неабиякий вплив на мережі 5G, особливо на основні, за рахунок величезного обсягу даних і потреб розвідки в мобільних точках доступу і в центрах передачі даних по всьому світу.

Вже приводяться плани нових мережевих архітектур. Недавні повідомлення показують, що Verizon зосереджена на створенні набору регіональних центрів обробки даних для надання NFV на edge мережі. І AT&T був великим прихильником NFV, заявивши, що планує стратегію центру обробки даних, яка буде віртуалізувати до 80 відсотків своєї мережі.

1.10 Реакція постачальників мереж на використання 5G

Зрушення матиме величезний вплив на мережеву архітектуру, де постачальники мережевого обладнання все більше зосереджуються на мобільних додатках, орієнтованих на дані, особливо на NFV.

Давайте подивимося, як деякі великі виробники реагують на цей масовий зсув:

- «Brocade» побачила, що бездротова можливість настільки велика, що призвела до недавнього придбання Ruckus Wireless, заявивши: «Придбання зміцнить здатність Brocade розвивати можливості ринків, що розвиваються навколо мобільних послуг 5G, Internet of Things (IoT), «розумних міст» і LTE / Wi-Fi конвергенції».

- Ericsson багато говорив про 5G, що вимагає мережевого зрізу, який може забезпечити зростання трафіку відео і машин-машин (M2M). Компанія просуває технологію SDN та NFV для цього.

- Cisco, тепер партнер Ericsson, відзначив наслідки, які 5G матиме для попиту на мобільний транспорт та іншу мережеву інфраструктуру. Перехід на 5G і IoT став великим драйвером партнерства між Cisco і Ericsson, гігантами в космосі.

- VMware прагне захистити свою ліцензію на віртуалізацію, впроваджуючи такі продукти, як vCloud для NFV, що дозволить постачальникам послуг створювати хмарні сервіси та інфраструктуру NFV, велика частина яких призначена для мобільних додатків.

- Nokia відзначає, що 5G стимулюватиме попит на SDN та NFV, щоб задовольнити величезну потребу в масштабуванні. Нещодавно він анонсував архітектуру 5G на основі SDN, яка може автоматично адаптувати радіодоступ та основні мережеві ресурси для задоволення потреб різних служб, зміни трафіку з плином часу і місцем розташування, і топологією мережі.

1.11 Приклад використання мобільної хмарної мережі

Мобільна хмарна мережа MCN -заснована на сервіс-орієнтованій архітектурі, яка надає наскрізні, складені служби з використанням хмарних обчислень і SDN технології. Ця архітектура сумісна з NFV, але за межами NFV, приносить нові поліпшення. Приклад використання мобільної хмарної мережі включає в себе дійсні реалізації телекомунікаційного обладнання в якості програмного забезпечення і хмарну інфраструктуру, забезпечуючи відповідне уявлення про те, як нове віртуалізоване середовище буде реалізовано [8].

Слідуючи тенденції віртуалізації мережевих функцій (NFV), проект Mobile Cloud Networking (MCN) розробляє платформу для підтримки хмарних сервісів для надання хмарних мережевих функцій і великого набору відповідних віртуальних мережевих сервісів. Для того, щоб скористатися перевагами технологій, запропонованих хмарними обчисленнями, сьогоденні мережі зв'язку повинні бути перероблені і адаптовані до нової парадигми як для розробки комплексної платформи підтримки послуг, так і шляхом відповідної програмної підтримки мережевих компонентів. В рамках проекту Mobile Cloud Networking ця нова парадигма була розроблена, і деякі ранні результати вже доступні для використання в співтоваристві. Зокрема, приклад спрямований на розгортання мобільної базової мережі в хмарній інфраструктурі і відображення автоматизованого, еластичного і гнучкого механізму, запропонованого такими технологіями для типових мережевих послуг.

Приклад відображає як мобільна базова мережа може бути запроектована на вимогу поверх стандартної хмарної інфраструктури.

Сценарій буде наступним:

(1) орендар (корпоративний кінцевий користувач (EEU) в термінології MCN) - може бути оператором мобільної віртуальної мережі (MVNO) або корпоративної мережі - запитує підтвердження екземпляра служби мобільної базової мережі через панель моніторингу MCN. Service Manager - службний інтерфейс, в який можуть заходити орендарі і запитувати автоматичне створення екземпляра служби через API або призначений для користувача інтерфейс. Зокрема, розгортання такої базової мережі буде здійснюватися поверх розподіленої хмари, розміщеної в кількох місцях у Європі. По завершенні процедур ініціалізації кінцеві точки мобільного базової мережі будуть передані в EEU.

(2) У EEU буде можливість мати доступ до веб-інтерфейсу Home Subscriber Server (HSS) і отримати нових абонентів. Інформація про абонентів буде також використовуватися для налаштування клієнтського пристрою (в нашому випадку ноутбуку).

(3) Клієнтський пристрій відправить запити на вкладення в мобільну базову мережу та встановить службу підключення. Оскільки клієнти будуть знаходитися в США, буде встановлено VPN-з'єднання з емулятором eNodeB, через який буде відправлений запит на вкладення. По завершенні процедури вкладення весь трафік даних буде перенаправлений до Європи. Можна буде показати, що публічні IP-адреси, призначені абоненту, є частиною діапазону IP-тестування європейського хмарного тесту.

(4) Клієнти, підключені до мережі, будуть встановлювати виклик з використанням підсистеми IP-мультимедіа (IP Multimedia Subsystem), що надається MVNO. Під час виклику адміністратор MVNO може відкрити засіб «Моніторинг як послуга», що надається платформою MCN, і перевірити поточну ситуацію з сервісами. Для цього два IMS клієнта будуть встановлені на демонстраційному пристрої.

(5) Паралельно для порівняння використання розгорнутих компонентів буде використовуватися інструмент зіставлення. Інтеграція модуля управління часом виконання в оркестр забезпечує інструменти для прийому

автоматичних сигналів тривоги, що дозволяють EEU налаштувати реакцію свого сервісу на такі події, як збій віртуальних ресурсів або раптовий пік використання ресурсів.

(6) У кінці демонстрації можна буде показати, що MVNO може позбутися створеної базової мережі і звільнити ресурси, які більше не потрібні. Після цієї операції MVNO отримає рахунок, який вказує витрати на запуск такої віртуалізованої базової мережі.



Рисунок.1.12 Загальна архітектура прикладу використання мобільної хмарної мережі

Конкретні деталі

Приклад включає в себе різні етапи:

- Комплексний надійний сервіс Orchestrator, керує всім життєвим циклом складових сервісів, від їх розгортання і конфігурації до їх видалення. Він здатний динамічно надавати повну віртуальну мережу, починаючи з радіотреку, через основну мережу до надання послуг IMS. Крім того, завдяки використанню хмарних стандартів гарантується сумісність між гетерогенними службами, розробленими декількома провайдерами. Послуга доставляє з радіостанції всю основну мережу до надання послуг IMS. (Оркестрація розроблена на основі відкритого вихідного коду, доступною під ліцензією Apache 2.0.)
- Хмарна інфраструктура на основі OpenStack забезпечує розгортання хмарних мережевих сервісів на вимогу.

- Базова система моніторингу для забезпечення миттєвої ємності і тригерів для адаптації віртуальної мережевої інфраструктури. Це буде частиною оркестрованої композиції.
- Система обліку та виставлення рахунків для надання послуг з оплати і виставлення рахунків назад абоненту або наданій службі. Це буде частиною оркестрованої композиції.
- Набір віртуалізованих мережевих функцій:
 - Реалістичне виконання підсистеми IP мультимедіа 3GPP (IMS) на основі open-source OpenIMSCore
 - Реалістичне виконання віртуального 3GPP EPC на основі інструментарію Fraunhofer FOKUS Open5GCore
 - Емуляція LTE ґрунтується на реалізації Fraunhofer FOKUS Open5GCore eNB
- Приклад створення викликів IMS за наданими віртуалізованими мережевими функціями на вимогу.

1.12 Висновки з розділу 1

LTE цілком може відповідати всім найбільш важливим вимогам 5G, включаючи підвищення швидкості передачі даних і підвищення пропускнуої спроможності системи за допомогою FD-MIMO, поліпшену підтримку неліцензованих операцій і скорочення затримки як в площині користувача, так і в сигналізації. Покращення, які заплановані в Rel-14 і Rel-15, не тільки гарантують, що LTE забезпечить кращу підтримку для масових MTC і ITS; Вони також дозволять LTE вирішувати нові варіанти використання, такі як критичні комунікації.

Кілька прикладів того, як найбільші мережеві постачальники в світі змінюються, щоб позиціонувати себе для 5G, коли вони зливаються з віртуалізацією. Великі виробники мікросхем, такі як Broadcom, Cavium і

Intel не просують ні однієї можливості на ринку мереж 5G. Наслідки на ринках чіпів для споживчих пристроїв, що також будуть величезним

Цей приклад підкреслює основні досягнення в проекті FP7 для мобільних хмарних мереж ЄС, надаючи сервіс-орієнтовану архітектуру, яка підтримує комплексні і складені служби з використанням технологій хмарних обчислень і SDN.

РОЗДІЛ 2 ВІРТУАЛІЗАЦІЇ МЕРЕЖЕВИХ ФУНКЦІЙ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ

2.1 Віртуалізована архітектура та її види

NF пов'язані з функціональними можливостями управління і призначеною для користувача / інформаційною площиною. Деякі з них можуть мати тільки керуючі функції (наприклад, функція виявлення і вибору доступу до мережі - ANDSF), а інші можуть мати обидві (наприклад, Public Data Network Gateway -PGW). При переміщенні NF в віртуалізовану архітектуру існує два можливих загальних підходи (див. рис. 2.1):

- Комплексний підхід до віртуалізації - всі функціональні об'єкти управління і призначені для користувача площини переміщуються в віртуальні ресурси.

- Частковий підхід до віртуалізації - віртуальні тільки об'єкти функціональної площини управління, а призначений для користувача трафік перенаправляється і обробляється фізичним обладнанням. У цьому підході SDN може зіграти вирішальну роль [2].

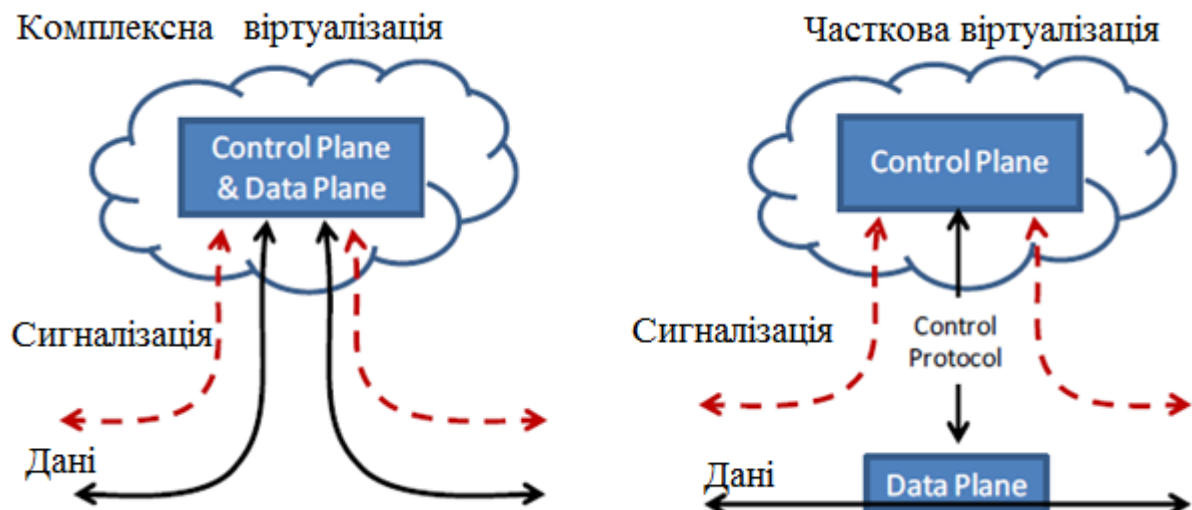


Рисунок. 2.1 Представлення комплексної і часткової віртуалізації

Розглянемо більш детально комплексний віртуалізований підхід.

Крім того, рішення про те, які NF для віртуалізації залежить від:

1. Чи може конкретна NF фактично може бути віртуалізована з технічної точки зору (наприклад, чи є апаратні обмеження або обмеження продуктивності);
2. Чи процес і зусилля при цьому приносять реальну цінність, як з економічної, так і з операційної точки зору.

Відповідь на ці питання прокладе шлях до перших розгортань VNF. У цьому сенсі ETSI визначив і почав вивчати прецеденти, які, як вважається, представляли комерційний і технічний інтерес, такі як CPE, EPC і IMS.

2.2 Історія NFV. Довідкова архітектура ETSI та процес віртуалізації

NFV виникла в співтоваристві постачальників послуг, оскільки оператори шукали способи скоротити витрати і прискорити розгортання прибуткових послуг для монетизації своїх мереж і збільшення доходів. Апаратні мережеві пристрої, які зазвичай є дорогими і складними для розгортання та управління, обмежують можливості постачальників об'єднати функціональність і швидко запускати нові служби.

Постачальники послуг відчували, що якщо вони зможуть відокремити мережеві сервіси від апаратного забезпечення, це дозволить їм розгортати мережеві компоненти, які можуть по-справжньому відповідати і підтримувати повністю віртуалізовану інфраструктуру, включаючи сервери, сховища і навіть інші мережі.

По мірі того, як NFV набрав силу, оператори зібралися в Європейському інституті стандартів телекомунікацій (ETSI) і створили групу галузевих специфікацій (ISG) для NFV для прискорення процесу віртуалізації мережевих функцій. ETSI ISG для NFV, запущений в січні 2013 року, що працює над розробкою вимог і архітектури віртуалізованих мережевих

функцій в мережі телекомунікацій. Він включав такі компоненти структури NFV:

- інфраструктура NFV (NFVI) - фізичні ресурси (обчислення, зберігання, мережа) і віртуальні екземпляри, що становлять інфраструктуру;
- віртуалізовані мережеві функції (VNF) - програмна реалізація мережевої функції.
- NFV Management and Orchestration (NFV MANO) - рівень управління та контролю, який фокусується на всіх завданнях управління віртуалізацією, необхідних протягом усього життєвого циклу VNF.

Кожен з цих компонентів містить ряд різних технологій NFV, які організації можуть розгорнути для досягнення гнучкості, масштабованості і ефективності, які їм потрібні.

Оскільки була випущена оригінальна модель ETSI для NFV, деякі оператори вимагали більш швидкого і органічного розвитку платформ NFV. Деякі оператори навіть склали свої власні технологічні програми, в тому числі розробляють свої власні проекти з відкритим вихідним кодом, які потім передають спільноті (наприклад, OSM / OpenMANO від Telefonica, ONAP, що включає ECOMP від AT & T і Open-O від China Mobile).

ETSI NFV надав посилання на базову архітектуру структури NFV, зображену на рис. 2.2. Він фокусується на функціях, необхідних для віртуалізації і подальшої експлуатації мережі оператора, визначаючи основні функціональні блоки і основні контрольні точки між цими блоками.

Аналізуючи рис. 2.2, нижня ліва частина зображення являє собою інфраструктуру NFV (NFVI), яка включає всі апаратне і програмне забезпечення, які підтримують середу, в якій VNF розгорнуті, управляються і виконуються. Ця інфраструктура забезпечує необхідні віртуальні ресурси для VNF і може фізично охоплювати кілька місць розташування. Середня ліва частина рисунку містить VNF, які використовують ресурси, надані NFVI. Права частина містить елементи управління і оркестровки. Менеджер віртуальної інфраструктури (VIM) відповідає за контроль / управління

ресурсами NFVI. Можуть бути розгорнуті кілька екземплярів VIM. Менеджер VNF відповідає за управління життєвим циклом примірників VNF (створення, налаштування, оновлення, масштабування вгору / вниз, завершення і т. д.). Orchestrator відповідає за організацію і управління ресурсами NFVI, а також за впровадження мережевих сервісів на NFVI.

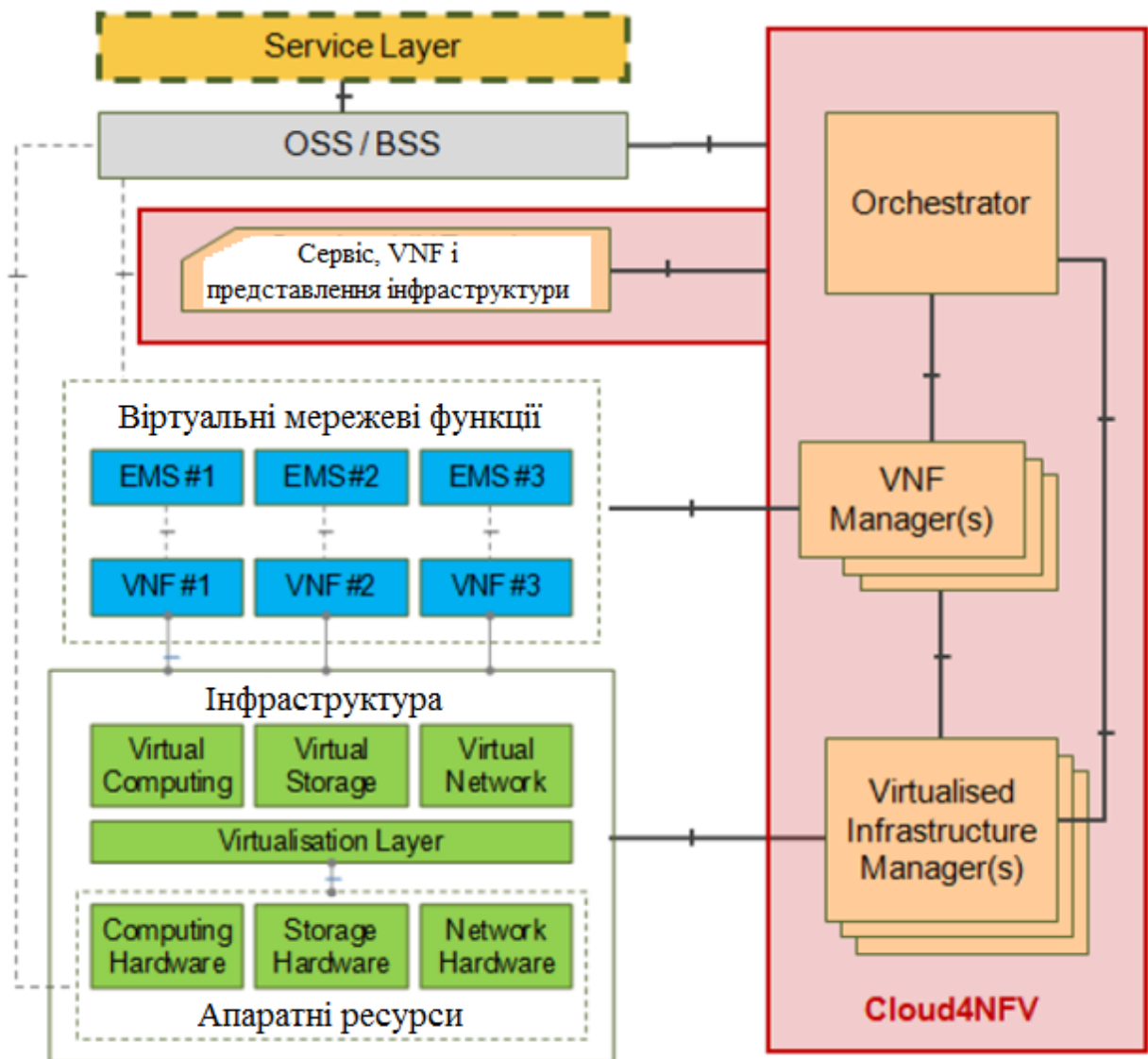


Рисунок. 2.2 Довідкова архітектурна структура NFV

У верхньому лівому кутку знаходяться OSS / BSS оператора, а також сервіс, VNF і представлення інфраструктури. Останнє містить набір даних з інформацією про шаблон розгортання VNF, пов'язаною з обслуговуванням інформації і інформаційних моделей NFVI. Також є додатковий блок, Service

Layer, який представляє, де моделюються сервіси, що надаються кінцевому користувачеві.

2.3 Інфраструктура NFV (NFVI)

На рис. 2.3 показаний приклад NFVI, що складається з: централізованого DC, точок присутності (POPs), тобто невеликих DC, а також сайту замовника, коли він вмонтував контрольовану провайдером інфраструктуру на місці для підтримки NFV. Мережа, що забезпечує зв'язок між цими місцями розташування, вважається частиною NFVI.

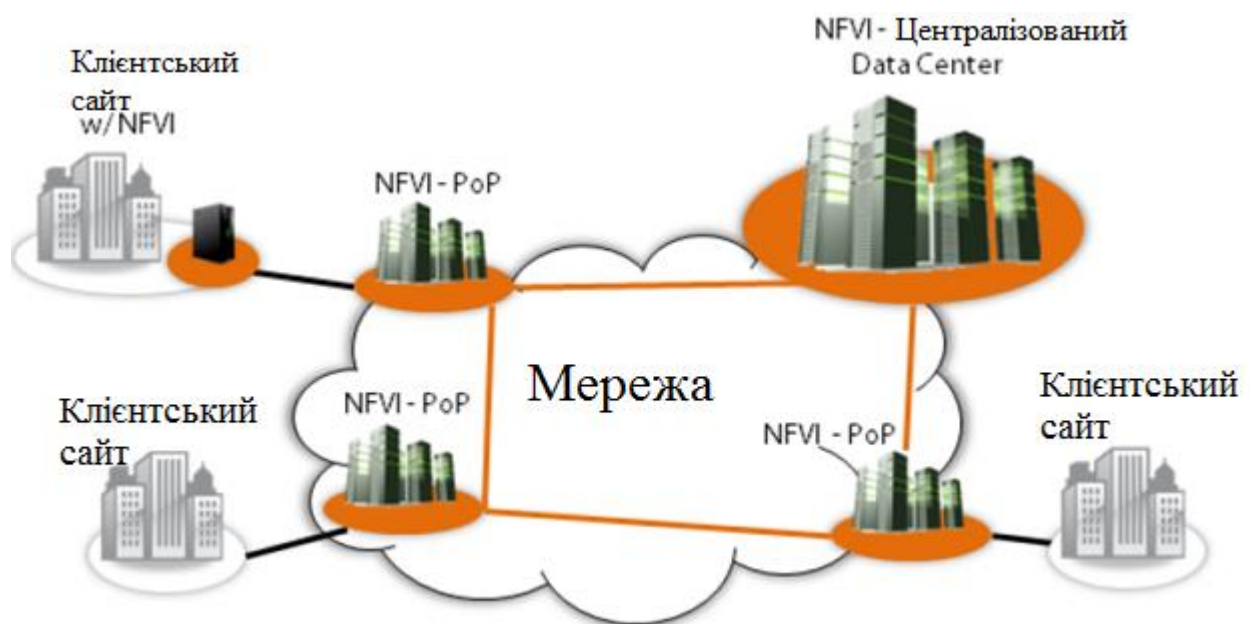


Рисунок. 2.3 Інфраструктура NFV

Попит на таку розподілену інфраструктуру пов'язаний з вимогами, що пред'являються різними NF. Наприклад, існують певні функції, які повинні знаходитися в певному географічному розташуванні, а інші - ні. Наприклад:

1. Блок базової смуги RAN має суворі вимоги щодо затримки радіоантени, що робить PoP найкращим підходящим кандидатом для розміщення цієї функції;

2. Деякі функції CPE, такі як телевізійні приставки, повинні бути ближче до сайту замовника, в PoP;
3. ANDSF може бути розташований в централізованому DC.

Зосередимося на двох основних аспектах інфраструктури NFV, які ми розглянемо в звіті: NFVI (сама платформа фундаменту, що включає апаратне забезпечення, операційні системи, гіпервізор і т. Д.) і VIM (компонент управління ресурсами NFVI).

Після того як фізичні мережеві функції перетворюються в програмне забезпечення, тобто віртуальні мережеві функції (VNF), їм потрібні ресурси для запуску. Інфраструктура NFV (NFVI) складається з фізичних та віртуальних обчислень, сховищ і мережевих ресурсів, які будуть використовувати VNF.

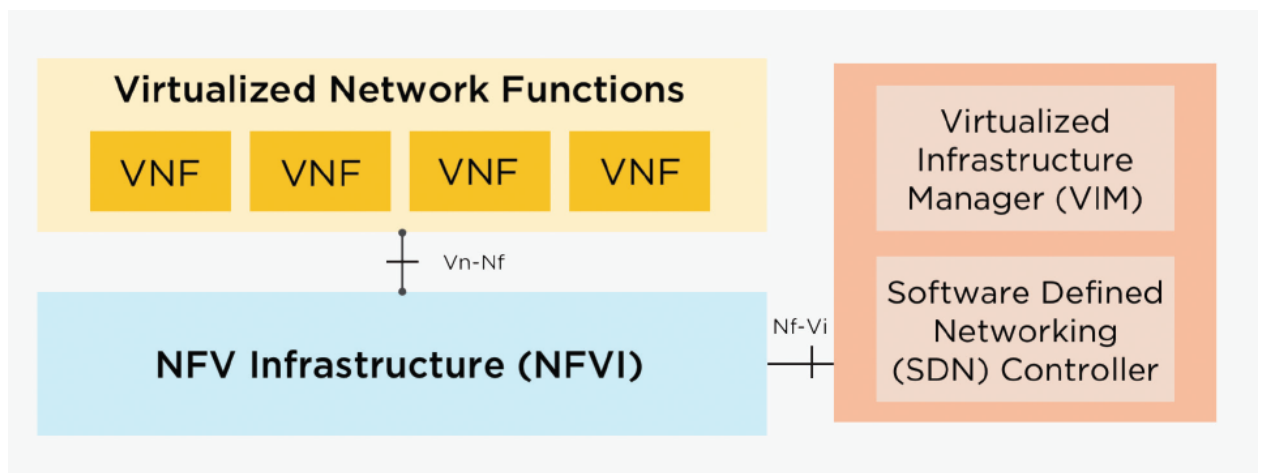


Рисунок. 2.4 NFVI в рамках NFV

Рівень NFVI в основному взаємодіє з двома іншими компонентами інфраструктури NFV: VNF і Virtualized Infrastructure Manager (VIM). Програмне забезпечення VNF працює на NFVI. VIM, з іншого боку, відповідає за надання та управління віртуальною інфраструктурою. Інтерфейс VNF для NFVI (Vn-Nf) являє собою datapath, через який проходить мережевий трафік, тоді як інтерфейс NFV для VIM (Nf-Vi) являє собою шлях управління, який використовується виключно для управління, але не для

будь-якого мережевого трафіку. NFVI складається з трьох різних шарів: фізичної інфраструктури, рівня віртуалізації і віртуальної інфраструктури.

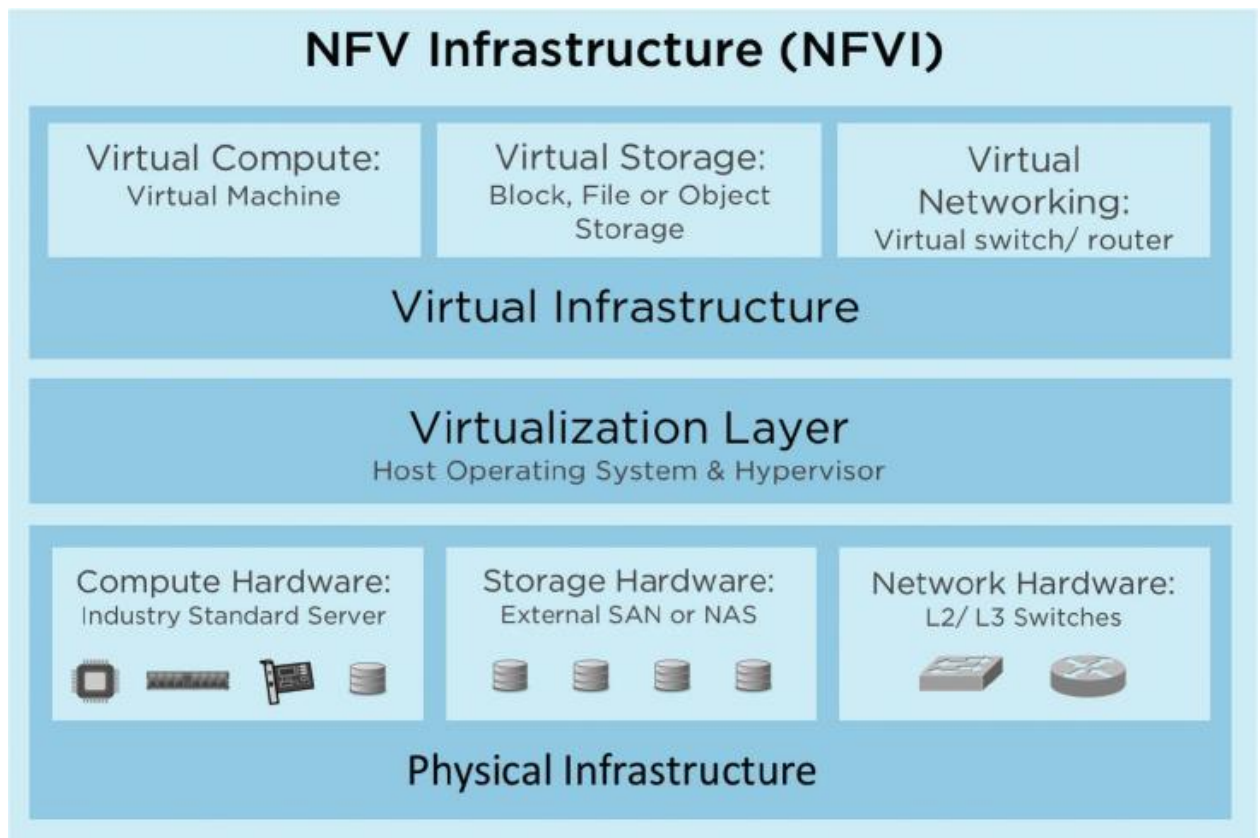


Рисунок. 2.5 Архітектура NFVI високого рівня

2.3.1 NFVI - Обладнання (дата-центр, edge і vCPE)

Устаткування NFVI складається з обчислювального, зберігального і мережевого устаткування. Ми розглянемо кожен окремо більш детально.

Обчислювальне обладнання: промислові стандартні сервери:

Ключовою перевагою NFV є перехід окремих ОДНОБЛОКОВОМУ ящиків на стандартні сервери. Існує величезна різноманітність варіантів, коли мова заходить про ці стандартних серверах, спочатку очікують більш одного.

Серверні стійки

Найбільш поширеними при використанні стандартних серверів є серверні стійки 1U або 2U. Сьогодні для багатьох розгортань NFV CSP як і раніше використовують фірмові сервери від таких великих виробників, як HPE, Dell EMC, Lenovo, Cisco, Huawei і IBM. Проте, завдяки зусиллям Open Compute Project (OCP), великі виробники оригінального дизайну (ODM) відправляють сертифіковані стійки, які добре працюють для хмарних інфраструктур, які часто добре працюють для робочих навантажень NFV (з винятком, що більшість хмарних робочих навантажень мають менше потреб введення-виведення).

Більшість архітектур NFVI мають тенденцію вибирати стандартну систему COTS (комерційний сервер) і створювати віртуальну хмарну інфраструктуру, що відтворює стійки цих серверів. Все частіше ці сервери також оцінюються за можливостями їх чистого металу та ємності контейнерів, оскільки ми бачимо більш змішані середовища.

Сьогоднішні галузеві стандартні сервери були синонімом архітектури Intel, хоча останнім часом спостерігається деяка активність навколо архітектур ARM для хмари і NFV, особливо з урахуванням недавнього оголошення Microsoft і Qualcomm. ARM пропонує різні характеристики з точки зору продуктивності, вартості, інтерфейсів, доступного програмного забезпечення і ще багато чого іншого. З іншого боку, OpenPOWER і AMD поки не проявили великого інтересу до NFV.

Блейд-шасі

Для багатьох робочих навантажень NFV, які пов'язані перш за все з обробкою пакетів і додатками, тобто в основному обчислювальними і мережевими робочими навантаженнями, без особливих потреб в сховищі, блейд-сервери розглядають деякі CSP як варіант з більш високою щільністю. Сервери Blade об'єднують кілька обчислювальних вузлів в одне шасі. Вони дбають про мережеве, енергетичне, охолоджуюче і апаратне управління для всього набору обчислювальних вузлів. Розмір цих серверів може становити

від двох до чотирьох обчислювальних вузлів в 2U до шістнадцяти серверів в 16U.

Рішення для Hyperconverged Infrastructure (HCI)

Для CSP, які шукають більше простоти, ніж блейд-сервери, і знаходять спільне обчислення, сховище та мережеву взаємодію, рішення HCI може бути доречним. Ці рішення об'єднують обчислення, сховища і мережі в один ящик і створюють більше можливостей, просто розширюючи їх, додаючи більше ящиків (в цьому сенсі вони більш гнучкі, ніж блейд-сервери). Гіперконвертоване (HCI) рішення може включати в себе додаткові функціональні можливості з доданою вартістю, такі як висока доступність, резервне копіювання, відновлення після збою, безпеку і багато чого іншого. Зручність цього рішення пов'язана з низькою гнучкістю в плані розгортання, налаштування, масштабування і поновлення різних рівнів незалежно. Це може привести до недовикористання обладнання та проблем з продуктивністю. Для робочих навантажень NFV існують великі постачальники CSP, які починають надавати конвергентні та гіперконвергентні робочі навантаження і починають налаштовувати їх для більшої кількості мережевих послуг з підтримкою введення-виведення.

Edge Devices і Vcpe

Розглянуте досі обчислювальне обладнання підходить для централізованої хмари. Проте, часто доводиться використовувати edge обчислення (також звані Fog Computing), де потрібно або сервер рівня несучої (наприклад, в центральному офісі), або невелика коробка для розміщення всього декількох віртуальних машин (наприклад, будинок для споживачів). На ринку доступні як сервери операторського класу, так і пристрої з більш легкою вагою та з різними варіантами підключення, включаючи Ethernet, Wi-Fi, LTE, DSL, GPON, TDM і Metro Ethernet. Ці пристрої можуть також включати спеціалізоване апаратне прискорення, наприклад GPU для транскодування відео або машинного навчання. Крім того, в цих edge місцях розташування більш низьке енергоспоживання і

щільність може бути краще сирі обчислювальної потужності; в даний час розглядаються такі альтернативні архітектури від стандартної архітектури Intel.

У житлових і комерційних розгортаннях клієнтський edge має морфінг з декількох виділених ящиків, які надають різні функції (переривання лінії, міжмережевий екран, маршрутизація, оптимізація WAN, IP PBX і т. д.) у базові платформи Virtual CPE (vCPE), які використовують програмне забезпечення для віртуалізації, включаючи різновидності гіпервізорів, ОС з підтримкою Linux Container, розділених ОС з можливістю установки пакетів VNF. В даний час розробляються і тестуються різні підходи в поєднанні з різними архітектурами (VNF можуть працювати в регіональному центрі обробки даних, в центральному офісі або на самому пристрої CPE). Також тестуються відповідні бізнес-моделі, котрі в тому числі мають CSP, які здають такі NFVI vCPE платформи своїм клієнтам, з можливістю оренди програмного забезпечення на вимогу (безпека, прискорення і т. д.) або ж придбання підприємствами цих платформ і оренди OTT послуг в хмарі.

Апаратне прискорення площини даних

Велика частина NFV стосується обробки пакетів, і продуктивність має першочергове значення. Продуктивність вимірюється за допомогою таких параметрів, як пропускна здатність пакетів, біт / сек, джиттер, обробка затримки і т. д., причому перші два є найбільш поширеними. Підвищення продуктивності апаратного забезпечення NFV здійснюється за допомогою двох механізмів: А) удосконалень процесора і В) апаратного розвантаження. Удосконалення процесора включає SR-IOV (де трафік NIC може обійти гіпервізор і перейти безпосередньо до віртуальної машини), величезна підтримка сторінки (скорочення пошуку), прив'язка процесора (прив'язка віртуальної машини до певного набору ядер) і т. д. Апаратне розвантаження поставляється у вигляді інтелектуальних мережевих адаптерів, які мають розвантаження для віртуального перемикачів, шифрування, стиснення або

віддзеркалення портів. Або це може бути FPGA, що діє як співпроцесор для прискорення захисту або стиснення.

Устаткування для зберігання даних: SAN, NAS або промислові стандартні сервери

У той час як більшість робочих навантажень NFV зосереджені на обчислювальних і мережевих потребах, деякі функції, такі як кешування (відео та інше) або аналітика і велике сховище даних, зажадають швидкого і великого резерву зберігання. Устаткування для зберігання даних поставляється у вигляді сховища з прямим підключенням (тобто сховища в обчислювальному вузлі) або традиційного сховища SAN (мережа зберігання даних) або NAS (мережеве сховище даних), яке може бути вирізано як віртуальне сховище для віртуальних машин. В якості альтернативи, програмне забезпечення з встановленим сховищем (SDS) може використовуватися там, де апаратне забезпечення є стандартним сервером, оптимізованим для зберігання. Типовим сервером зберігання може бути блок 2U або 4U з великою кількістю (36, 48, 60 +) твердотільних дисків (SSD) або жорстких дисків (HDD) і набагато меншою обчислювальною потужністю і пам'яттю в порівнянні з обчислювальним сервером (наприклад, один роз'єм з пам'яттю 64 ГБ). Сервери зберігання можуть бути підключені до зовнішніх JBODs / JBOF (просто набір дисків / флеш-накопичувачів) для додаткової ємності.

Мережеве обладнання (власні або відкриті комутатори)

Мережеве обладнання поставляється у вигляді власних комутаторів L2 / L3 або bare-metal комутатори. Комутатори Bare-metal призначені для підключення до комп'ютера, на якому промисловий стандартний сервер повинен обчислювати або зберігати. Ці комутатори використовують загальнодоступні комутатори ASIC і дезагрегують апаратне забезпечення з програмного забезпечення. Вони запускають програмне забезпечення, таке як Open Network Linux або інші мережеві комутатори сторонніх виробників.

2.3.2 NFVI - рівень віртуалізації

Рівень віртуалізації NFVI знаходиться поверх апаратного забезпечення і являє собою програмну платформу, яка зазвичай включає гіпервізор. Однак, оскільки контейнери стають життєздатними для розгортання виробництва NFV, рівень віртуалізації NFVI, швидше за все, буде розширюватися, включаючи операційні системи, які є механізмами часу роботи контейнера.

Гіпервізори поділяють ресурси фізичної машини і пропонують еквівалент фізичної машини для додатка. Насправді три основні функції гіпервізора: А) поділяють ресурси фізичної машини, В) забезпечують ізоляцію між різними віртуальними машинами (це робиться за допомогою CPU) і С) наслідують всі необхідні периферійні пристрої, наприклад, карти NIC. Основними гіпервізорами, що використовуються для NFV, є VMware vSphere, KVM і KVM в реальному часі (є й інші технології VM від таких постачальників, як Microsoft і Oracle):

- vSphere: vSphere - це власний гіпервізор VMware. Він дуже зрілий (15 років з моменту першого випуску) і стабільний. Використовується для робочих навантажень, які не є хмарними. vSphere - це те, що називається гіпервізором Type 1 або bare-metal, оскільки він працює безпосередньо на обладнанні.

- KVM (віртуальна машина на основі ядра): KVM - це зрілий 10-річний проект гіпервізора. Це гіпервізор з відкритим вихідним кодом типу 2 або розміщений поверх операційної системи Linux (так званої операційної системи хоста, щоб відрізнити її від операційної системи, яка працює у віртуальній машині, званої гостьовою ОС). З цієї причини при використанні KVM вам також необхідно вибрати ОС хоста. Зазвичай використовуються операційними системами для хост-систем є RHEL, SUSE або Ubuntu.

- KVM в реальному часі: цей гіпервізор з відкритим вихідним кодом є варіантом KVM, який забезпечує детерміноване поведіння. Незважаючи на

те, що ви очікуєте, що в основному це допоможе з такими параметрами, як джиттер, було показано, що він значно покращує пропускну здатність невеликого пакета (Nokia продемонструвала це на недавньому саміті OpenStack).

2.3.3 NFVI - Віртуальна інфраструктура

Віртуальна інфраструктура складається з віртуальних машин, віртуального сховища і віртуальних мереж.

Віртуальні машини

Віртуальні машини (VM) створюються гіпервізором, як обговорювалося вище. Гіпервізори представляють API для створення, знищення, міграції та управління віртуальними машинами. У разі KVM це робиться через бібліотеку libvirt, а для vSphere - через диспетчер віртуальної машини, так званий vCenter. Ці віртуальні машини - це те, що розміщуються всередині VNF, і які діють як інструмент, що забезпечує ядро мережевих сервісів концепції NFV.

Віртуальне сховище

Відповідальність за віртуалізацію блочного або файлового сховища лежить на програмному забезпеченні SAN, NAS або SDS, а віртуальна машина просто представлена з номером логічного пристрою (LUN) або файлом. Рівень віртуалізації може додавати додаткові функції, такі як моментальні знімки, резервні копії, тонке надання ресурсів, спільне використання VM, реплікація і т. д. У разі зберігання об'єктів, найбільш придатних для масового зберігання, він вже віртуалізований і відкритий через API REST, тому нічого особливого не потрібно щоб зробити це.

Віртуальна мережа

Гіпервізор містить віртуальний комутатор або маршрутизатор, який має чотири функції: А) перемикання між віртуальними машинами в одному і

тому ж фізичному вузлі без необхідності переходу на зовнішній комутатор, найбільш часто використовується для створення ланцюгів сервісних функцій; В) надання мережевих послуг накладання, де мережі, що піддаються впливу віртуальних машин, повністю абстрагуються з власним адресним простором і топологіями, відмінними від фізичної або основної мережі; С) надання служб безпеки, наприклад, примусове виконання списків контролю доступу; D) дозволяють підключати шлюзи до Інтернету. У NFV використовується безліч технологій віртуальних мереж. Деякі з поширених відкритих технологій:

- OVS або OpenVSwitch: OVS - це стабільний проект віртуального комутатора, який був запущений в 2009 році. Це багаторівневий комутатор, який може охоплювати гіпервізор (через обчислювальні вузли) і включає в себе безпеку, моніторинг, накладення мережевої інкапсуляції, зв'язок NIC і т. д. OpenFlow є основним методом програмування OVS, а також OVSDB для конфігурації, хоча підтримуються і додаткові протоколи. Більшість контролерів SDN (наприклад, ODL, OVN, ONOS) підтримують OVS.

- DVR (Distributed Virtual Router): розподілений маршрутизатор в OpenStack, доповнює OVS, надаючи функції маршрутизації і шлюзу, використовуючи повністю розподілену архітектуру.

- vRouter: OpenContrail - це один SDN-контролер з відкритим вихідним кодом, який не підтримує OVS. Замість цього для цього потрібно власне віртуальне мережеве програмне забезпечення під назвою vRouter. VRouter також підтримує маршрутизацію і може бути запрограмований через BGP.

- FD.io: Fast Data Project (FD.io) - це високопродуктивна заміна OVS. Він використовує механізм обробки векторних пакетів (VPP) для паралельної обробки пакетів, щоб домогтися поліпшення продуктивності 5x-39x в порівнянні з OVS, хоча є версія OVS з DPDK, яка працює значно краще, ніж запаси OVS.

Віртуальне мережеве рішення для VMware:

- VDS: розподілений комутатор vSphere дуже схожий по функціональності на проект OVS з відкритим вихідним кодом, за винятком

того, що він тісно інтегрований з vSphere і vCenter. Ця схожість не повинна дивувати, тому що OVS і VDS були спочатку створені командою Nicira, тепер є частиною VMware.

Рівень віртуальної мережі також є домом для технологій прискорення площини даних, таких як:

- DPDK: DPDK - це набір програмних бібліотек, які обходять ядро, таким чином уникаючи дорогого режиму ядра для перемикача призначеного для користувацького режиму. DPDK також використовує режим опитування замість переривань, щоб запобігти безлічі непотрібних і трудомістких контекстних перемикачів. У дослідженні, проведеному Intel, OVS з DPDK продемонструвала 75% -ве збільшення пропускної здатності порівняно з OVS без DPDK. DPDK був просто переміщений в Linux Foundation і тепер розроблений декількома постачальниками за межами Intel;

- ODP (OpenDataPlane): цей проект групою Linaro Networking створює стандартні API для прискорення мережевої площини даних. Обґрунтування полягає в тому, що розробники додатків не хочуть використовувати власні API, але замість цього будуть використовувати відкриті API. Під API постачальники обладнання можуть впроваджувати інновації за допомогою різних апаратних прискорень.

- IO Visor: в певному сенсі IO Visor є протилежним DPDK, оскільки він знаходиться в ядрі, а не обходить його. Він має дві підтехнології, які прискорюють обробку пакетів в ядрі Linux. EBPf - це програмована вбудована віртуальна машина для розширення функціональності ядра. XDP або розширений шлях передачі даних використовують eBPf для створення програмованого високопродуктивного пакетного процесора для розширення стека Linux Networking. На додаток до eBPf і XDP доступні кілька плагінів, інструменти для розробки і управління.

2.3.4 NFVI - Майбутні тенденції

NFV - дуже молода сфера з високою швидкістю інновацій. Деякі з найбільш перспективних тенденцій NFVI, які очікуються.

Апаратні засоби

Нові спеціалізовані процесори введення-виведення, які націлені на NFV, швидше за все, з'являться. Зрозуміло, тяга цих технологій залежатиме від величини вартості, яку вони забезпечують з точки зору підвищення продуктивності і розвантаження ядра процесора. Mellanox Bluefield і SmartNIC від Neterion є ранніми прикладами процесорів вводу-виводу.

Рівень віртуалізації

Рівень віртуалізації переживає відродження. На рис 2.6 показані кілька цікавих майбутніх альтернатив.

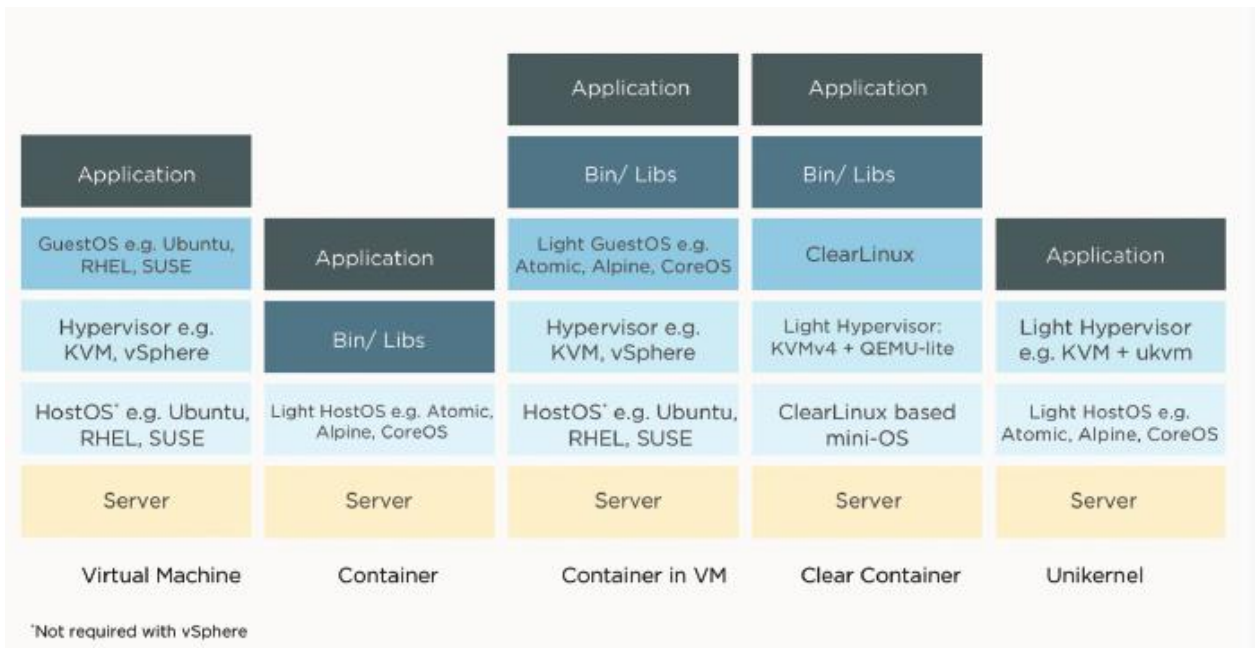


Рисунок.2.6 Майбутні альтернативи рівня віртуалізації

Поточний рівень віртуалізації за замовчуванням складається з гіпервізору, що створює віртуальні машини (самий лівий елемент на наведеному вище малюнку). З'являються чотири нові моделі. Ми порівняємо

ці майбутні технології з традиційною моделлю VM і подивимося, які їхні сильні і слабкі сторони в табл.2.1.

Таблиця.2.1 Порівняння майбутніх технологій з традиційною моделлю VM

Технологія	Опис	Розмір зображення	Час завантаження
Virtual Machine	Для програми потрібна гостьова ОС, яка, в свою чергу, працює з гіпервізором, який працює на ОС хоста. В даний час це технологія за замовчуванням в NFVI.	ГБ	10-100 с
Container	Контейнери використовують віртуалізацію ОС замість віртуалізації машин. В результаті виходить набагато легший VNF, так як всі контейнери в обчислювальному вузлі спільно використовують ядро. Контейнери більш доброзичливі до архітектури додатків DevOps і мікросерверів. Базова ОС також може бути спрощена, що призводить до створення більш легких контейнерних ОС, наприклад, Atomic, Alpine, CoreOS або Clear Linux. Менше зображення контейнера означає, що щільність обчислювального примірника на вузол може збільшитися до 10х. Незважаючи на те, що контейнери вивезли на	МБ	100 мс

	підприємство, вони		
--	--------------------	--	--

Продовження таблиці 2.1

Технологія	Опис	Розмір зображення	Час завантаження
	дещо проблематичні в NFV. Спільне використання ядра створює проблеми ізоляції і безпеки. Контейнерні структури оркестровки, також ігнорували такі теми, як надійні мережеві моделі і розрахований на багато користувачів режим. Незважаючи на сильне бажання запуску VNF в контейнерах, ці проблеми необхідно вирішити в першу чергу.		
Container in VM	У цьому підході контейнери запускаються в віртуальній машині. Цей підхід забезпечує переваги ізоляції віртуальної машини і переваги DevOps / cloud для контейнера. Однак продуктивність схожа на VM, можливо, навіть трохи гірше. Цей підхід є дуже поширеним способом запуску контейнерів в NFV і навіть в корпоративних додатках сьогодні.	ГБ	10-100 с

Clear Containers	Clear Containers є частиною проекту Clear Linux, створеного	МБ	100 мс
------------------	---	----	--------

Продовження таблиці 2.1

Технологія	Опис	Розмір зображення	Час завантаження
	Intel. Ідея полягає в тому, щоб запускати контейнер у віртуальній машині, але різко скорочувати оболонку віртуальної машини. Clear Containers також використовують контейнерні дієслова, щоб вся конструкція VM відображалася як контейнер, а не віртуальна машина. Оболонка VM - це легкий хост (Clear Linux), KVM і легша версія QEMU. Час завантаження і розмір зображення ближче до розміру контейнера. Оскільки Clear Containers надають кращі контейнери і віртуальні машини, вони можуть бути дуже цікавою розробкою для NFV.		(~25-30% вище, ніж у контейнерів)
Unikernels	Unikernels - віртуальні машини. Однак, замість того, щоб перетягувати всю гостьову ОС, Unikernels пов'язують додаток	МБ	10 мс

	тільки з бібліотеками, використовуваними додатком. Таким чином, зображення дуже маленькі, схожі на розміри контейнерів. Час завантаження		
--	--	--	--

Продовження таблиці 2.1

Технологія	Опис	Розмір зображення	Час завантаження
	навіть швидше, ніж у контейнерів (на порядок більше). При невеликих розмірах зображень і часу завантаження, які знаходяться в діапазоні часу проходження пакета (RTT), для NFV можуть бути досить цікаві одноядерні процесори. Недоліком є те, що кожен VNF доведеться перекомпілювати.		

2.4 Висновки з розділу 2

Деякі оператори вимагали більш швидкого і органічного розвитку платформ NFV. Деякі оператори навіть склали свої власні технологічні програми, в тому числі розробляють свої власні проекти з відкритим вихідним кодом, які потім передають спільноті (наприклад, OSM / OpenMANO від Telefonica, ONAP, що включає ECOMP від AT & T і Open-O від China Mobile).

Рівень NFVI складається з програмного забезпечення для апаратного забезпечення і віртуалізації, яким керують VNF. Цей рівень - це просто канал передачі даних і не має відношення до будь-яких аналізу планування, організації, забезпечення, моніторингу, обслуговування і т. д.

РОЗДІЛ 3 АНАЛІЗ ВПРОВАДЖЕННЯ ТЕХНОЛОГІЙ ВІРТУАЛІЗАЦІЇ В ОПОРНІЙ МЕРЕЖІ ОПЕРАТОРІВ ЗВ'ЯЗКУ

3.1 Ericsson Virtual Evolved Packet Core

Virtual Evolved Packet Core надає перевірені рішення, орієнтовані на велику кількість вертикальних додатків, відкриваючи тим самим нові можливості оператора.

Переваги віртуального Evolved Packet Core від Ericsson включають всі переваги NFV, але з диференціацією Ericsson - повне комплексне рішення, що означає віртуалізацію всіх компонентів Evolved Packet Core. Підтримка гнучкого розгортання цілої низки віртуальних мережевих сервісів, підтримуваних vMME, vSGSN, vPGW, vSGW, vGGSN, vPCRF, vDPI, vProbe, vePDG і vTWAG. Забезпечення повної сумісності з власним Evolved Packet Core підтримуючи сумісність від оточуючих пристроїв і RAN до зарядних систем і сервісів. Власні та віртуальні мережеві вузли будуть безперешкодно співіснувати в таких областях, як об'єднання, резервування геоданих і розподіл навантаження.

Ericsson Evolved Packet Core надає перевірені рішення, призначені для великої кількості вертикальних прецедентів. До основних початкових послуг віртуальної мережі відносяться:

- Інтернет речей
- Розподілений мобільний широкопasmовий зв'язок
- Виробництво
- Зв'язок (виклик VoLTE і Wi-Fi)
- Оператор мобільного віртуальної мережі (MVNO)
- Мобільний широкопasmовий доступ.

Віртуальні мережеві сервіси можуть використовуватися в якості мережевих частин, що забезпечують гнучкість, дозволяючи створювати безліч логічних мереж поверх загальної фізичної інфраструктури та в якості

кроку до архітектури ядра 5G. Як приклад він може використовуватися для швидкого оптимізованого введення-виведення без впливу на послугу мобільної широкопasmової мережі.

Повний портфель рішень Evolved Packet Core від Ericsson забезпечує безпрецедентну масштабованість і гнучкість від невеликих локальних розгортань з використанням ЕРС в коробці до широкомасштабних розгортань центрів обробки даних. Це означає, що віртуальне Evolved Packet Core може бути розгорнуто в великих централізованих центрах обробки даних, але також розподілено поруч з радіомережею.

Evolved Packet Core забезпечує плавні, безшовні шляхи міграції з унікальним співвідношенням функцій на власних і віртуалізованих мережевих вузлах. Все для підтримки спрощених операцій і швидкого часу для обслуговування мережевих сервісів.

Успішне розгортання віртуалізованої функції залежить від повноти рішення, а також від відкритості рішення. Ericsson Evolved Packet Core підтримується в системі Ericson Cloud System, а також в перевірених середовищах виконання третьої сторони.

3.2 Huawei CloudCore

CloudCore - це рішення для мережевої віртуалізації, яке змінює архітектуру базової мережі NE зі спеціалізованих апаратних засобів до загального обладнання через Network Virtualization (NFV), відокремлюючи програмне забезпечення від апаратного забезпечення. CloudCore забезпечує віртуалізовану функцію управління NE для автоматичного розгортання ресурсів і планування, підвищення ефективності мережі. Крім того, CloudCore забезпечує гнучке налаштування обслуговування і міжмережеву взаємодію, скорочуючи ТТМ. Huawei CloudCore підтримує віртуалізацію всіх

мережевих елементів. Постачальники можуть надавати послуги FMC для будь-яких абонентів і терміналів через CloudCore.

Рішення Huawei CloudCore має три функції: удосконалені продукти, засновані на хмарі, просте розгортання та ефективна робота.

Удосконалені продукти, засновані на хмарі:

1. Відкритість: підтримка апаратного обладнання та CloudOS;
2. Висока продуктивність: кращі можливості передачі з високою швидкістю, продуктивність віртуалізації краще, ніж у виробництва;
3. Висока доступність: 99,999% доступність.

Простота розгортання:

1. Служба інтеграції E2E;
2. Повний набір інструментів для підвищення ефективності інтеграції;
3. Повна економічна система партнера для забезпечення оптимізації за межами майданчика.

Ефективна робота:

1. Швидке розгортання служби в годинах на основі MANO;
2. E2E усунення несправностей;
3. Підтримка єдиної операції, включаючи хмарну і традиційну платформу.

3.2.1 Service Oriented Core (SOC)

Service Oriented Core (SOC), рішення Huawei по 5G Core Network, допомагає операторам створювати абсолютно нові мережеві архітектури та бізнес-моделі, відкриваючи нові можливості для ринку і домагаючись успішних цифрових перетворень.

Великі оператори в даний час активно прагнуть зробити цифрове перетворення. Нові технології, такі як SDN, NFV і MEC, використовуються для побудови мережі наступного покоління 5G, що призводить до появи

нових бізнес-моделей, таких як B2B, B2B2C і B2V. Це відкриває нові можливості та виклики для операторів і галузі в цілому. Мережі 5G потребують сильної Core Network для підтримки постійно мінливих сервісів в майбутньому. SOC забезпечує весь доступ і всі послуги, що виникають, в міру необхідності.

Забезпечення доступу усіх користувачів і всіх служб за допомогою 5G Core Networks

Для операторів SOC забезпечує доступ і реалізує всі мобільні режими доступу 2G / 3G / 4G / 5G і інші режими доступу, такі як доступ xDSL, Wi-Fi і LPWAN, через одну базову мережу, тим самим допомагаючи операторам максимізувати ROI і розширити масштаб зв'язків.

Крім того, SOC використовує архітектуру Cloud Native для надання мережевих частин, які відповідають різноманітним і диференційованим вимогам до обслуговування через одну фізичну мережу. Наприклад, для моніторингу інтелектуального середовища потрібно низьке енергоспоживання з високою щільністю з'єднання, послуги AR / VR вимагають високої пропускної здатності, а автономне водіння вимагає надзвичайно низькою затримки. SOC допомагає досліджувати вертикальні галузеві ринки і підвищує інновації в сфері послуг, щоб домогтися успіху в майбутньому.

Нові послуги 5G від концепції до реальності від SOC

Huawei і China Mobile спільно продемонстрували функції, дизайн і бізнес-логіку мережевих частин, а також те, як оператори використовують мережеві частини на основі SOC для задоволення різних вимог до мережі великою кількістю додатків в майбутньому.

Також Huawei і Vodafone продемонстрували розумне водіння на основі SOC, а Huawei і Deutsche Telekom представили три наскрізні орієнтовані на 5G мережеві частини, засновані на ОС 5, для комерційного використання: частина xMBB для сервісів AR / VR, GLA частина для IoV і промислового

контролю, а також частина FMC для забезпечення безперервного обслуговування відео.

3.2.2 SOC 2.0 будівництво 5G Core Networks

Після того, як Huawei випустила перше Core Networks рішення 5G Service Oriented Core (SOC) 2.0.

У майбутньому телекомунікаційні мережі будуть не тільки надавати користувачам послуги зв'язку, але також забезпечувати зв'язок і послуги для різних галузей, таких як автономне керування автомобілем, промисловий контроль, AR / VR і розумне місто. В якості базового рішення наступного покоління 5G SOC2.0 має наступні функції:

1. Включення всього доступу: підтримує кілька режимів мережевого доступу і забезпечує підключення до безлічі терміналів з різних галузей.
2. Надання всіх послуг: надання послуг на вимогу для вертикальних галузей.

Huawei SOC 2.0 використовує такі методи для реалізації майбутньої базової мережі, що забезпечують всі доступи і всі послуги:

- Управління та поділ користувальницьких площин (CUPS): площину управління і користувацька площина відділені розподіленням розгортанням. Це важлива основа для SOC 2.0 для реалізації гнучкої архітектури. Мережеві функції розгортаються в центральних DC і локальних DC на основі вимог до обслуговування або розгортаються на периферійних DC, які ближче до абонентів, використовуючи технологію Mobile Edge Computing (MEC).

- Архітектура, заснована на обслуговуванні (SBA): мережеві елементи розподілені для створення незалежних і модульних функцій. Завдяки методам обслуговування на вимогу ці функції організовані в єдиній архітектурі відповідно до вимог до обслуговування. Підтримуються гнучкі режими множинного доступу і вимоги до сервісу. Кожна функція може бути незалежно повторена і оновлена для швидкого задоволення нових вимог сервісу.

- Нарізка мережі: щоб включити всі служби, мережа повинна забезпечувати взаємно ізольовані труби для обслуговування кожної вертикальної галузі. Мережева нарізка на основі хмарних обчислень є попередньою умовою для інтелектуального розподілу сервісних ресурсів і можливостей, пов'язаних з обслуговуванням. Таким чином, оператори можуть надавати диференційовані послуги в різних сценаріях.

- Гнучкість доступу: дозволяє операторам використовувати кілька режимів доступу, щоб забезпечити відмінний сервіс. Наприклад, коли пропускна здатність мережі недостатня в зонах «гарячих точок», оператори можуть пов'язувати мобільні мережі і сигнали Wi-Fi для забезпечення високошвидкісного доступу до Інтернету.

Розвиток базової мережі 5G складається з трьох етапів:

1. По-перше, створення хмарної платформи, орієнтованої на DC, розгортається хмарні VNF, впровадження розгортання з використанням проміжного DC і без фіксації дизайну та перенесення традиційних мережевих сервісів на хмарну платформу.

2. По-друге, впровадження CUPS і використання MEC для створення розподілених мереж для забезпечення низької затримки.

3. SBA, мережева нарізка і доступність для надання послуг 5G на вимогу.

В даний час більше 70% світових операторів почали хмарну реконструкцію мережі.

3.3 Впровадження 5G

Три суттєвих постачальника телекомунікацій Південної Кореї, SK Telecom, KT і LG Uplus, офіційно розгорнули перші в світі комерційні поставки п'ятого покоління 1 грудня 2018 року запусивши нове обладнання та відправивши в мережах нового покоління свої перші 5G сигнали в Корею

разом. Мобільні оператори спочатку планують поставити послуги 5G за допомогою мобільних маршрутизаторів - пристроїв, що дозволяють підключення до пристроїв Wi-Fi - а не смартфонів. 5G послуги для смартфонів стануть доступними в березні наступного року. Обіцяна швидкість завантаження до 20 гігабіт в секунду та час відгуку менше ніж 0,0001 секунди. Оскільки 5G може обробляти великі обсяги даних з наднизьким затримкою, він стає основою багатьох послуг, що вимагають безперервного підключення.

SK Telecom планує розгорнути послуги 5G у столичному районі Сеула, а також у шести найбільших містах Кореї: Пусан, Інчхон, Тэгу, Тэджон, Ульсан і Кванджу, при цьому мережа стає доступною першим корпоративним клієнтам у виробничому секторі. Першим клієнтом 5G мобільного оператора є компанія Myunghwa Ind., виробник автомобільних компонентів, яка прийме рішення SKV 5G-AI для машинного бачення, яке буде збирати та відправляти фотографії високої чіткості з виробничої лінії на сервер хмар, де їх буде переглянуто на дефекти через штучно інтелектуальні алгоритми. Щоб забезпечити максимальну безпеку, SKT також планує вперше розгорнути свою систему квантової криптографії в районі Сеул-Ансан, а потім застосувати цю технологію до серверів клієнтів. Фірма також може похвалитися мережевою технологією AI, яка забезпечує швидку передачу даних та якісне обслуговування.

Аналогічним чином LG Uplus буде розгорнути послуги 5G в столичному регіоні Сеула і вибрати великі міста в Кореї і починати пропонувати пакети мережевого маршрутизатора 5G і плани платежів Wi-Fi 5G. Оператор мобільного зв'язку планує встановити більш ніж 7000 5G мережевих станцій по всій Кореї до кінця грудня та розширити свою 5G-інфраструктуру до великих міст, перш ніж запуснути сервіси 5G з підтримкою смартфонів у березні наступного року.

KT, який все ще оговтується від наслідків пожежі, яка заплямував репутацію мобільного оператора і викликала стурбованість з приводу

стабільності мережі, також прагне продовжити заявлені зусилля з надання стабільних послуг 5G тут. Компанія планує провести внутрішню подію в Центрі інтелектуальної мережевої контролю КТ у м. Гвачон, провінція Кенджі, щоб відзначити своє перше комерційний розгортання 5G.

3.4 Висновки з розділу 3

Однією з проблем для операторів зв'язку в розгортанні NFV є забезпечення продуктивності користувацької площини. У рішеннях NFV від Huawei робляться численні заходи для оптимізації продуктивності апаратних, хмарних ОС і шарів VNF, що в кінцевому підсумку підвищує продуктивність користувацької площини у три рази.

Huawei і China Mobile спільно продемонстрували функції, дизайн і бізнес-логіку мережевих частин, а також те, як оператори використовують мережеві частини на основі SOC для задоволення різних вимог до мережі великою кількістю додатків в майбутньому.

Також Huawei і Vodafone продемонстрували розумне водіння на основі SOC, а Huawei і Deutsche Telekom представили три наскрізні орієнтовані на 5G мережеві частини, засновані на ОС 5, для комерційного використання: частина xMBB для сервісів AR / VR, GLA частина для IoV і промислового контролю, а також частина FMC для забезпечення безперебійного обслуговування відео.

В даний час більше 70% світових операторів почали хмарну реконструкцію мережі. Компанія Huawei і її партнери по галузі планують продовжити просування промислового розвитку NFV вперед, щоб допомогти більшій кількості компаній трансформуватися в хмару.

Три суттєвих постачальника телекомунікацій Південної Кореї, SK Telecom, KT і LG Uplus, офіційно розгорнули перші в світі комерційні поставки п'ятого покоління 1 грудня 2018 року запустивши нове обладнання

та відправивши в мережах нового покоління свої перші 5G сигнали в Корею разом. Мобільні оператори спочатку планують поставити послуги 5G за допомогою мобільних маршрутизаторів.

Обіцяна швидкість завантаження до 20 гігабіт в секунду та час відгуку менше ніж 0,0001 секунди.

РОЗДІЛ 4 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВІРТУАЛЬНОЇ МЕРЕЖЕВОЇ ФУНКЦІЇ

4.1 Оптимальна організація віртуальних мережевих функцій

Поява віртуалізації мережевих функцій (NFV) призводить до появи безлічі нових алгоритмічних проблем в роботі мереж зв'язку. NFV вводить волатильність в управлінні мережевими функціями, які можуть бути динамічно організовані, тобто розміщені, змінені та інше. Віртуальні мережеві функції (VNF) можуть належати цілям VNF, де вузли в ланцюжку можуть обслуговувати безліч запитів, що виходять з країв мережі. В роботі визначається проблема розміщення і маршрутизації VNF (VNF-PR), пропонуючи гнучку формулювання лінійного програмування, яке здатне враховувати специфічні особливості і обмеження інфраструктур NFV і суттєво відрізняється від існуючих форм впровадження віртуальної мережі в сучасний стан.

Також розробляється математична евристика, здатна масштабуватися з кількома цілями і великими екземплярами. Завдяки великим симуляціям ми робимо висновки про компроміси, що досягаються між завданнями підвищення ефективності інфраструктури класичного трафіку (TE) і NFV, оцінюючи вимоги до доступу до Інтернету та віртуальної приватної мережі (VPN). Ще кількісно порівнюється ефективність запропонованої евристики VNF-PR до класичного підходу до віртуальної мережі (VNE), запропонованим для оркестровки NFV, що показує обчислювальні відмінності і як модернізований підхід може забезпечити більш стабільне і більш близьке до оптимального рішення.

Після десяти років фундаментальних досліджень з мережевої віртуалізації і впровадження віртуальних мереж, віртуалізація мережевих функцій стає реальністю завдяки величезним інвестиціям, які роблять телекомунікаційні провайдери, постачальники хмарних обчислень і

постачальники. Гранична точка знаходиться в 2012 році, коли потрібно було експериментувати і розгорнути те, що було придумано як «Віртуалізація мережевих функцій (NFV)», це призвело до створення дослідницької групи NFV в Європейському інституті стандартів телекомунікацій (ETSI). З тих пір прикладні дослідження і розробки прискорили інвестиції, тому попередні прототипи були продемонстровані і розгорнуті (що призводить до комерціалізації в деяких випадках) з кінця 2014 року. З NFV увага досліджень мережевої віртуалізації тепер зосереджена на ключових аспектах систем NFV, які або не вважалися доречними, або не задумувалися перед галузевими зусиллями в країнах, що розвиваються організаціях стандартів (SDO). Ключові аспекти, про які варто згадати, це:

Цільовий ланцюжок послуг NFV, тобто проблема дозволу потоку трафіку, що проходить через попередньо обчислений або динамічно обчислюваний список вузлів VNF, можливо, з огляду на той факт, що вузли VNF можуть бути розміщені і переміщені по кластерам віртуалізації як функція призначення попиту існуючим ланцюгам VNF або підланцюгам;

- зміни вхідного / вихідного бітового потоку в VNF, обумовлені конкретними операціями VNF (такими як стиснення, як при кодуванні, декомпресія, як при тунелюванні);
- затримка обробки та пересилання VNF як параметр оркестровки.

Головною метою постає заповнення недоліку ефективності витрат в розумний час виконання.

ETSI є де-факто посиланням SDO для специфікації функціональної архітектури високого рівня NFV. Високий рівень означає, що його ідентифікована роль - це специфікація основних функціональних блоків, їх архітектура і взаємозв'язок, елементи реалізації яких потім можуть бути точно адресовані іншими SDO. ETSI визначає три компоненти для архітектури NFV:

- Віртуальні мережеві функції (VNF);

- Інфраструктура NFV (NFV), включаючи елементи, необхідні для запуску VNF, такі як вузол гіпервизора і кластери віртуалізації;
- Управління та оркестровка (MANO), обробка операцій, необхідних для запуску, міграції, оптимізації вузлів і ланцюгів VNF, можливо, в координації з оркестрами транспортних мереж.

Таким чином, процедури MANO підтримують економію масштабу NFV, так що фізичні ресурси віртуалізації NFVI (сервери та кластери), призначені для операцій NFV, ефективно використовуються як для операторів NFVI, так і для крайніх користувачів. Перспективним варіантом використання NFV для мереж-операторів є віртуальне клієнтське обладнання (vCPE), яке спрощує обладнання CPE за допомогою віртуалізованих індивідуальних мережевих функцій, розміщених в місцях доступу і мережі агрегації, як показано на рис., а також інші перспективні програми, такі як віртуалізація кластера Evolved Packet Core (EPC) в стільникових основних мережах і віртуалізація стільникових базових станцій.

Операції MANO численні і варіюються від розміщення і створення екземплярів VNF, щоб краще відповідати вимогам користувача до ланцюжка і маршрутизації ланцюгів VNF через транспортну мережу, що має в своєму розпорядженні кілька місцерозташування NFVI. Частина рішення по оркестровці також може бути конфігурацією VNF, щоб ділитися ними між активними потребами, одночасно виконуючи спільні завдання Traffic Engineering (TE) в мережах IP-транспорту, а також нові цілі ефективності NFV, такі як мінімізація кількості примірників VNF. У цьому контексті внесок в документ виглядає наступним чином:

- визначається і формулюється за допомогою математичного програмування задача оптимізації VNF Placement and Routing (VNF-PR), включаючи обмеження стиснення / декомпресії і два режими затримки пересилання (з і без швидкого шляху), як в рамках цілей TE, так і NFV.

- якісно і кількісно порівнювати підхід VNF-PR до існуючого підходу Virtual Embedding (VNE);
- розроблено математичний евристичний підхід, що дозволяє також запускати експерименти для великих випадків проблеми протягом прийняттого часу виконання.

4.2 Сучасний стан оркестровки NFV

Велика кількість робіт в літературі стосується оптимізації розміщення віртуальних машин (VM) щодо, наприклад, балансування навантаження на сервер або енергозбереження. Віртуалізація мережі між віртуальними машинами також є проблемою, що розглядається в цій області.

У NFV мережеві функції, які колись керували апаратними базами, тепер повинні бути віртуалізувати як VNF. VNF можуть бути пов'язані один з одним, щоб забезпечити конкретну послугу, також відому як ланцюжок обслуговування VNF. Постачальники послуг можуть розгортати певні ланцюжка послуг для забезпечення вимог до мережевих послуг, запитуваних клієнтами.

Попередні роботи по організації оркестровки NFV, як правило, вирішують проблему оркестровки NFV як проблему VNE, яка розглядає запити віртуальної мережі як логічні графи, які повинні бути вбудовані в мережу субстрату. VNF розглядаються як звичайні віртуальні машини, які відображаються в мережі контейнерів VM, які взаємопов'язані через фізичні посилення, на яких розміщуються логічні посилення віртуальних мереж. Аналогічним чином пропонується розміщення ланцюгів VNF, яке об'єднує проблеми маршрутизації розташування і проблеми VNE, вирішуючи спочатку розміщення, а потім ланцюжок. Застарілу проблему VNE відокремлюють на дві проблеми впровадження: вбудовування VM і впровадження ланцюжка обслуговування, де ланцюжок послуг

впроваджений на віртуальних машинах і кожна VM на фізичних серверах. Кожен ланцюжок обслуговування має особливі вимоги, зокрема, вимога наскрізної затримки.



Рисунок. 4.1 Традиційне обладнання для клієнтів (CPE) з ланцюжком VNF



Рисунок. 4.2 Віртуалізоване CPE (vCPE) з ланцюжком VNF

Розміщення і маршрутизація VNF - це проблема, принципово відмінна від проблеми VNE. Як і в VNE, віртуальні мережеві вузли повинні бути

розміщені в базовій фізичній інфраструктурі. Однак, на відміну від VNE, в розміщенні VNF і маршрутизації: (i) попит не є багатоточковим мережевим запитом на з'єднання, а вимагає маршрутизації потоку-адресата-точка-точка, і (ii) конкретні аспекти NFV, такі як латентна поведінка пересилання, змінити пропускну здатність самостійно / виходу та швидкість передачі, не розглядаються в VNE. Їх включення додатково збільшило б тимчасову складність VNE. У цьому сенсі проблема розміщення та маршрутизації VNF ближче до проблеми місцерозташування об'єкта, тоді як VNE ближче до проблеми зіставлення.

Відповідний спосіб вирішення проблем вирішення NFV MANO полягає у визначенні проблеми VNF Placement and Routing (VNF-PR), безпосередньо адаптованої до середовища NFV, заради тимчасової складності, точність моделювання і практичність використання. Робота фокусується на надання більш загального формулювання проблеми розміщення VNF і маршрутизації. Крім ланцюжка з гарантіями замовлення VNF, також фіксується і досліджується практична функція стиснення і декомпресії трафіку, яка може бути нав'язана VNF уздовж маршруту трафіку.

Загальний підхід полягає в тому, щоб покладатися на властивості графа, знайти найкраще використання обмежених ресурсів при обслуговуванні більшої набору вимог. Проблема розміщення вузлів VNF з глибоким пакетом (DPI) (без ланцюжка), з формальним визначенням проблеми і жадібним евристичним алгоритмом для її вирішення. В роботі розглядається не тільки ефективність використання ресурсів, а й якість надання послуг (наприклад, затримка пересилання трафіку). Крім того, розглянуто компроміс між цілями ефективності ресурсів і цілями розвитку мережевого трафіку.

4.3 Мережева модель

4.3.1 Опис проблеми

Визначення: проблема розміщення та маршрутизації функції віртуальної мережі (VNF-PR). Мережа представлена графіком $G(N, A)$, де N - це набір вузлів комутації, A представляє можливі спрямовані з'єднання між вузлами. Маршрутизатор $i \in N$ і пов'язаний з ним кластер NFVI представлені одним і тим же вузлом; цей вибір дозволяє обмежити розмір графіка і зменшити обчислювальні зусилля. Ми представляємо з $N_v \subset N$ безліч вузлів N , які мають кластерами серверів NFVI. Розглянемо набір вимог D , кожен попит $k \in D$ характеризується вихідним ок, точкою призначення t_k , номінальною шириною смуги b_k (статистично репрезентативною для попиту k) і послідовністю VNF різних типів, які повинні задовольняти попит (необхідно подолати попит). Для кожного VNF зарезервована одна віртуальна машина, тому ми можемо говорити про розподіл віртуальної машини або VNF на кластері NFVI, що означає, що ми резервуємо необхідні ресурси (наприклад, ЦП, ОЗУ) для розміщення віртуальної машини з VNF. Завдання оптимізації VNF-PR полягає в тому, щоб знайти:

- оптимальне розміщення вузлів VNF над кластерами NFVI;
- оптимальну маршрутизацію для вимог і їх призначення цілям вузлів VNF.

За умови:

- обмеження пропускної здатності каналу;
- обмеження пропускної здатності кластера NFVI;
- обмеження стиснення / декомпресії потоку VNF;
- обмеження часу очікування пересилання VNF;
- обмеження обміну вузлами VNF;
- ланцюга VNF (повного або часткового) для кожного попиту.

Мета оптимізації повинна включати як показники рівня мережі, так і показники продуктивності на рівні NFVI. В розглянутій мережевий моделі пропонується як метрику рівня мережі класична метрика TE, тобто мінімізація максимального використання посилань. Як метрика рівня NFVI пропонується мінімізація виділених обчислювальних ресурсів. Крім того, допускається, що:

- Кілька VNF одного і того ж типу (тобто одна і та ж функціональність) можуть бути виділені на одному вузлі, але кожен запит не може розділити потік на кілька VNF одного і того ж типу;
- Споживання обчислювальних ресурсів VNF може бути виражено в термінах живої пам'яті (наприклад, ОЗУ) і обчислювальних модулів (CPU), проте модель повинна бути досить універсальною для інтеграції інших обчислювальних ресурсів;
- Затримка, що вводиться VNF, може слідувати одному з двох наступних режимів (як показано на рис. 4.3):

Стандарт: VNFs накопичують трафік на вхідних і вихідних віртуальних і фізичних мережевих інтерфейсах, так що затримку пересилання можна розглядати як опуклу кусочно-лінійну функцію сукупної швидкості передачі бітів в VNF через збільшення використання буфера і втрати пакетів у вигляді бітрейт зростає. Це відноситься до стандартних VNF, які працюють зі стандартними буферами ядра і гіпервизора.

Fastpath: VNF використовують оптимально розраховані і відносно невеликі буфери і зменшують кількість копій пакетів в пам'яті, так що затримка пересилання постійна до максимальної сукупної швидкості передачі біт, після якої пакети видаляються (наприклад, це відбувається для Intel / 6WIND DPDK - швидкі рішення).



Рисунок. 4.3 Приклад профілів затримки пересилання VNF.

На рис. 4.3 наведено приклади профілів зптримки пересилання для двох випадків.

- Для кожного запиту і кластера NFVI може бути встановлена тільки одна VNF стиснення / декомпресії. Це дозволяє нам зберегти час виконання на прийнятних рівнях, що не зменшуючи надмірно альтернативи розміщення VNF. Це припущення може бути ослаблене за рахунок роботи над розширеним графіком і, отже, збільшення обчислювального часу алгоритму.

4.3.2 Математичне формулювання

Спочатку вводиться базова модель, яка не враховує обмеження затримки і функції стиснення / декомпресії. Причина цього вибору двояка. По-перше, це дозволяє більш чітко пояснення моделі і покрокове введення технік, які дозволяють утримувати модель лінійною. Без цих двох функцій модель являє собою комбінацію дизайну мережі та місця розташування об'єкта. По-друге, в алгоритмічній фазі, яка використовувалась для

вирішення моделі, застосовується послідовність моделей зі зростаючою складністю (базова модель з затримкою, з обробкою / декомпресією).

1) Базова модель VNF-PR: математичні позначення, що використовуються в наступному змішаному цілочисельному лінійному програмуванні (MILP), що являє собою основне формулювання проблеми VNF-PR. Використовується чотири сімейства двійкових змінних: $x(k, ij)$ являє собою використання послань на вимогу, отже, шлях, який використовується вимогою; $y(f, i, n)$ являє собою розподіл копії n VNF типу f на даному вузлі; $w(f, ik)$ являє собою привласнення цієї вимоги VNF, а $z(f, ik)$ являє собою привласнення цієї вимоги конкретною копією VNF. Безперервна змінна π_{ik} використовується для представлення позиції вузла i в шлях, який використовується для маршрутизації запиту k . Це сімейство змінних необхідно для накладення (повного або часткового) порядку в ланцюжку VNF. Розглядаються дві об'єктивні функції:

- Мета TE: мінімізувати максимальне використання мережевих послань:

$$\min U \quad (4.1)$$

- Мета NFV: мінімізувати кількість ядер (ЦП), які використовуються інстанційованими VNF:

$$\min \sum_{i \in N_v} \sum_{f \in F} \sum_{n \in 1..c_i^f} r_{CPU} y_i^{fn} \quad (2)$$

Загальні математичні примітки:

- Групи: N -всі вузли; $N_v \subseteq N$ -вузли, обладнані кластером NFVI; $A \subseteq N \times N$ -всі дуги (посилання); D -вимоги; R -типи ресурсів (CPU, RAM, ...); F -типи VNF;

- Параметри мережі: γ_{ij} -пропускна спроможність; G_i -ємність вузла $i \in N_v$ в термінах ресурсу $r \in R$;

- Параметри попиту: o_k -джерело попиту $k \in D$; t_k -призначення попиту $k \in D$; b_k -номінальна пропускна здатність попиту $k \in D$; $m_{f, k-1}$,

якщо запит $k \in D$ запитує VNF типу $f \in F$; s_f^k - коефіцієнт порядку для VNF f , запитаний на вимогу k ;

- Параметри VNF / VM: r_r - вимога ресурсу $r \in R$ для віртуальної машини; c_{fi} - максимальну кількість копій VNF f на вузлі i ;
- Бінарні змінні: $x_{ij}^k = 1$, якщо дуга (i, j) використовується на вимогу $k \in D$; $z_{fnik} = 1$, якщо вимога $k \in D$ використовує копію n -го VNF типу $f \in F$, поміщеного на вузол $i \in N_v$; $y_{fni} = 1$, якщо призначена n -я копія VNF f до вузла $i \in N_v$; $w_{fik} = 1$, якщо запит k використовує VNF f на вузлі $i \in N_v$;
- Неперервні змінні: $U \geq 0$ - максимальне використання послань; $\pi_{ik} \geq 0$ - становище вузла i в шлях, який використовується вимогою k .

Перша мета дозволяє враховувати внутрішні коливання, пов'язані з інтернет-трафіком, і, таким чином, мінімізувати ризик раптового вузького місця в мережевих посланнях. Останнє передбачає той факт, що сьогодні перша модульна вартість серверів віртуалізації, особливо з точки зору споживання енергії, - це процесор. Тепер ми представляємо обмеження.

Обмеження на обмеження руху по одному шляху:

$$\sum_{j:(i,j) \in A} x_{ij}^k - \sum_{j:(j,i) \in A} x_{ji}^k = \begin{cases} 1 & \text{if } i = o_k \\ -1 & \text{if } i = t_k \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in D, \forall i \in N \quad (3)$$

Обмеження швидкості використання:

$$\sum_{k \in D} b_k x_{ij}^k \leq U \gamma_{ij} \quad \forall (i, j) \in A \quad (4)$$

Обмеження ресурсів вузла (VNF):

$$\sum_{f \in F} \sum_{n \in 1..c_f^i} r_r y_i^{fn} \leq \Gamma_{ir} \quad \forall i \in N_v \quad (5)$$

Кожну вимогу використовує рівно один VNF кожного необхідного типу:

$$\sum_{i \in N_v} \sum_{n \in 1..c_i^f} z_{ik}^{fn} = 1 \quad \forall k \in D, f \in F : m_k^f = 1 \quad (6)$$

Обмеження (7) - (9) - це обмеження послідовності між двійковими змінними. VNF можна використовувати, тільки якщо він присутній, для даного вузла:

$$z_{ik}^{fn} \leq y_i^{fn} \quad \forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f \quad (7)$$

Якщо вимога не передається VNF, воно не може його використовувати:

$$z_{ik}^{fn} \leq \sum_{j:(j,i) \in A} x_{ji}^k \quad \forall k \in D, i \in N_v, f \in F : m_k^f = 1 \quad (8)$$

Допоміжні змінні для забезпечення узгодженості:

$$\sum_{n \in 1..c_i^f} z_{ik}^{fn} = w_{ik}^f, \quad \forall k \in D, i \in N_v, f \in F \quad (9)$$

Нарешті, вводяться обмеження, щоб уникнути нездійсненою маршрутизації і нав'язувати порядок ланцюга VNF:

Попередження утворення ізольованих циклів:

$$\pi_{jk} \geq \pi_{ik} + x_{ij}^k - |N_v|(1 - x_{ij}^k) \quad \forall k \in D, (i, j) \in A \quad (10)$$

Накладення порядку на віртуальні функції:

$$\pi_{jk} \geq \pi_{ik} - (|N_v| + 1)(2 - w_{ik}^{f_1} - w_{ik}^{f_2}) \quad \forall k \in D, \\ \forall i, j \in N_v, f_1, f_2 \in F : s_k^{f_2} \geq s_k^{f_1} \quad (11)$$

Якщо ми розглянемо обмеження балансу потоку (3) і обмеження пропускної здатності каналу (4), для кожного запиту вибір дуг, що утворюють шлях плюс ізольований цикл, може бути прийнятним рішенням. В умовах чистої маршрутизації ці рішення є еквівалентними рішенням, де видаляються змінні маршрутизації протягом циклу та зберігаються лише ті, що проходять по шляху. Фактично, обидва обмеження (3) і обмеження (4)

будуть дійсні для цього нового рішення. Завдання інтегрує функції маршрутизації в проблему місця розташування об'єкта, тому такі рішення не завжди можуть бути перетворені простим шляхом, просто видаливши цикл. Фактично, якщо об'єкт (VNF), який використовується попитом, знаходиться на циклі, видалення циклу дасть неможливе рішення. Тому необхідно видалити такі рішення безпосередньо в моделі, з цією метою були введені обмеження (10). Змінна π_{ik} є порядок вузла i за запитом обслуговування k , тому, якщо існує дуга (i, j) , то π_{jk} буде не менше $\pi_{ik} + 1$. З іншого боку, якщо $\text{arc}(i, j)$ не існує ($x_{ij} = 0$), то обмеження перестав бути активним: π_{ik} завжди менше, ніж $|N_v|$, так як шлях може містити не більше всіх вузлів на графіку. Таким чином, дозволені тільки рішення, що містять прості шляхи. Ці змінні також використовуються в рівнянні (11), щоб дозволити накладення порядку на VNF уздовж маршруту попиту k . Вони накладають, що якщо попит k використовує VNF f_1 , розташований на вузлі i і його наступника VNF f_2 ($s_{f_2 k} \geq s_{f_1 k}$), який розташований на вузлі j , то в шляху маршрутизації попиту k вузол i повинен бути попереднім вузлом j .

Примітки до моделі передачі затримки:

- Параметри: L - максимально допустима затримка для попиту; λ_{ij} - затримка, введена посиланням $(i, j) \in A$;
- Стандартна модель затримки: $g_{fj}(b)$ - j -й компонент лінеаризованої функції затримки для VNF f і агрегованої смуги частот b ; ng - кількість штучних компонентів lin. функції затримки;
- Модель затримки fastpath: l_f - затримка, введена VNF f ; $B_f \max$ - максимально допустима ширина смуги пропускання для VNF f ;
- Змінні: $l_{f ik} \geq 0$ - затримка, що вимагає $k \in D$, використовуючи VNF f на вузлі i типу $f \in F$, що розміщується вузлом $i \in N_v$.

2) Затримка пересилання VNF: будемо вважати, що для кожної вимоги $k \in D$ допускається максимальна затримка L , щоб гарантувати певний рівень QoS. Затримка залежить від двох компонентів: затримки каналу,

представленої параметром λ_{ij} для кожної дуги (i, j) і затримки VNF. Затримка VNF залежить від використовуваної моделі затримки (стандартної або швидкої). Щоб зберегти позначення якомога більш однорідним, було введено додаткову змінну $l(f, ik)$, щоб представити затримку, що випробовується запитом k , які перетинають VNF f , розташовану на вузлі i . Тому ми використовується набір обмежень, загальних для обох моделей, що обмежують загальну затримку:

$$\sum_{(i,j) \in A} \lambda_{ij} x_{ij}^k + \sum_{i \in N_v} \sum_{f \in F} l_{ik}^f \leq L \quad \forall k \in D \quad (12)$$

Набір обмежень, що залежать від обраної моделі затримки, дозволяє обчислити значення змінної $l(f, ik)$.

Стандартна: затримка, введена на вимогу k для використання VNF f , залежить від загального трафіку, що проходить через VNF (його власні та інші вимоги). Нехай $g_j^f(\cdot)$ - j -та компонента кусково-лінеарізуючої функції включає зсув для VNF типу f , то отримаємо:

$$l_{ik}^f \geq g_j^f \left(\sum_{d \in D} b_k z_{id}^{fn} \right) - L(1 - z_{ik}^{fn}) \quad (13)$$

$$\forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f, j \in 1..n_g$$

Обмеження (13) діє тільки тоді, коли вимога використовує копію n VNF f на вузлі i ($z(f, ik) = 1$). В іншому випадку, оскільки загальна затримка обмежена L (обмеження (12)), будь-який член $g_j^f(\cdot)$ повинен бути менше L , і тому обмеження є надлишковим обмеженням ($l_{ik} \geq 0$). Можна помітити, що навіть якщо в стандартній моделі затримки немає меж дозволеної пропускної здатності, обмеження (12), що обмежує загальну затримку i , як наслідок, затримку для VNF, накладає неявну межу дозволеної смуги пропускання.

Fastpath: затримка фіксована, але накладається обмеження в загальному трафіку, що підтримується VNF. Тому ми отримуємо наступні два набору обмежень:

$$l_{ik}^f = \bar{l}^f \quad \forall k \in D, i \in N_v, f \in F \quad (14)$$

$$\sum_{d \in D} b_d z_{id}^{fn} \leq B_{max}^f \quad \forall i \in N_v, f \in F, n \in 1..c_i^f \quad (15)$$

Зміни моделі:

- Набори: N_a - вузли доступу; N_0 - дублювання вузлів доступу, де знаходяться вимоги;
- Параметри: μ_f - коефіцієнт стиснення / декомпресії для VNF $f \in F$; b_{min}^k - мінімальна ширина смуги попиту $k \in D$; b_{max}^k - максимальна ширина смуги попиту $k \in D$; M_i - максимальний обсяг трафіку, який може бути переключено вузлом i ;
- Змінні: $\varphi_{kij} \geq 0$ - потік для попиту $k \in D$ на дузі (i, j) ; $\psi_{fnik} \geq 0$ - потік для попиту $k \in D$, який запроваджує вузол i , i використовуючи копію n VNF $f \in F$.

Обмеження (14) можна підставити безпосередньо в обмеження (12).

3) Стиснення / декомпресія з бітовою швидкістю. Щоб уявити можливість стиснення / декомпресії потоків для деяких VNF, потрібно внести деякі зміни в опис моделі. Введено параметр стиснення / декомпресії μ_f для кожного типу VNF, $\mu_f > 1$ означає, що декомпресія виконується VNF f . Коли попит проходить через VNF з $\mu_f \neq 1$, його пропускна здатність змінюється, тому знання тільки маршрутизації (x змінних) недостатньо для визначення загального потоку уздовж дуги. З цієї причини введено змінну φ_{kij} , яка явно представляє потік на дузі (i, j) для попиту k . Іншим наслідком є те, що класичні рівняння балансу потоку вже недійсні. Щоб розширити модель, не вводячи надлишкову складність, припущено, що при заданому вузлі i і вимозі k такий попит використовує не більше VNF f з коефіцієнтом стиснення / декомпресії ($\mu_f \neq 1$). Працюючи над розширеним графіком, щоб розрізнити вузли доступу (вузли походження / призначення) і вузли NFVI. Кожен вузол i доступу дублюється в вузлі i' . Дуга (i, i') буде додана, і всі

дуги (i, j) , що відбуваються з вузла доступу i , будуть перетворені в дуги (i', j) . Тому функціональність маршрутизації знаходиться на вузлі i , а функціональність NFVI може бути виділена на вузлі i' . Крім того, було додано змінну ϕ ($f_n ik$), яка представляє потік попиту k , що входить у вузол i , і використовуючи копію n VNF типу f . Якщо попит проходить через VNF з коефіцієнтом стиснення / декомпресії μ_f , то вихідний потік вузла пропорційний потоку:

$$\sum_{j \in N: (i,j) \in A} \phi_{ij}^k = \mu_f \sum_{j \in N: (j,i) \in A} \phi_{ji}^k$$

або еквівалентний:

$$\sum_{j \in N: (i,j) \in A} \phi_{ij}^k - \sum_{j \in N: (j,i) \in A} \phi_{ji}^k = \sum_{j \in N: (j,i) \in A} (\mu_f - 1) \phi_{ji}^k$$

Це рівняння справедливо тільки в тому випадку, якщо попит k використовує копію n VNF f на даному вузлі i (затримка залежить від проходження смуги пропускання однієї копії). Тому для отримання правильного рівняння потрібно написати:

$$\begin{aligned} \sum_{j \in N: (i,j) \in A} \phi_{ij}^k - \sum_{j \in N: (j,i) \in A} \phi_{ji}^k = \\ \sum_{j \in N: (j,i) \in A} \phi_{ji}^k \sum_{n \in 1..c_i^f} (\mu_f - 1) z_{ik}^{fn} \end{aligned}$$

Коли $\sum_{n \in 1..c_i^f} (\mu_f - 1) z_{ik}^{fn} = 0$ обмеження говорить про те, що вхідний потік та вихідний потік є однаковими, тобто якщо VNF не проходить, то потік залишається незмінним. Той же результат виходить для всіх VNF f таких, що $\mu_f = 1$ (без стиснення / декомпресії). З метою лінеаризації цього обмеження введено змінну ψ ($f_n ik$):

$$\psi_{ik}^{fn} = \left(\sum_{j \in N: (j,i) \in A} \phi_{ji}^k \right) z_{ik}^{fn}$$

Обмеження можуть бути лінеароване з використанням рівнянь (20) - (22), причому параметр M_i дорівнює $\sum_{(j,i) \in A} \gamma_{ji}$, який представляє собою максимальну кількість потоку, що входить у вузол i . Якщо $(\mu_f - 1)z_{ik}^{fn} = 1$ тоді ψ_{ik}^{fn} представляє потік попиту k , що входить у вузол i і проходить через копію n VNF f (обмеження (20) - (21)), в іншому випадку воно дорівнює нулю (обмеження (22)). Тепер можна представити нові обмеження, які необхідно додати в базову модель VNF-PR: баланс потоку для вузлів доступу:

$$\begin{aligned} & \sum_{j \in N: (i,j) \in A} \phi_{ij}^k - \sum_{j \in N: (j,i) \in A} \phi_{ji}^k = \\ & = \begin{cases} b^k & \text{if } i = o_k \\ 0 & \text{otherwise} \\ -b^k \prod_{f \in F: m_k^f = 1} \mu_f & \text{if } i = t_k \end{cases} \quad \forall k \in D, i \in N_a \end{aligned} \quad (16)$$

Баланс витрати і стиснення / декомпресії для вузлів NFVI і для кожного вимоги:

$$\begin{aligned} & \sum_{j \in N: (i,j) \in A} \phi_{ij}^k - \sum_{j \in N: (j,i) \in A} \phi_{ji}^k = \\ & \sum_{f \in F, n \in 1..c_f^i} (\mu_f - 1) \psi_{ik}^{fn} \quad \forall k \in D, i \in N_v \end{aligned} \quad (17)$$

Когерентність між змінними шляху і потоку:

$$\phi_{ij}^k \leq b_k^{\max} x_{ij}^k \quad \forall k \in D, (i, j) \in A \quad (18)$$

$$\phi_{ij}^k \geq b_k^{\min} x_{ij}^k \quad \forall k \in D, (i, j) \in A \quad (19)$$

Обмеження лінеаризації стиснення / декомпресії VNF:

$$\psi_{ik}^{fn} \leq \sum_{j \in N: (j,i) \in A} \phi_{ji}^k + M_i(1 - z_{ik}^{fn})$$

$$\forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f \quad (20)$$

$$\psi_{ik}^{fn} \geq \sum_{j \in N: (j,i) \in A} \phi_{ji}^k + M_i(1 - z_{ik}^{fn})$$

$$\forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f \quad (21)$$

$$\psi_{ik}^{fn} \leq M_i z_{ik}^{fn}$$

$$\forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f \quad (22)$$

Одне стиснення / декомпресія VNF на вузол і вимога:

$$\sum_{f \in F} \sum_{n \in 1..c_i^f: \mu_f \neq 1} z_{ik}^{fn} \leq 1 \quad \forall k \in D, \forall i \in N_v \quad (23)$$

Рівняння (16) являє собою баланс потоку для вузлів доступу. На вузлі призначення кількість потоків встановлюється рівним попиту, помноженому на все коефіцієнти стиснення всіх необхідних VNF. Рівняння (17) являє баланс потоку для даного вузла, який має можливість розміщення VNF (NFVI). Рівняння (18) - (19) дозволяють пов'язувати змінні x і ϕ таким чином, що тільки тоді, коли дуга (i, j) використовується на вимогу k , тобто $x(k, ij) = 1$, то змінна ϕ може бути відмінна від нуля. У міру того як попит проходить через VNF, який може стискати або розширюватися потік, можливо визначити верхню і нижню межі попиту: $b_k^{max} = b_k \prod_{f \in F: \mu_f \geq 1}$ і $b_k^{min} = b_k \prod_{f \in F: \mu_f \leq 1}$.

Змінні x і раніше необхідні для накладення виключення ізольованих циклів і порядку в ланцюжку VNF. Обмеження швидкості використання повинні бути змінені таким чином:

$$\sum_{k \in D} \phi_{ij}^k \leq U \gamma_{ij} \quad \forall (i, j) \in A \quad (24)$$

Щоб врахувати комбінований ефект стиснення / декомпресії і затримки VNF, необхідна деяка модифікація.

Для стандартної моделі обмеження (13) змінюються таким чином:

$$l_{ik}^f \geq g_j^f \left(\sum_{d \in D} \psi_{id}^{fn} \right) - L(1 - z_{ik}^{fn})$$

$$\forall k \in D, i \in N_v, f \in F, n \in 1..c_i^f, j \in 1..n_g \quad (25)$$

Для моделі fastpath обмеження (15) змінюються таким чином:

$$\sum_{d \in D} \psi_{id}^{fn} \leq B_{max}^f \quad \forall i \in N_v, f \in F, n \in 1..c_i^f \quad (26)$$

4.3.3 Багатоцільове математичне евристичне рішення

Багатоцільова проблема: мінімізація максимального використання посилань, що відображає орієнтоване на ISP бачення, щоб поліпшити якість обслуговування користувачів (строго пов'язано з перевантаженням каналів зв'язку, особливо для служб реального часу) і мінімізація загальної вартості інфраструктури віртуалізації на рівень NFVI, який відображає цілі постачальника NFVI. Такий багатоцільовий підхід має особливе значення, коли постачальник NFVI є інший об'єкт, ніж постачальник послуг Інтернету. Ці дві мети знаходяться в конкуренції; насправді, щоб отримати низьке використання, необхідно виділити велику кількість VNF.

Пріоритети цілей: спочатку слід мінімізувати максимальне використання посилань (U), а потім вартість NFV (загальна кількість використовуваних CPU). Це розглядається як мета TE-NFV. На практиці виконується перший крок оптимізації, щоб знайти оптимальне рішення, що відповідає максимальному використанню посилань (U *), а потім, зберігаючи оптимальне значення, знайдене на першому кроці в якості параметра (тобто додаючи обмеження $U \leq U^*$), мінімізуємо другу мету (вартість NFV). Фактично, при заданому оптимальному значенні першого кроку на другому етапі доступні різні можливі конфігурації, і на цьому другому етапі може

бути досягнуто значне зниження первинної вартості без втрати по відношенню до основної мети (максимальне використання посилань).

Щоб зрозуміти вплив максимального використання посилань на вартість NFV, було вивчено чутливість другого етапу оптимізації по оптимальному значенню U^* . Тому повторно оптимізовано другу мета, яка послаблює обмеження на максимальне використання посилань параметром α , тобто було застосоване обмеження $U \leq \alpha + U^*$ замість $U \leq U^*$. Збільшуємо α крок за кроком, поки вартість NFV не зменшиться більше. Це значення відповідає спочатку мінімізації вартості NFV, а потім максимальної вартості використання каналу (NFV-TE).

Оптимізація повної моделі дуже дорога і що обчислювальний час може бути значно зменшено, виконуючи послідовність оптимізації, починаючи з базової моделі і закінчуючи повною. Результат кожного кроку використовується в якості відправної точки для наступного, це дозволяє скоротити час обчислень і / або створити кращі рішення або прогалини (коли оптимізація зупинена до досягнення оптимального рішення). Щоб бути більш точним, послідовність моделей, які було оптимізовано, спочатку є базовою (враховуються тільки маршрутизація попиту, місце розташування і ємності VNF), потім базовий (додається затримка) і, нарешті, затримка і стиснення / декомпресія.

Найбільш складною моделлю з точки зору оптимізації є остання. З цієї причини для цього кроку необхідно забезпечити допустиме початкове рішення. Для цієї мети попередній крок, що оптимізує модель, повинен виконуватися з деякою незначною модифікацією. Функція стиснення / декомпресії змінює кількість потоку, що проходить за графіком, тому, щоб гарантувати, що рішення другого кроку можливо для останнього, необхідно направити найгіршу кількість потоку, задану в випадку, коли всі VNF з декомпресією вже застосовуються до потоку попиту.

Цільова функція NFV призводить до більш складних обчислень, ніж TE. Тому для отримання оптимального рішення цілі NFV використовується

процедура поділу за кількістю виділених VNF / VM, щоб гарантувати оптимальність рішення, навіть якщо за один крок вирішувач не може цього гарантувати: тобто при кожному кроці ділення навпіл, якщо знайдено допустиме рішення, кількість VNF / VM ділиться на два, і якщо не існує допустимого рішення (доказ того, що проблема нездійсненна, призводить до більш простої обчислювальної задачі, ніж пошук оптимального рішення), то вона подвоюється.

4.3.4 Подальші уточнення моделі

Модель, яка надана вище, може бути уточнена і налаштована відповідно до конкретних вимог. Нижче перераховані можливі варіанти, а також відповідні варіації моделювання.

Правила прив'язки VNF: через конфіденційність, надійність або з інших причин провайдер може захотіти накласти правила розміщення певних типів VNF: розміщуватися або не розміщуватися на певних серверах, групуватися або не групуватися разом та інше. Такі конкретні правила розміщення VNF називаються афінними і / або анти-афінними правилами. Щоб розширити модель, щоб врахувати їх, найпростішим способом є введення нової змінної, що представляє наявність певного типу VNF f на даному вузлі i (модель дозволяє мати кілька копій одного і того ж типу VNF на тому ж сайті). Назвемо цю змінну v_i^f , вона буде дорівнює одиниці, якщо VNF типу f знаходиться на вузлі i . Щоб ці змінні відповідали вже певним змінним y_i^{nf} , потрібно додати:

$$\sum_{n \in 1..c_i^f} y_i^{nf} \leq c_i^f v_i^f \quad \forall i \in N_v, f \in F$$

Точніше, загальні правила афінності / анти-афінності: правила приєднання VNF-VNF, якщо два VNF часто спілкуються і повинні спільно використовувати вузол, треба зберегти VNF разом, щоб зменшити трафік по

мережах і підвищити ефективність трафіку. Нехай $AffVV_{f_1 f_2}$ - параметр, що дорівнює одиниці, якщо f_1 і f_2 повинні спільно використовувати один і той же вузол. Потім:

$$v_i^{f_1} = v_i^{f_2} \quad \forall (f_1, f_2) : AffVV_{f_1 f_2} = 1$$

Правила приєднання VNF-Server: деякі засоби запобігання вторгнень VNF повинні знаходитися в краях мережі для захисту від вірусів, трафіку з відмовами обслуговування (DoS) і спрямованих атак. Нехай $AffVS(f, i)$ - параметр, що дорівнює одиниці, якщо f має бути встановлений на i . Потім:

$$v_i^f = 1 \quad \forall (i, f) : AffVS_i^f = 1$$

або обмеженою підмножиною вузлів $S \in N_v$:

$$\sum_{i \in S} v_i^f = 1 \quad \forall (i, f) : AffVS_i^f = 1$$

Правила анти-афінності VNF-VNF: може знадобитися встановити кілька екземплярів одного VNF на кілька серверів, щоб підвищити надійність VNF від збоїв. Нехай $AAff(f)$ - антиафінний параметр; ми потім накладаємо, що принаймні $nbMin$ в вузлах містить VNF:

$$\sum_{n \in N_s} v_n^f \geq nbMin \quad \forall f : AAff_f = 1$$

Якщо різні VNF не можуть бути розташовані один з одним, нехай $AAffVV_{f_1 f_2}$ є параметром анти-афінності і накладає:

$$v_i^{f_1} + v_i^{f_2} \leq 1 \quad \forall (f_1, f_2) : AAffVV_{f_1 f_2} = 1$$

Правила анти-афінності VNF-сервера: може знадобитися для уникнення ресурсномістких VNF, що знаходяться на певних критично важливих серверах. Нехай \bar{AAffVS}_i^f є антиафінним параметром і накладає:

$$v_i^f = 0 \quad \forall (i, f) : \text{AAffVS}_i^f = 1$$

Всі обмеження, що встановлюють деякі змінні на один або нуль, просто зменшують кількість змінних; тому очікується, що такі обмеження не збільшують обчислювальний час. Для обміну VNF між різними вимогами може бути введено кілька інших умов (ізоляція VNF).

Ізоляція VNF: якщо один і той же VNF не може бути розділений між двома конкретними вимогами, можна додати обмеження, щоб накласти цю умову. Досить ввести параметр несумісності $inc_{k_1 k_2}$, що дорівнює одиниці, якщо потреба k_1 повинна бути ізольована від попиту k_2 ; то нам потрібно додати:

$$z_{ik_1}^{fn} + z_{ik_2}^{fn} \leq 1 \quad \forall i \in N_v, f \in F, \\ n \in 1..c_i^f, k_1, k_2 \in D$$

Кілька стиснень /декомпресій VNF на вузол NFVI: припустимо, що в кожному вузлі NFVI є трохи більше одного VNF, який може стискати /декомпресувати потік, тобто з коефіцієнтом стиснення $\mu_f \neq 1$. Це припущення можна пом'якшити, використовуючи розширений граф, в якому кожен вузол, який може розміщувати VNF (N_v), розширюється в декількох примірниках, по одному для кожного типу VNF, який може бути виділений в вузлі. В іншому випадку слід уявити всі можливі комбінації різних VNF, виділених для одного і того ж вузла, і додати додаткові виконавчі змінні для подання обраної комбінації.

Часткове впорядкування VNF: можна помітити, що частковий порядок може бути накладено з тією ж формою обмежень, які використовуються для загального впорядкування (11), просто обмежуючи їх число існуючими умовами пріоритету. Досить ввести обмеження для кожної пари VNF, яка має відношення пріоритету. Більш формально для кожного запиту k можна ввести орієнтований ациклічний граф $O_k (V_k, P_k)$, де вузли V являють собою набір VNF, які повинні обслуговувати попит ($V = \{i \in F: m(f_k) = 1\}$), а дуги P являють собою відношення порядку між такими VNF, тобто дугою $(i, j) \in P$,

якщо VNF j слід використовувати після VNF i . Тоді обмеження (11) можна переписати як:

$$\pi_{jk} \geq \pi_{ik} - (|N_v| + 1)(2 - w_{ik}^{f_1} - w_{ik}^{f_2})$$

$$\forall k \in D, \forall i, j \in N_v, f_1, f_2 \in V_k : (f_1, f_2) \in P_k$$

Додаткові обчислювальні обмеження: їх можна легко включити, налаштувавши існуючі параметри, оскільки запити обчислювальних ресурсів можуть бути виражені адитивним способом (наприклад, для зберігання).

Балансування навантаження: в поточній моделі кожен запит може використовувати один VNF для кожного типу. Модель може бути розширена, щоб забезпечити балансування навантаження на VNF. Якщо балансування навантаження локальна для кластера NFVI, зміна моделі невелика, насправді просто необхідно мати деякі безперервні змінні, що враховують кількість попиту, пов'язане з кожним VNF. Якщо балансування навантаження може бути між різними кластерами, тоді необхідно розширити модель, що дозволяє використовувати кілька шляхів для кожного запиту. Однак очікується, що таке розширення значно збільшить час виконання.

Різні шаблони VM: наприклад, розгляд моделі при одноточковій відповідності між шаблонами VNF і VM (єдиний шаблон). Тим не менше, кілька шаблонів VM можна розглядати в моделі за ціною збільшення одного виміру / індексу всіх змінних, індексованих по ідентифікаторам VNF.

Основний маршрутизатор як VNF: якщо функцію маршрутизації ядра також віртуалізувати, тобто, якщо вузол NFVI і мережевий маршрутизатор можна розглядати як один фізичний вузол, який виконує функцію базової маршрутизації, обробляючи сукупний трафік незалежно від попиту, як VNF, тоді нам потрібно додати член , пропорційний притік і відтік, до (5):

$$\begin{aligned}
& \sum_{k \in D} \sum_{f \in F} \sum_{n \in 1..c_i^f} r r_r y_i^{fn} \\
& + \sum_{k \in D} \sum_{j: (i,j) \in A} b_k x_{ij}^k \\
& + \sum_{k \in D} \sum_{j: (j,i) \in A} b_k x_{ji}^k \leq \Gamma_{ir} \quad \forall i \in N_v, r \in R
\end{aligned}$$

Якщо розглядається стиснення / декомпресія швидкості передачі даних, обмеження (5) має бути змінено таким чином:

$$\begin{aligned}
& \sum_{k \in D} \sum_{f \in F} \sum_{n \in 1..c_i^f} r r_r y_i^{fn} \\
& + \sum_{k \in D} \sum_{j: (i,j) \in A} \phi_{ij}^k \\
& + \sum_{k \in D} \sum_{j: (j,i) \in A} \phi_{ji}^k \leq \Gamma_{ir} \quad \forall i \in N_v, r \in R
\end{aligned}$$

4.4 Результати математичного моделювання

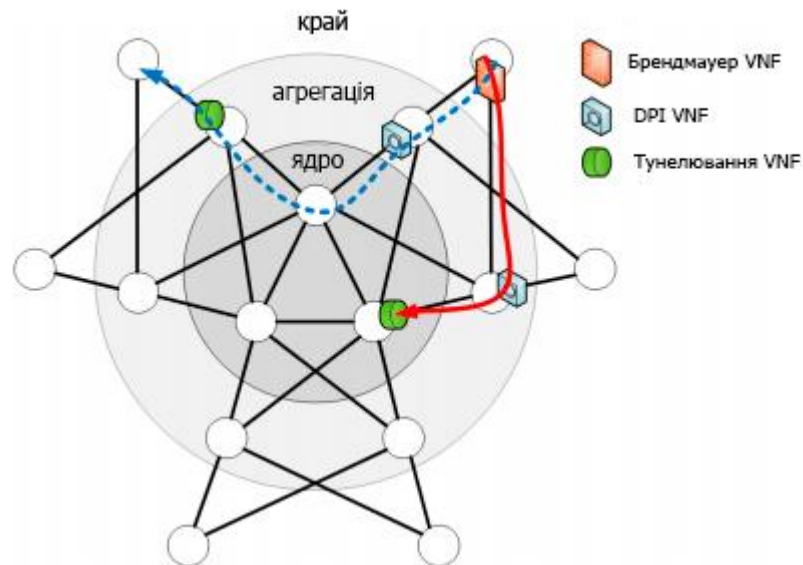


Рисунок. 4.4 Прийнята топологія мережі і приклад вирішення VNF-PR

Була прийнята трирівнева топологія, що представлена на рис. 4.4 для обчислювальних результатів. Кожен крайовий вузол з'єднаний з двома вузлами агрегації, кожен вузол агрегації з'єднаний з двома основними вузлами, а вузли ядра повністю зачеплені. Було розглянуто всі вузли як вузли NFVI, які можуть приймати VNF.

Тестування проведено з використанням двох різних тематичних досліджень для розподілу попиту: Інтернету і віртуальної приватної мережі (VPN). В інтернет-тематичному дослідженні (наприклад, потік червоним на рис. 4.4) вимоги до трафіку відправляються кожним крайовим вузлом (наприклад, кінцевим користувачем) на кожен основний вузол (наприклад, центр обробки даних) і з кожного основного вузла для досягнення кожного крайового вузла, що означає, що в цьому випадку обидва крайових вузла і вузли ядра є вузлами доступу (тобто, коли генеруються вимоги); в той час як в разі тематичного дослідження VPN (наприклад, синій потік на рис. 4.4) крайові вузли відправляють запити трафіку один одному, а це означає, що набір граничних вузлів відповідає набору вузлів доступу. Загальна кількість

вимог до трафіку по-різному для двох тематичних досліджень (36 для Інтернету та 30 для VPN), але було збережено постійний загальний обсяг трафіку (сума вимог) в мережі для порівняння. Це робиться для коригування інтервалу генерації випадкового попиту $[a, b]$: дійсно, вимоги генеруються випадковим чином з рівномірним розподілом в даному інтервалі $[a, b]$, таким чином, що вимоги до краю не можуть створити вузьке місце на краю посилення, тобто $a = 0,1$ і $b = 0,14$ в інтернет-дослідженні, $a = 0,13$ і $b = 0,17$ в тематичному дослідженні VPN. Ці значення дозволяють зберегти загальний обсяг трафіку на одному рівні для двох випадків. Для кожного випадку розглядаються 10 матриць випадкового попиту.

З'єднання агрегування розраховані таким чином, що існує ризик насичення каналу (тобто використання посилення вище 100%), якщо розподіл трафіку не оптимізовано. Основні посилення такі, що існує дуже низький ризик вузьких місць. Затримки каналу встановлюються таким чином, щоб впоратися з різними географічними областями: 1 мс для крайових посилення, 3 мс для посилення на агрегування і 5 мс для основних посилення. Використовується один шаблон віртуальної машини, що вимагає 1 процесор і 16 ГБ оперативної пам'яті. Запускаються установки тестів для наскрізної затримки (L) зі строгими і вільними значеннями (15 мс і 20 мс, відповідно). Розглянуто три типи VNF для кожного запиту: брандмауер VNF (стиснення, оскільки він блокує трафік), VNF з глибоким пакетом (DPI) і тунелювання VNF (декомпресія) із суворим порядком: спочатку брандмауер VNF, потім DPI VNF і нарешті, тунелювання VNF. Вузли NFVI мають збільшену пропускну здатність від краю до ядра: 3 процесора і 40 ГБ оперативної пам'яті на кожному крайовому вузлі, 5 процесорів і 80 ГБ оперативної пам'яті на кожному вузлі агрегації і 10 процесорах і 160 ГБ оперативної пам'яті на основних вузлах. Реалізація алгоритму VNF-PR була проведена з використанням AMPL і CPLEX 12.6.3.0.

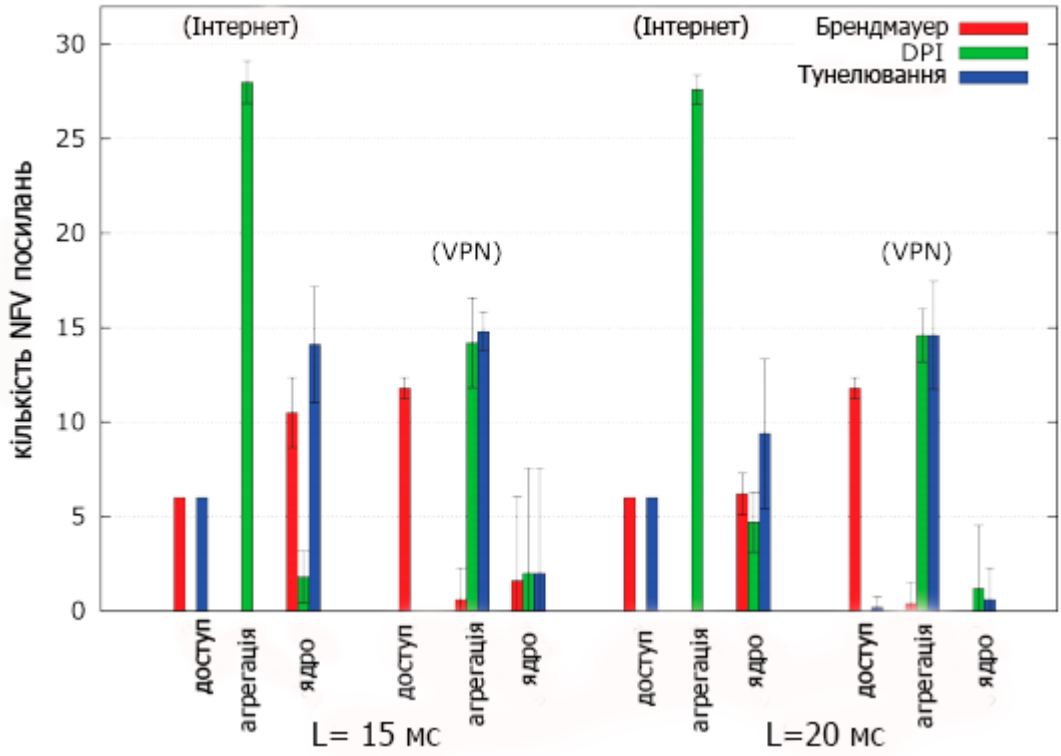


Рисунок. 4.5 Розподіл вузлів VNF за рівнями NFVI (стандартний випадок) при TE

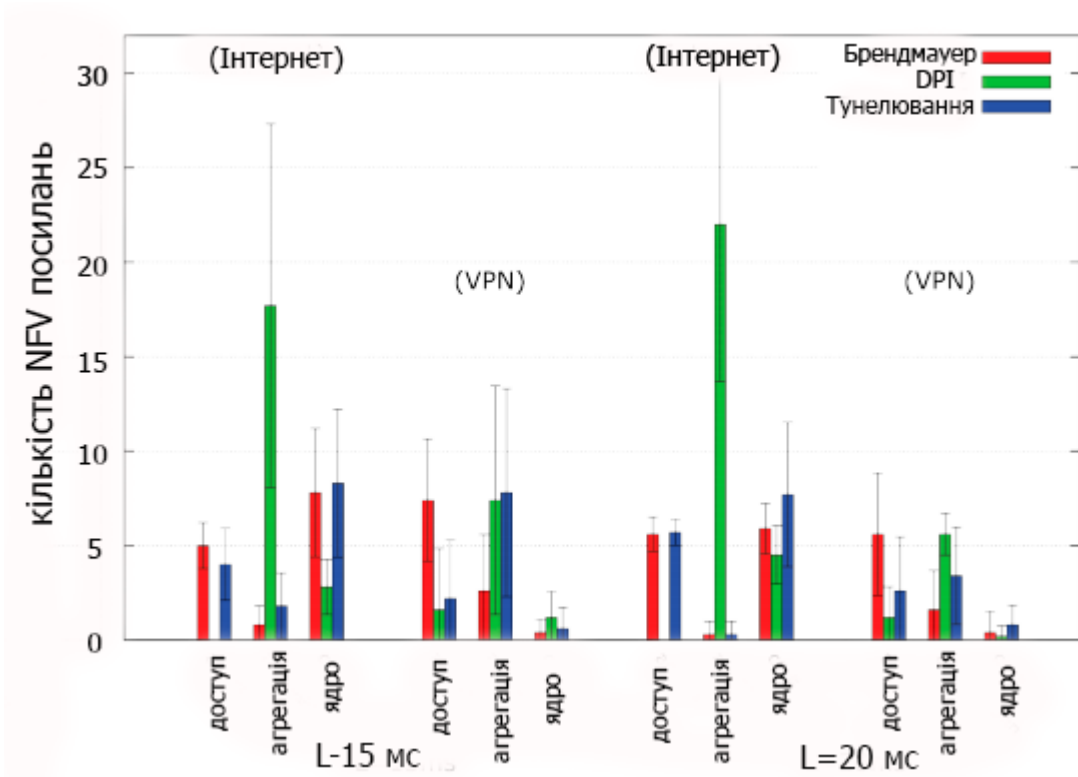


Рисунок. 4.6 Розподіл вузлів VNF за рівнями NFVI (стандартний випадок) при TE-NFV

4.4.1 Загальні результати

Тестувалися як інтернет-тести, так і VPN-тести в стандартному, а також в профілях затримки fastpath, час затримки обробки VNF задано як на рис.. Час виконання було обмежено до 600 с для кожної базової фази оптимізації TE і 800 с для повної TE і NFV фази оптимізації.

У всіх тестах найгірший розрив оптимальності з метою TE становить 25%. Деякі рішення виявилися оптимальними протягом 800 (тобто граничний час виконання фази оптимізації TE), інші були з середнім інтервалом оптимальності 15%. Однак ці субоптимальні рішення були швидко знайдені (протягом декількох секунд) і були б оптимальні, якщо для майже всіх випробувань було задано більше часу (близько 1 або 2 годин).

Для мети NFV було важко досягти оптимального значення в межах 800 с. Результати залежать від завдання: до вимог VPN, як в стандартних, так і в швидких профілях затримки, було отримано більш низький розрив в оптимальності і менша варіація, ніж до вимог Інтернету. Можливим поясненням є збільшення кількості вимог до трафіку за допомогою інтернет-тематичного дослідження, яке, як видається, значно впливає на обчислювальні зусилля. Аналіз поведінки рішень: порівнюючи два різних тематичних дослідження попиту (наприклад, Інтернет та VPN) з двома точками зору: 1) що відбувається, коли розглядається вартість NFV в цільовій функції замість TE; 2) що відбувається, коли підсилюється прив'язка на кінцевій затримці. Також порівнюється поведінку по відношенню до профілів затримки.

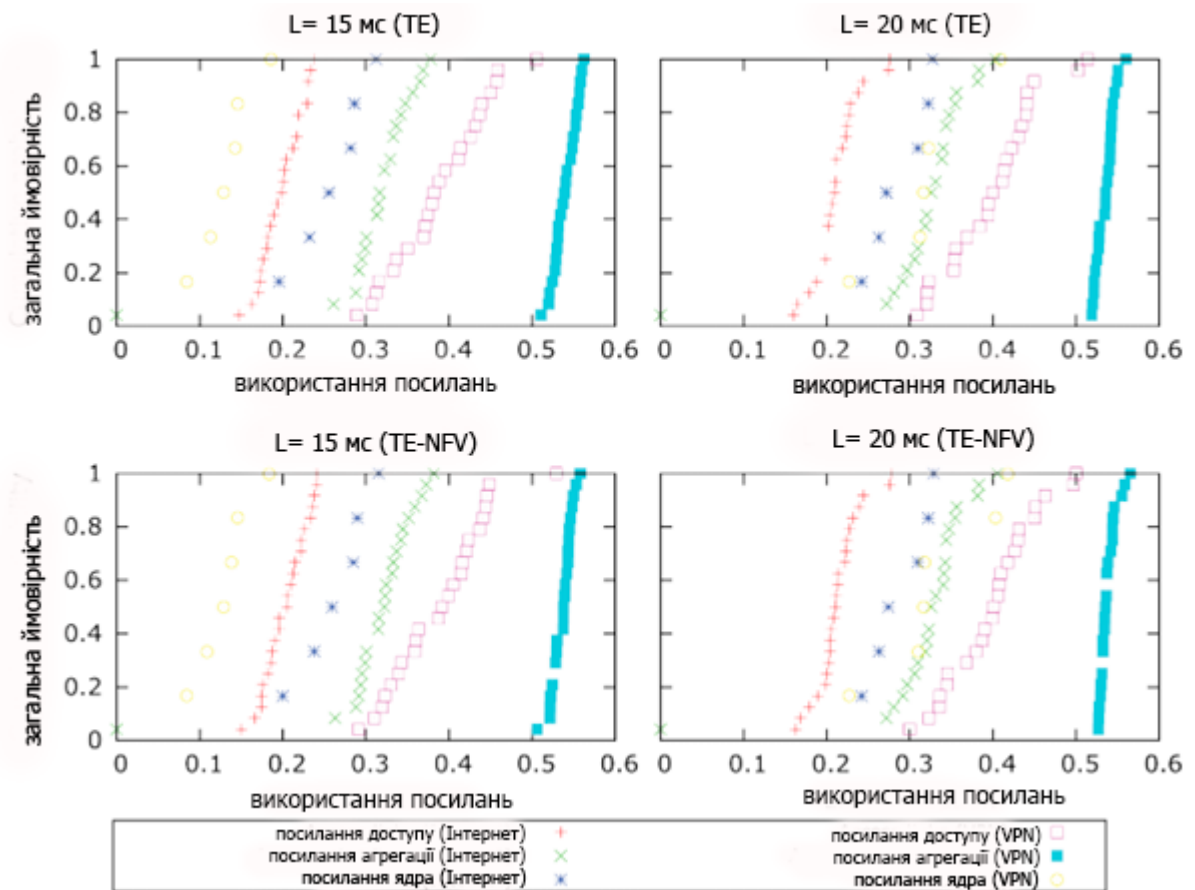


Рисунок. 4.7 Використання емпіричних CDF (стандартний випадок)

1) Порівняння TE та TE-NFV: аналіз різниці між результатами з метою TE і результатами з складовою метою TE-NFV.

Кількість NFV посилок (рис. 4.5 та рис. 4.6): значно знижується у випадку TE-NFV, а зниження за допомогою вимог VPN є більш значним, ніж вимоги до Інтернету, особливо з обмеженою прив'язкою до наскрізної затримки ($L = 20$ мс).

Використання посилок (рис. 4.7): на нього не робить істотного впливу включення мінімізації кількості NFV посилок в ціль оптимізації.

Затримка переадресації VNF (рис. 4.8): прийняття мети TE-NFV дозволяє більш високий рівень спільного використання VNF для обох пов'язаних із затримкою ситуацій, як для Інтернет випадку, так і VPN.

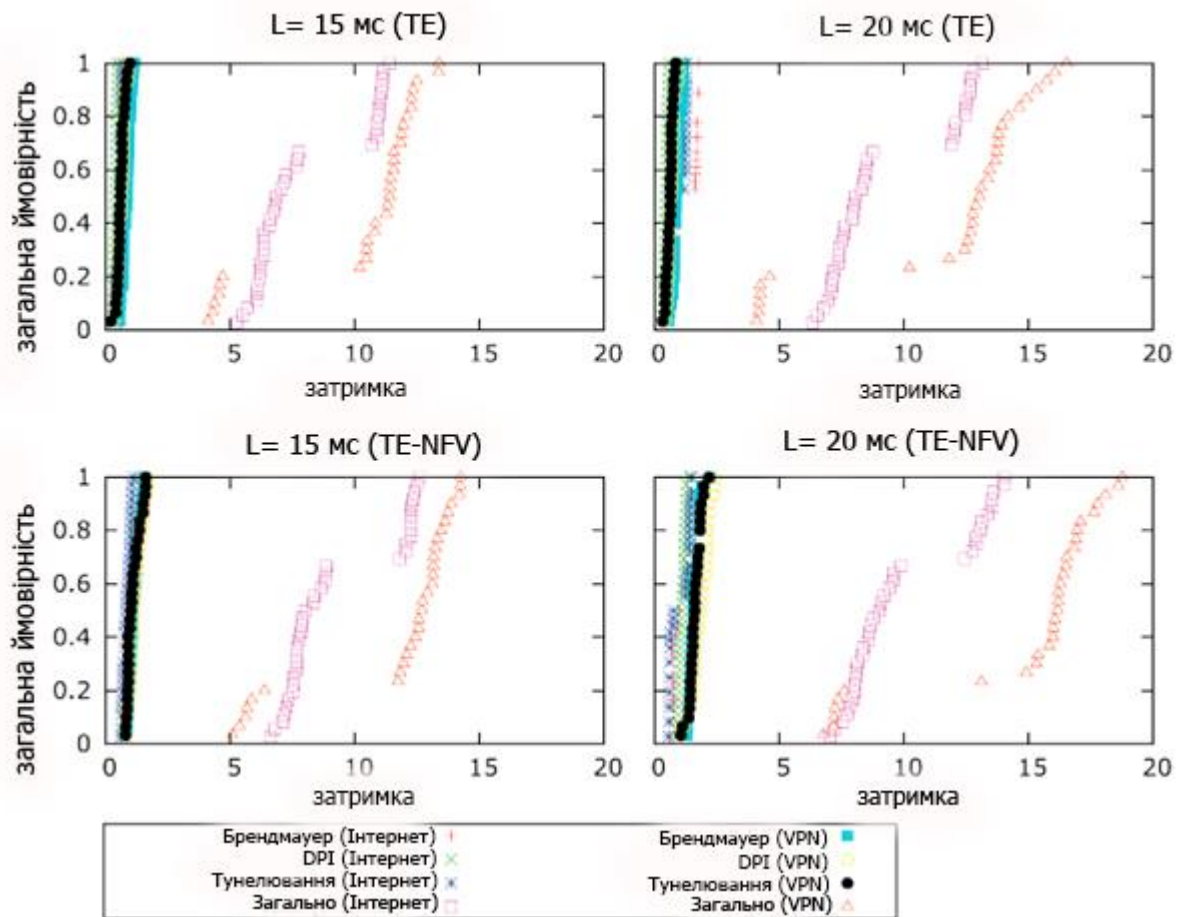


Рисунок. 4.8 Емпіричні CDF-компоненти компонентів з затримкою
(стандартний випадок)

2) Ослаблення обмеження TE: чутливість до максимального використання послань: було проведено аналіз чутливості, щоб продемонструвати ефект ослаблення мети TE щодо оптимальної кількості NFV з метою подальшого доказу взаємозв'язку між двома випадками. З метою TE-NFV, навіть якщо кількість розміщення VNF мінімізована, гарантується мінімальне максимальне використання послань. Щоб аналізувати вплив прив'язки TE на об'єктну оптимізацію витрат NFV, необхідно починати з оптимального значення TE. Для цього виконується серія кроків оптимізації цільової функції затрат NFV, дозволяючи все це пом'якшуватися.

Кількість NFV (в середньому 10 виконань) при різній межі максимального використання послань (U). Обчислюється кількість NFV при

$U = U^* + \alpha$, причому α змінюється від 0 до 0,4. Для випадків Інтернету і VPN, коли $\alpha = 0$, ТЕ-кордон (U), яка використовується для фази ТЕ-NFV, є U^* , знайденої в ТЕ-фазі; коли $\alpha = 0,4$, U , який використовується для фази ТЕ-NFV, становить близько 1 (тобто досягнута насиченість каналу). Результати показують, що вільна ТЕ-прив'язка (використання посилянь) дозволяє поліпшити рішення ТЕ-NFV. У більшості випадків практично відсутня скорочення (або взагалі відсутня) від $\alpha = 0,2$ до $\alpha = 0,4$, що свідчить про наявність стелі між метою ТЕ і метою ТЕНFV: краще використовувати ресурси NFV (тобто мета ТЕ -NFV), дозволяючи обмеження використання вільної лінії (тобто мета ТЕ), проте це не завжди вірно, коли досягаємо стелі (наприклад, інші обмеження, такі як ємність VM, також впливають на вартість NFV). У той час як для випадку Інтернет Fastpath $L = 20$ мс відбувається зниження кількості NFVI з ростом α від 0,2 до 0,4. Помітно, що для одного і того ж тематичного дослідження з $L = 15$ найкраща знайдена мета - 31 (як для $\alpha = 0,2$ і 0,4), тому припускаємо це зміна поведінки до неоптимальності рішення в разі $L = 20$ мс, а не на іншу поведінку системи (було накладено короткий термін).

Змінна	$\alpha = 0$	$\alpha = 0.2$	$\alpha = 0.4$
	L= 15 мс		
Стандартний Інтернет	48.2	25.8	24.9
Fastpath Інтернет	37.125	31	31
Стандартний VPN	31	28.8	28.6
Fastpath VPN	39.1	37.7	37.7
	L= 20 мс		
Стандартний Інтернет	52	23.7	23.1
Fastpath Інтернет	38.7	34.8	31.2
Стандартний VPN	20.9	20.4	20
Fastpath VPN	34.9	34.8	33.8

Рисунок. 4.9 Кількість NFV для різних рівнів релаксації ТЕ, з оптимізацією ТЕ-NFV

3) Чутливість до затримки: аналіз впливу затримки ланцюга VNF на зв'язок (L) з результатами.

Кількість NFV (рис.4.5 та рис. 4.6): світова ціна NFV залишається майже незмінною, переходячи від слабкої до сильної затримки. Загальна вартість знижується завдяки вимогам VPN по обидва цілям оптимізації з обмеженою затримкою, особливо з метою TE-NFV. Це відбувається через те, що з обмеженою затримкою трафік може проходити через посилення з високою затримкою (наприклад, основні посилення), щоб розділити більше VNF. Навпаки, є невелике збільшення вартості до потреб Інтернету по цілі TE-NFV. Аналізуючи більш детально результати, очевидно, що для інтернет-тематичного дослідження вирішувач (CPLEX) має більше труднощів для зменшення розриву. Було зроблено висновок про те, що спрощення слабкої затримки робить компонент проблеми місця розташування (визначення цілей VNF і NFV) переважаючим по відношенню до маршрутизації, і тому проблема стає більш складним завданням. Це також підтверджується помітною мінливістю результатів, з великим довірчим інтервалом в кількості VNF з вимогами до Інтернету, перехідним від строгого латентного стану до вільних латентним кордонів. Більш того, з урахуванням вимог VPN існує більш висока залежність від затримки, ніж до вимог Інтернету. Це відбувається тому, що, загалом, далі відправляти запити трафіку з прикордонного вузла на прикордонний вузол, ніж з вузла edge (core) node на core (edge), тобто наскрізний шлях пересилання запитів VPN в загальному випадку довше, ніж вимог до Інтернету, що призводить до більш високої залежності від затримки.

Використання посилень (рис. 4.7): в підтримку вищезгаданого аналізу помітно, що при обмеженій затримці прив'язки основні посилення стають більш використовуваними до вимог VPN.

Затримка VNF-пересилання (рис.4.8): таке ж спостереження можна отримати, переглянувши компоненти наскрізної затримки, затримка кожного VNF і загальна затримка стають явно довшими до вимог VPN з обмеженою затримкою.

Ці спостереження підтверджують важливість прив'язки рішень ланцюжка VNF і розміщення.

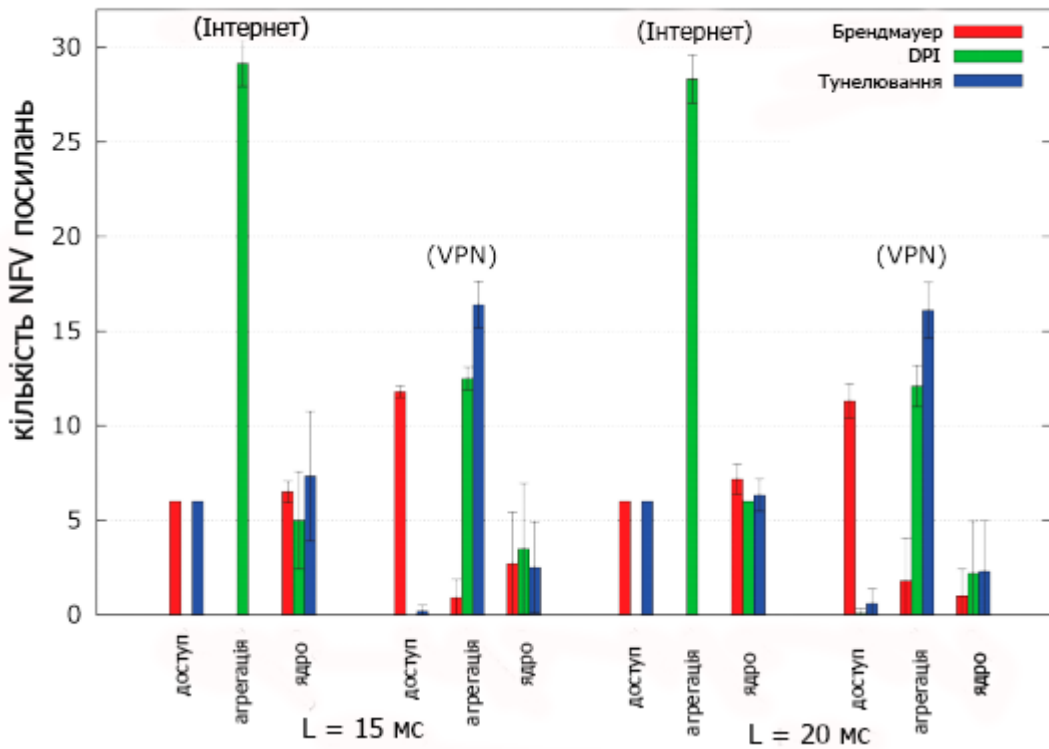


Рисунок. 4.10 Розподіл вузлів VNF за рівнями NFVI (випадок fastpath) при TE

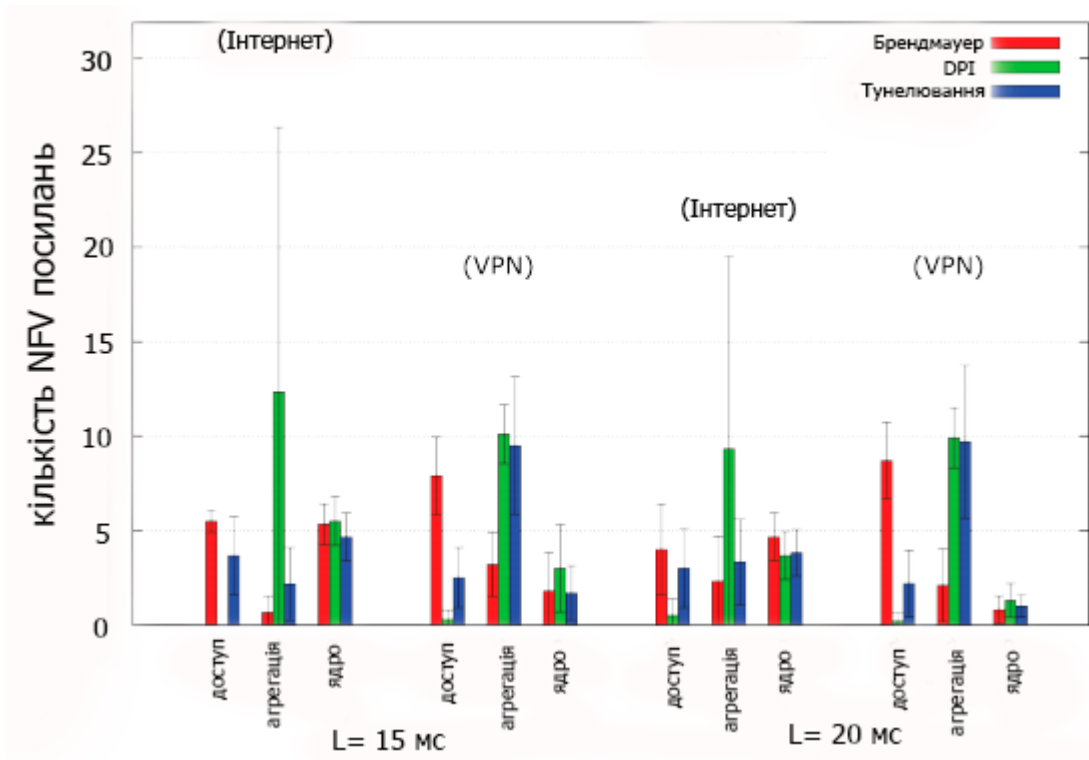


Рисунок. 4.11 Розподіл вузлів VNF за рівнями NFVI (випадок fastpath) при TE-NFV

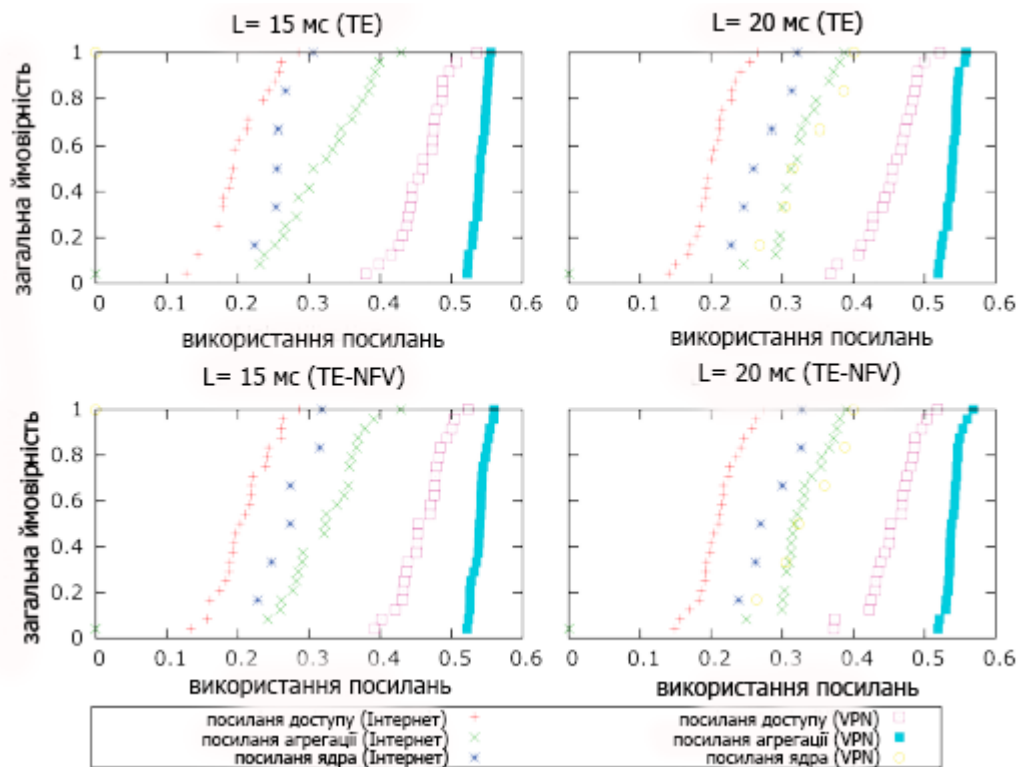


Рисунок. 4.12 Використання емпіричних CDF-посилань (випадок fastpath)

4) Перемикання VNF в стандартному або fastpath випадку: порівнюється результати зі стандартним профілем затримки пересилання VNF з даними з профілем fastpath.

Кількість NFV (рис. 4.5 та рис. 4.6 в порівнянні з рис.10 та рис. 4.11): за метою TE-NFV прискорена пересилання VNF з прискореної швидкістю більша, ніж стандартна пересилання з вимогами VPN, особливо з обмеженою прив'язкою до наскрізної затримки ($L = 20$ мс), в той час як це суперечить вимогам Інтернету. Це відбувається через максимальну прив'язку трафіку, яка задана у випадку fastpath, і яка не встановлена для стандартного випадку (що, однак, призводить до більш високої наскрізної затримки).

Використання посилок (рис. 4.7 в порівнянні з рис. 4.12): між двома профілями затримки немає помітних відмінностей.

Затримка пересилання VNF (рис. 4.8 в порівнянні з рис. 4.13): VNF краще розподіляються в стандартному випадку, це пов'язано з максимальним обмеженням трафіку, встановленим у випадку з fastpath.

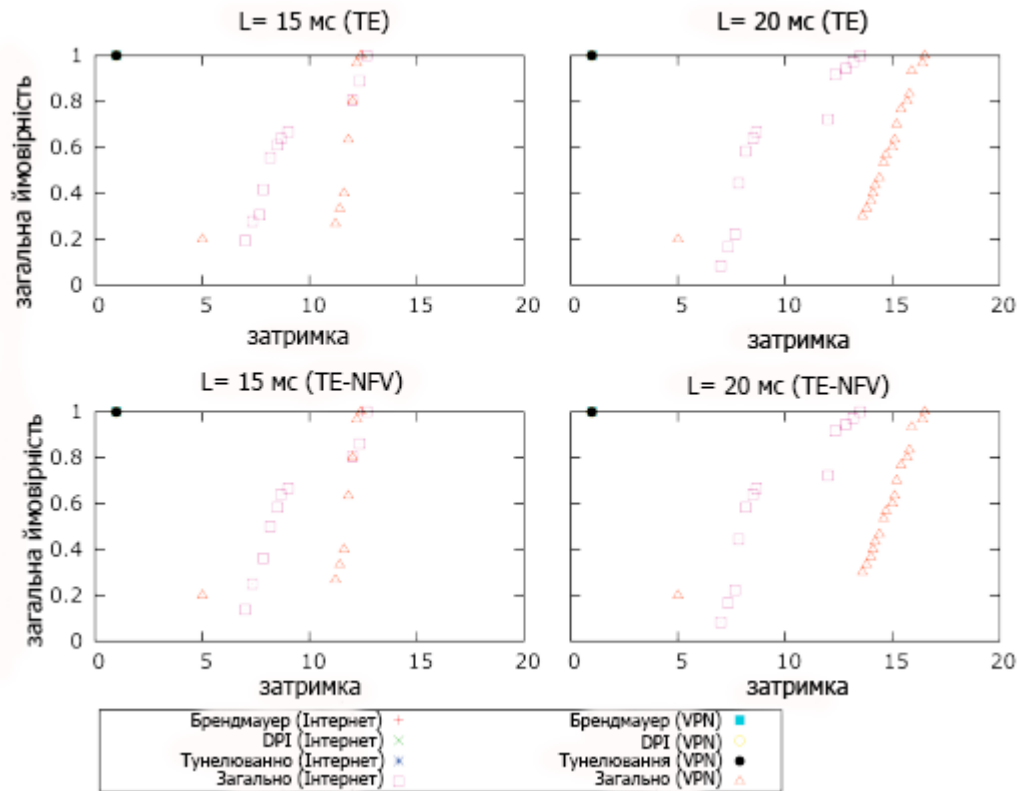


Рисунок. 4.13 Емпіричні CDF-компоненти латентних компонентів (випадок fastpath)

Ці спостереження показують, що, виконуючи оптимізацію в два етапи (спочатку TE, а потім мінімізацію витрат NFV), може дозволити значно знизити кількість NFV посилань, не впливаючи на розподіл використання посилань.

4.4.2 Підходи, засновані на VNF-PR і VNE

Було порівняно запропонований підхід VNF-PR з підходом VNE Based (VNE-B). Пропонується підхід моделювання VNE-B для загальної проблеми оркестровки VNF: кожен запит на трафік розглядається як віртуальний граф ($G(N, L)$, де N - це набір трафіку вузла, тобто комутатори або VNF, а L позначає посилання між ними), які повинні бути вбудовані в графік субстрату, представлений комутаторами / маршрутизаторами і кластерами NFVI. Відображення шляху віртуального трафіку вимагає фізичної мережі,

реалізується шляхом впровадження VNF на фізичних серверах і встановлення шляху для віртуальних посилань. Вже згадана мета полягає в мінімізації загальної кількості ОППХ (операційних витрат): вартості розгортання VNF, вартості енергії, вартості пересилки трафіку і додаткового штрафу, щоб брати до уваги порушення рівня обслуговування (SLO). В якості мети оптимізації розглядається зважена сума чотирьох вищезазначених витрат. В запропонованому порівнянні був фокус на статичній версії проблеми і її запропонованому підході до вирішення. Він заснований на процедурі, яка вирішує послідовність ILP, де для кожної ітерації кількість VNF обмежена і час виконання також обмежений. Виконання ILP вирішуються за допомогою CPLEX.

Перерахуємо спрощення і адаптації до запропонованої моделі, щоб використовувати ті ж параметри, що і в підході VNE-B:

- було знижено поставлене в роботі завдання до однієї мети: мінімізації загальних експлуатаційних витрат мережі, використовуючи ті ж параметри підходу VNE-B;
- було розглянуто тільки режим затримки fastpath, тобто фіксувалась затримка пересилання VNF;
- було відкинуто аспекти стиснення / декомпресії, тобто була застосована модель «basic-lat»;
- оскільки підхід VNE-B використовує шаблони VNF, було пов'язано шаблон VNE з кожним типом VNF відповідно до запитаної VNF кількості vCPU (наприклад, шаблон з ємністю 4 процесора пов'язаний з VNF, робить запити 4 процесора);
- було додано параметр штрафу для кожного запиту трафіку, щоб прийняти до уваги порушення SLO.

Що стосується тестових даних, вони були встановлені відповідно з установкою імітації, прийнявши топологію Internet2 (12 комутаторів і 15 посилань), встановивши один і той же фізичний зв'язок і можливості сервера

NFVI, використовуючи ту ж специфікацію VNF, послідовність запитів VNF, і т. д., і приймаючи ті ж дані про затрати. Що стосується даних трафіку, було створено 5 груп тестів з різними наборами запитів трафіку (6, 12, 18, 24 і 30). Для кожної групи випадковим чином вибирається з набору матриць трафіку з 10 матриць. Потім ці групи даних тестувалися за допомогою трьох методів (VNE-B, VNF-PR і VNF-PR-D). Запропоновані алгоритми реалізовані в AMPL, а алгоритм VNE-B реалізований на C++ з використанням CPLEX для вирішення ILP. CPLEX 12.5.1 був використаний для цих випробувань. На рис. показано порівняння запропонованого рішення VNF-PR і рішення VNE-B з точки зору глобальної вартості в залежності від кількості розглянутих вимог. Можна помітити, що VNF-PRD може знайти кращі рішення в порівнянні з алгоритмом VNE-B, особливо коли кількість вимог до трафіку невелика: він завжди може знайти оптимальні рішення. У той час як алгоритм VNE-B завжди здатний знаходити прийнятне рішення дуже швидко (приблизно протягом 10 секунд), результат може бути набагато гірше, і він не забезпечує оптимальних рішень навіть по 6 вимогам. Що стосується VNF-PR, то це точний метод для знаходження оптимального рішення, тому він в цілому повільний. Він запускається, щоб знайти оптимальне рішення, щоб показати розрив між оптимальними рішеннями і можливими рішеннями. На рис. показано, що розрив можливих рішень стає дуже важливим, оскільки запити трафіку збільшуються. Однак розрив можливих рішень, який був знайдений VNF-PR-D, менше, ніж у VNE-B, що говорить про те, що для досягнення компромісу під час виконання і оптимальності рішення необхідні більш складні формулювання проблем.

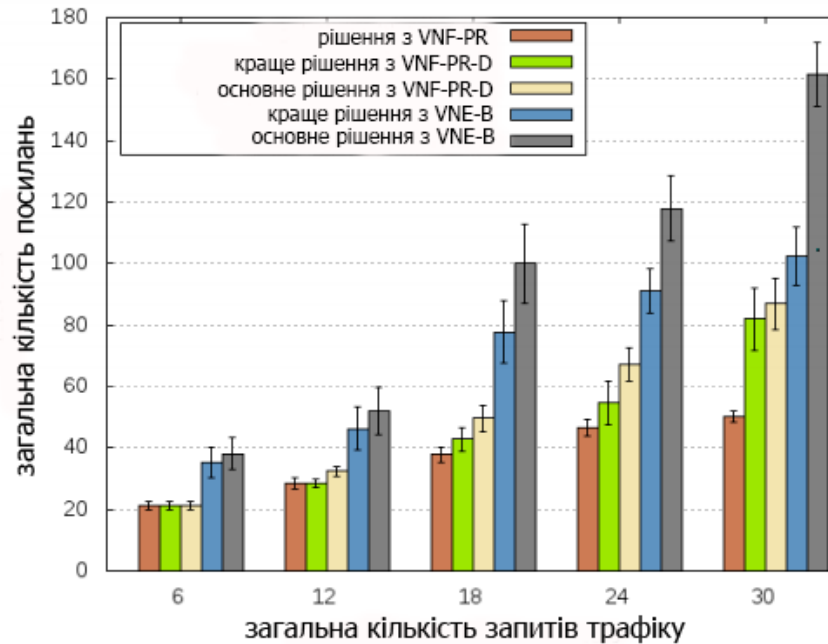


Рисунок. 4.14 Порівняння алгоритмів VNE-PR і VNE з точки зору об'єктивної функції

4.5 Висновки з розділу 4

У цій роботі було запропоновано модель ланцюжка і розміщення VNF, включаючи алгоритм, сформульований як змішана цілочисельна лінійна програма. Запропонована модель враховувала специфічні режими пересилки VNF (стандартні і швидкодіючі режими), обмеження порядку ланцюжків VNF, а також зміни бітрейта потоку. Ці обмеження роблять виділення крайових вимог по цілям VNF унікальними, але складними. Також було згадано, як додаткові властивості, спеціалізовані для NFV-систем, можуть бути інтегровані в запропоноване формулювання. Для того, щоб освоїти часову складність алгоритму вирішення, розглядаючи дві різні цілі оптимізації - єдину мету трафіку (TE) та мету TE в поєднанні з ціллю мінімізації витрат на інфраструктуру NFV (TE-NFV) - було спроектовано та розглянуто математико-евристичний метод розв'язання. Крім того, було порівняно запропонований підхід VNF-PR до класичного підходу VNE, який часто пропонується в літературі для оркестровки NFV.

Було проведено великі тести для оцінки запропонованого алгоритму на тривірневій топології, яка представляє топологію ISP. Результати показали, що об'єднана мета TE-NFV значно знижує кількість VNF в мережі в порівнянні з метою TE, практично не впливаючи на використання посилянь та затримку. Крім того, було відзначено, що поряд з різними цілями оптимізації (TE і TE-NFV) різні розподіли попиту на трафік (інтернет-і VPN-тематичні дослідження) і різні типи VNF (з точки зору функції по бітрейту) можуть привести до різного місцерозміщення вузлів VNF і різних ланцюгів VNF.

Також кількісно порівняно запропоновану модель VNF-PR із застарілою моделлю VNE. Експериментальні результати показали, що запропонований алгоритм VNF-PR був більш стабільним і близьким до оптимального, ніж алгоритм VNE. Дослідження також показало, що штраф за порушення SLO становив 0 майже у всіх тестах, виконаних з використанням підходу VNF-PR, тоді як штраф за порушення SLO завжди існував для всіх тестів з використанням VNE. Це підтвердило, що запропонований алгоритм VNF-PR краще визначає наскрізну маршрутизацію обслуговування. Крім того, середня вартість пересилки всіх тестів, які були вирішені за допомогою алгоритму VNE, була приблизно в 4 рази дорожче, ніж алгоритм VNF-PR, що вказує на те, що в порівнянні з рішенням VNE було значно менше надлишкових маршрутів трафіку при зв'язуванні VNF з підходом VNF-PR.

ВИСНОВКИ

LTE цілком може відповідати всім найбільш важливим вимогам 5G, включаючи підвищення швидкості передачі даних і підвищення пропускну здатності системи за допомогою FD-MIMO, поліпшену підтримку неліцензованих операцій і скорочення затримки як в площині користувача, так і в сигналізації. Покращення, які заплановані в Rel-14 і Rel-15, не тільки гарантують, що LTE забезпечить кращу підтримку для масових MTC і ITS, вони також дозволять LTE вирішувати нові варіанти використання, такі як критичні комунікації.

Кілька прикладів того, як найбільші мережеві постачальники в світі змінюються, щоб позиціонувати себе для 5G, коли вони зливаються з віртуалізацією. Великі виробники мікросхем, такі як Broadcom, Cavium і Intel не просують ні однієї можливості на ринку мереж 5G. Наслідки на ринках чіпів для споживчих пристроїв, що також будуть величезним

Цей приклад підкреслює основні досягнення в проекті FP7 для мобільних хмарних мереж EC, надаючи сервіс-орієнтовану архітектуру, яка підтримує комплексні і складені служби з використанням технологій хмарних обчислень і SDN.

Деякі оператори вимагали більш швидкого і органічного розвитку платформ NFV. Деякі оператори навіть склали свої власні технологічні програми, в тому числі розробляють свої власні проекти з відкритим вихідним кодом, які потім передають спільноті (наприклад, OSM / OpenMANO від Telefonica, ONAP, що включає ECOMP від AT & T і Open-O від China Mobile).

Рівень NFVI складається з програмного забезпечення для апаратного забезпечення і віртуалізації, яким керують VNF. Цей рівень - це просто канал передачі даних і не має відношення до будь-яких аналізу планування, організації, забезпечення, моніторингу, обслуговування і т. д.

Однією з проблем для операторів зв'язку в розгортанні NFV є забезпечення продуктивності користувацької площини. У рішеннях NFV від Huawei робляться численні заходи для оптимізації продуктивності апаратних, хмарних ОС і шарів VNF, що в кінцевому підсумку підвищує продуктивність користувацької площини у три рази.

Huawei і China Mobile спільно продемонстрували функції, дизайн і бізнес-логіку мережевих частин, а також те, як оператори використовують мережеві частини на основі SOC для задоволення різних вимог до мережі великою кількістю додатків в майбутньому.

Також Huawei і Vodafone продемонстрували розумне водіння на основі SOC, а Huawei і Deutsche Telekom представили три наскрізні орієнтовані на 5G мережеві частини, засновані на ОС 5, для комерційного використання: частина xMBB для сервісів AR / VR, GLA частина для IoV і промислового контролю, а також частина FMC для забезпечення безперебійного обслуговування відео.

В даний час більше 70% світових операторів почали хмарну реконструкцію мережі. Компанія Huawei і її партнери по галузі планують продовжити просування промислового розвитку NFV вперед, щоб допомогти більшій кількості компаній трансформуватися в хмару.

Було визначено і сформульовано за допомогою математичного програмування задачі оптимізації VNF Placement and Routing (VNF-PR), включаючи обмеження стиснення / декомпресії і два режими затримки пересилання (з і без швидкого шляху), як в рамках цілей TE, так і NFV. Також було проведене якісне і кількісне порівняння підходу VNF-PR до існуючого підходу Virtual Embedding (VNE). Розроблено математичний підхід, що дозволяє також запускати експерименти для великих випадків проблеми протягом прийнятеного часу виконання. Оцінка рішення була проведена шляхом великого моделювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. P. Gutierrez and J. Carapinha, "Cloud Networking: Implications of Agile Virtualisation on Provider Relationships", Electronic Communications of the EASST, 2011
2. ETSI, Network Functions Virtualisation (NFV): Architectural Framework, Technical Report ETSI GS NFV 002 v1.1.1, Oct. 2013.
3. J. Carapinha, et.al. "D2.1: Reference Scenarios and Technical System Requirements Definition", Mobile Cloud Networking FP7 project, April 2013.
4. G. Xilouris¹, et al., "T-NOVA: A Marketplace for Virtualized Network Functions", European Conference on Networks and Communications (EUCNC), June 2014.
5. Monteleone, G.; Paglierani, P., "Session Border Controller Virtualization Towards "Service-Defined" Networks Based on NFV and SDN," Future Networks and Services, IEEE SDN, Nov. 2013
6. Batalle, J.; et al., "On the Implementation of NFV over an OpenFlow Infrastructure: Routing Function Virtualization," Future Networks and Services (SDN4FNS), IEEE SDN, vol., no., pp.1,6, 11-13 Nov. 2013
7. MCN, "MCN," <http://blog.zhaw.ch/icclab/files/2012/06/MCNPartners.png>, Nov 2015.
8. FOKUS, "OpenIMSCore," <http://www.openimscore.org/>, Nov 2015.
9. "OpenIMSCore toolkit," <http://www.openimscore.org/index.html>, Nov 2015.
10. "Open5gCore," www.open5Gcore.net, Nov 2015.
11. MCN, "MCN demonstration preview",
<https://www.youtube.com/user/MobileCloudNetworkin>, Nov 2015.
12. Network Computing, First Commercial LTE Network Goes Live, available at: <http://www.networkcomputing.com/networking/first-commercial-lte-network-goes-live/752107374>
13. Ericsson, Ericsson Mobility Report 2016, November 2016

14. David Astšly et al., LTE: The Evolution of Mobile Broadband, IEEE Communications Magazine, April 2009
15. Stefan Parkvall et al., Evolution of LTE toward IMT-Advanced, IEEE Communications Magazine, February 2011
16. David Astšly et al., LTE Rel-12 and Beyond, IEEE Communications Magazine, July 2013
17. Juho Lee et al., LTE-advanced in 3GPP Rel-13/14: an evolution toward 5G, IEEE Communications Magazine, March 2016
18. ITU-R, IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M.2083-0, September 2015
19. 3GPP Technical Report 38.913, Study on Scenarios and Requirements for Next Generation Access Technologies, October 2016
20. C. Hoymann et al., LTE Rel-14 Outlook, IEEE Communications Magazine, June 2016