

ВСЕУКРАЇНСЬКА ПРОГРАМА ПІДТРИМКИ СТУДЕНТІВ
«Професіонали майбутнього»
ДЛЯ СТУДЕНТІВ ПРОФІЛЬНИХ ВИЩИХ НАВЧАЛЬНИХ ЗАКЛАДІВ
за напрямом підготовки «ТЕЛЕКОМУНІКАЦІЇ»



Державний вищий навчальний заклад
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ



ЗВІТ

про виконання проекту на тему
**«Створення технологічних ресурсів у конвергентній мережі на базі
мобільного оператора для надання послуги Triple Play»**

Узгоджено:
декан факультету
«Комп'ютерні інформаційні
технології та автоматика» ДонНТУ
_____ В.В. Турупалов

Виконавці проекту:

Студенти IV курсу напрямку підготовки «Телекомунікації» спеціальності «Телекомунікаційні системи і мережі»:

Абраменко Олексій;

Будішевський Денис;

Зюмін Сергій;

Патрушев Сергій;

Чекунков Олександр.

Науковий керівник: к.т.н., доцент кафедри «Автоматика і телекомунікації» (АТ) Дегтяренко Ілля Вячеславович.

Консультанти:

д.т.н., професор кафедри АТ Воронцов Олександр Григорович;

к.т.н., доцент, завідувач кафедри АТ Бессараб Володимир Іванович;

студент групи ТКС-10м факультету КІТіА Шахов Дмитро Сергійович.

Виконавці проекту висловлюють подяку керівництву та співробітникам Східного територіального управління ПрАТ «МТС Україна» за підтримку проекту.

ЗМІСТ

ВСТУП.....	4
1 НОВИЗНА ПРОЕКТУ	5
2 МЕТОДИ ДОСЛІДЖЕННЯ.....	5
3 ОПИС ПРОЕКТУ	6
3.1 Вибір концепції організації та синтез структури технологічних ресурсів різних рівнів конвергентної мережі.....	6
3.1.1 Технологічні ресурси на рівні доступу	6
3.1.2 Технологічні ресурси на рівні транспорту.....	6
3.1.3 Технологічні ресурси на рівні керування.....	7
3.1.4 Технологічні ресурси на рівні послуг	8
3.2 Розробка критерію оцінки якості надання послуги Triple Play	9
3.3 Розробка математичної моделі якості обслуговування на рівні керування та послуг конвергентної мережі	10
3.3.1 Залежність параметрів QoS від завантаженості серверів.....	10
3.3.2 Математична модель роботи кластера серверів	11
3.4 Синтез структурно-алгоритмічних рішень для підвищення якості надання послуги Triple Play	12
3.4.1 Статичне балансування навантаження на сервери.....	12
3.4.2 Динамічне балансування навантаження на сервери.....	13
3.4.3 Використання нейромережі для прогнозування навантаження	13
3.4.4 Розробка структури та алгоритму роботи системи динамічного балансування навантаження на сервери.....	14
3.5 Імітаційне моделювання системи балансування навантаження ..	16
4 ПОТРЕБИ ФІНАНСОВИХ ВИТРАТ	17
РЕЗУЛЬТАТИ ВИКОНАННЯ ПРОЕКТУ	17
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	18

ВСТУП

Значне зростання інтересу користувачів до різноманітних інтелектуальних послуг у мережах мобільних операторів — одна з основних тенденцій розвитку ринку телекомунікацій. Нині стрімко набувають популярності послуги Triple Play — надання абонентові одночасно трьох сервісів («тріади»): передача аудіо реального часу, відео реального часу, даних із високою швидкістю. Згідно з концепцією Triple Play, на базі тріади комбінуються різноманітні мультимедійні послуги: телефонія (VoIP), високошвидкісний доступ до Інтернету, телевізійне мовлення (IPTV), відеоконференції, онлайн-ігри, відео за запитом (VoD), хмарні обчислення (Cloud Computing), інтелектуальне відеоспостереження, керування системами «розумного будинку», спеціалізовані послуги для бізнес-абонентів тощо. Різні послуги надаються абонентові через єдину інфраструктуру мультисервісної конвергентної мережі.

Необхідність розгортання послуг Triple Play спонукає операторів мобільного зв'язку модернізувати власні мережі згідно з найсучаснішими стандартами, отож, наразі окреслюється тенденція модифікації існуючої мережної інфраструктури операторів зв'язку згідно з концепцією NGN [1]. Перехід до конвергентної NGN означає побудову мережі згідно з ієрархічною архітектурою, яка має чотири рівні: абонентського доступу, транспорту, керування, логіки послуг. Функції цих рівнів виконує окреме спеціалізоване апаратне та програмне забезпечення. На кожному з рівнів оператору необхідно створити технологічні ресурси, які уможливають надання послуги Triple Play.

В умовах конкурентної економіки прибуток оператора зв'язку у значній мірі визначається ступенем задоволення абонентів якістю та асортиментом послуг у мережі. Тому технологічні ресурси у конвергентній мережі мають забезпечити можливість надання широкого асортименту сучасних послуг Triple Play з високою якістю (необхідно дотримання вимог QoS, описаних в [2]).

Отже, **проблема**, що виникає під час побудови та експлуатації конвергентної мережі з наданням послуги Triple Play, полягає у необхідності вибору технологічних ресурсів та узгодження їх параметрів з вимогами QoS, які суттєво відрізняються для трафіка різних послуг.

Тому актуальною для сучасного мобільного оператора є **мета** проекту: надання високоякісних послуг Triple Play у конвергентній мережі за рахунок формування та ефективного використання технологічних ресурсів різних рівнів мережі.

Для досягнення мети необхідним є вирішення таких **задач**:

- 1) вибір концепції організації та синтез структури технологічних ресурсів різних рівнів конвергентної мережі, щодо надання послуги Triple Play;
- 2) розробка критерію оцінки якості надання послуги Triple Play;
- 3) розробка математичної моделі якості обслуговування на рівні керування та послуг конвергентної мережі;
- 4) синтез структурно-алгоритмічних рішень, щодо підвищення якості надання послуги Triple Play;
- 5) оцінка ефективності розроблених засобів та розробка рекомендації щодо їх впровадження.

1 НОВИЗНА ПРОЕКТУ

- Запропоновані технічні рішення для формування технологічних ресурсів на різних рівнях конвергентної мережі, що надає послуги Triple Play.
- Запропоновано інтегральний критерій оцінки ефективності роботи конвергентної мережі з використанням функцій корисності, сформованих на основі вимог QoS для послуг Triple Play.
- Розроблена математична модель роботи елементів конвергентної мережі рівня керування та послуг, що дозволяє оцінити параметри QoS.
- Отримали подальший розвиток методи динамічного балансування навантаження на сервери рівня послуг та керування, завдяки їх адаптації до характеристик потоку запитів та впровадження прогностичного керування з використанням нейромережі.
- Розроблена структура та алгоритм роботи системи динамічного балансування навантаження на сервери рівня послуг та керування, використання якої дозволяє покращити якість надання послуг Triple Play.
- Розроблено оригінальне програмне забезпечення та апаратний макет для імітаційного моделювання роботи системи балансування навантаження.

2 МЕТОДИ ДОСЛІДЖЕННЯ

Для розв'язання поставлених задач використовувалися методи теорії телетрафіка, методи теорії самоподібності, методи теорії мереж та системного аналізу, методи математичного моделювання, методи теорії нейронних мереж.

3 ОПИС ПРОЕКТУ

3.1 Вибір концепції організації та синтез структури технологічних ресурсів різних рівнів конвергентної мережі

3.1.1 Технологічні ресурси на рівні доступу

Для надання послуг Triple Play з високою якістю до технологічних ресурсів рівня доступу висувається вимога — створення високошвидкісних (бажано від 24 Мбіт/с) безпроводових та фіксованих каналів на ділянці «останньої милі». Для цього оператор може застосовувати у мережах фіксованого доступу технології GPON/FTTx; у мережах радіодоступу — UMTS/HSPA+MIMO, LTE, WiMAX.

Широкопasmові технології радіозв'язку переважно використовують більш високі несучі частоти, ніж GSM/EDGE (від 1900 МГц), тому загострюється проблема недостатнього рівня сигналу усередині приміщень. А оскільки, згідно з дослідженнями (зокрема, Cisco VNI), близько 60-70% мобільних з'єднань здійснюється із будинків, офісів, установ тощо, дуже важливо забезпечити у приміщеннях якісне покриття.

Використання фемтостільників [3] на рівні доступу дозволяє вирішити зазначену проблему. Фемтостільники випромінюють сигнал малої потужності, призначені для обслуговування територій від

15 до 50 м. Вони встановлюються безпосередньо у приміщенні, під'єднуються до проводового широкопasmового каналу Інтернет, через який здійснюється вихід у стільникову мережу (рис. 3.1). У мережі оператора інтегрується спеціальний шлюз RAN-Gateway, що працює з інформаційними потоками від фемтостільників.

Застосування фемтостільників на «останній милі» дозволяє оператору частково розвантажити радіомережу доступу та перекласти задачі обслуговування з'єднань і підтримання потужності сигналу усередині приміщень з зовнішніх базових станцій на фемтостільники.

3.1.2 Технологічні ресурси на рівні транспорту

Технологічні ресурси транспортного рівня мають забезпечити високу швидкість та гарантовану якість доставки трафіка з різними вимогами до QoS.

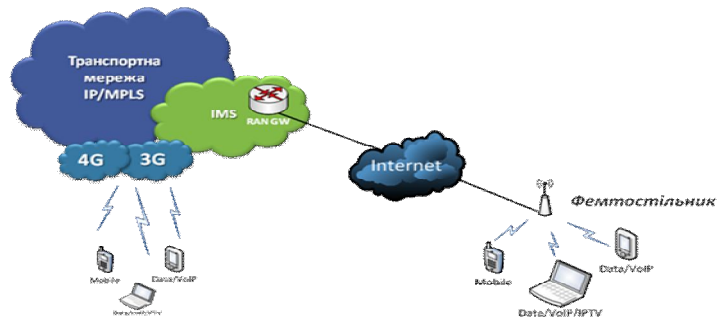


Рис.3.1. Схема підключення фемтостільника

Отже, транспортна мережа оператора повинна бути мультисервісною, побудованою на базі технології IP/MPLS (над NG-SDH, ATM, GE тощо). Методи підвищення ефективності роботи транспортної мережі IP/MPLS було розглянуто у проектах конкурсу «Професіонали майбутнього» у 2009 році.

3.1.3 Технологічні ресурси на рівні керування

Для надання послуги Triple Play на рівні керування мобільного оператора необхідно використовувати технологію Softswitch/IMS [4]. Головна функція IMS полягає у керуванні мультимедійними сесіями. Керування здійснюється, як правило, на базі протоколу SIP. Детально архітектура та функціонування підсистеми мультимедійного зв'язку IMS описана у [4].

Ефективним рішенням для оператора є впровадження мережі IMS, що самоорганізується — SO-IMS [5]. SO-IMS дозволяє автоматизувати мережні операції з метою підвищення надійності мережі та зменшення витрат оператора. Мережа SO-IMS безпосередньо впливає на параметри QoS, оскільки вона керує процесами передачі сигнальних та інформаційних пакетів, тому забезпечення надійної та ефективної обробки інформації у мережі IMS є необхідним для оператора зв'язку (див. рис. 3.2).

Для забезпечення стабільності сесій в мережі SO-IMS слід використовувати систему балансування SIP-навантаження на сервери P-CSCF [5]. Користувацьке обладнання надсилає повідомлення до балансира, надалі

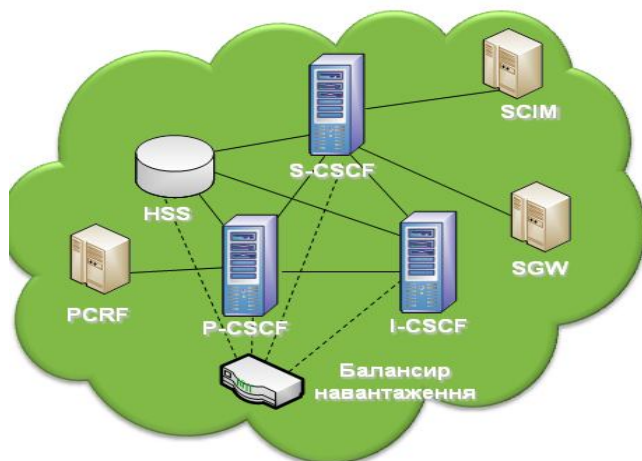


Рис. 3.2. Архітектура мережі SO-IMS із застосуванням балансування навантаження

система самостійно обирає потрібний P-CSCF в мережі за певними параметрами: територіальне розташування, завантаженість, справність тощо. Рішення дозволяє непомітно для користувача проводити реконфігурацію мережі (зокрема, зміна P-CSCF) та керувати завантаженістю P-CSCF [5].

Для підвищення якості обробки інформаційних пакетів у мережі SO-IMS слід використовувати систему балансування для вибору S-CFCS, що обслуговує сесії. Вибір сервера здійснює I-CSCF або головний вузол SO-IMS на підставі інформації про необхідні користувачеві послуги і можливості окремих S-CSCF,

враховуючи політику оператора і топологічну інформацію мережі [6]. Якість обслуговування сесій підвищується, якщо I-CSCF робить вибір S-CSCF за допомогою алгоритму балансування навантаження, спираючись на поточне завантаження серверів.

Отже, для підвищення якості роботи рівня керування актуальною є задача розробки інтелектуальних засобів балансування навантаження на сервери P-CSCF та S-CSCF. Засоби щодо її вирішення наведено в п. 3.4.

3.1.4 Технологічні ресурси на рівні послуг

Технологічні ресурси на рівні логіки послуг відіграють провідну роль у процесі надання послуги Triple Play мобільним оператором та практично визначають його конкурентоздатність на сучасному ринку зв'язку.

Ефективним рішенням для оператора зв'язку є застосування на рівні послуг спеціального програмно-апаратного рішення — платформи для розгортання послуг SDP [7]. Платформа являє собою архітектурне рішення нового покоління, націлене на швидку розробку і запровадження конвергентних мультимедійних послуг. Будується на базі стандартних серверних платформ. Різні виробники використовують свої підходи до побудови цього рішення. У структурі платформи існують програмно-апаратні інструменти створення послуг (SCE), виконання послуг (SEE), впровадження послуг (SDE), менеджмент послуг (SME). Доступ до SDP з боку мережних інтерфейсів та з боку програмістів здійснюється через стандартні відкриті програмні інтерфейси (API), наприклад, Parlay API, Parlay X API, SIP API.

Концептуально SDP являє собою міст між програмним кодом послуг та телекомунікаційною мережею. Застосування SDP на рівні логіки послуг надає оператору можливість зручного керування життєвим циклом послуги (PLM — Product Lifecycle Management), що включає етапи розробки, тестування, впровадження, модернізації, експлуатації, закриття. Оператор має можливість швидко реагувати на зміни у попиті на послуги, позаяк використання платформи дозволяє значно знизити собівартість та значно скоротити час виведення на ринок нових послуг [7]. Програмування послуг може здійснюватися персоналом оператора або замовлятися у сторонніх розробників, оскільки платформа забезпечує просту інтеграцію запрограмованих послуг у мережу.

Застосовуючи SDP, оператор забезпечує абонентам великий вибір сучасних послуг, що є дуже важливою перевагою компанії на ринку зв'язку.

Якість обслуговування на серверах SDP значно впливає на параметри QoS. Для підвищення якості обслуговування заявок необхідно використовувати методи балансування навантаження на сервери рівня послуг (див. п.3.3 та 3.4).

3.2 Розробка критерію оцінки якості надання послуги Triple Play

Аналіз результатів досліджень [8,9], показав що у сучасній конвергентній мережі з наданням послуги Triple Play трафік за типом вимог до QoS можна поділити на 3 класи: еластичний (Інтернет, дані); жорсткий реального часу (VoIP, сигналізація); адаптивний до затримок (IPTV, VoD).

Кожен з класів трафіка має граничні вимоги до QoS [2]. На основі даних вимог сформовано функції корисності U [8,9], які відображають ступінь задоволення абонентів в залежності від параметрів QoS, з якими надаються

послуги: пропускна здатність (B), затримка (τ), джиттер (Δt), ймовірність втрат (P). Графіки функцій U наведено на рис. 3.3.

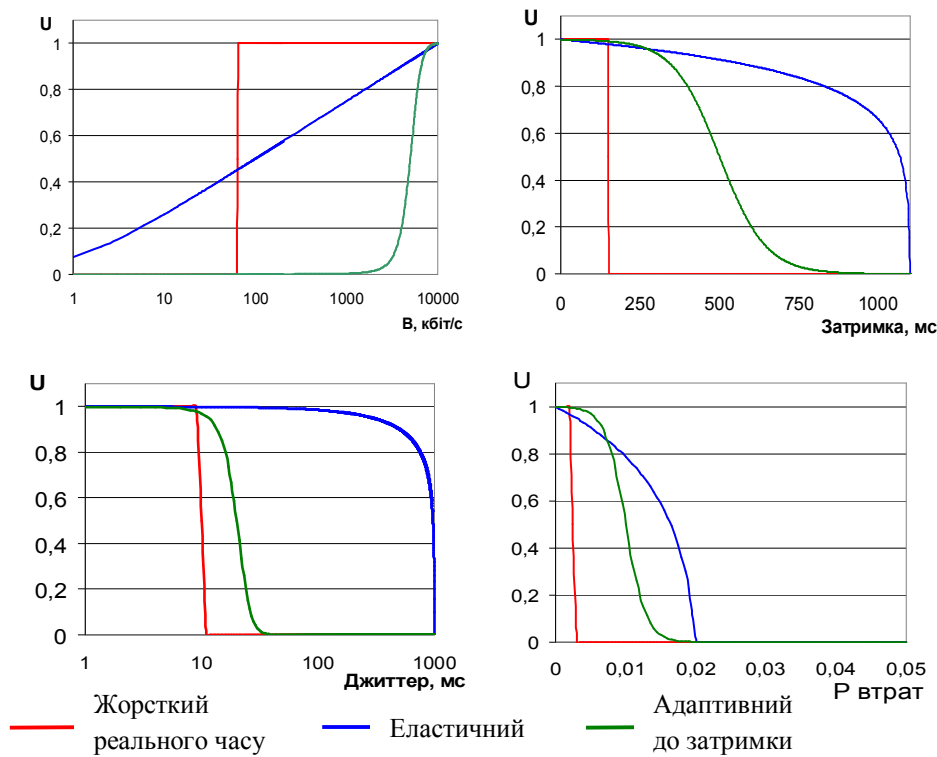


Рис. 3.3. Графіки залежності функцій корисності від параметрів QoS для різних класів трафіку

Економічним критерієм оцінки ефективності роботи конвергентної мережі є дохід оператора від надання послуг. Він значною мірою залежить від задоволеності абонентів якістю послуг. Дану залежність можливо

описати використовуючи функції корисності:

$$D = Y \sum_{i=1}^3 p_i \cdot V_i \cdot u_{B_i}(B) \cdot u_{\tau_i}(\tau) \cdot u_{\Delta t_i}(\Delta t) \cdot u_{P_i}(P) = Y \cdot U(B, \tau, \Delta t, P), \quad (3.1)$$

де D — сумарний дохід оператора за певний проміжок часу, Y — обсяг трафіка мережі; p_i — частка трафіка i -го класу; V_i — доходність трафіка i -го класу; $u_{B_i}(B), u_{\tau_i}(\tau), u_{\Delta t_i}(\Delta t), u_{P_i}(P)$ — функції корисності від пропускної здатності, затримки, джиттера, ймовірності втрат для i -го типу трафіка; $U(B, \tau, \Delta t, P)$ —

інтегральний показник якості обслуговування у мережі (інтегральна функція корисності).

Формула (3.1) ілюструє залежність показника прибутку оператора від досягнення параметрів якості обслуговування.

3.3 Розробка математичної моделі якості обслуговування на рівні керування та послуг конвергентної мережі

Надання будь-якої послуги у конвергентній мережі супроводжується послідовною обробкою інформації обладнанням рівнів NGN. Технологічні ресурси кожного рівня роблять внесок у формування затримки, джиттера, ймовірності втрати пакетів. Далі розглянуті компоненти параметрів QoS, спричинені серверним устаткуванням рівня послуг та керування.

3.3.1 Залежність параметрів QoS від завантаженості серверів

Параметри QoS на рівні послуг та керування передусім пов'язані з роботою серверів кластера. Час обробки заявки на сервері $\tau_{обробки}$ загалом є випадковою величиною, що залежить від параметрів кластерних ресурсів та потоку запитів. Аналіз літературних джерел [10] показав, що ключовий вплив на цю величину має значення поточної утилізації сервера W і час, необхідний на виконання однієї операції без урахування черг $\tau_{запиту}$ (тривалість самого запиту). Аналітична залежність для $\tau_{обробки}$ описується формулою [10]:

$$\tau_{обробки} = \tau_{запиту} + \frac{W \cdot \tau_{запиту}}{1 - W} \quad (3.2)$$

У свою чергу, $\tau_{обробки}$ впливає на такі параметри QoS як затримка, джиттер та ймовірність втрати пакетів. Час обробки заявки на сервері є безпосередньою складовою підсумкової затримки після обробки інформації на рівнях NGN. Джиттер є варіацією затримки. Втрати заявок зумовлюються перевищенням часу обслуговування припустимого значення для послуги. Дослідження доводять, що ймовірність втрат (P) в системі масового обслуговування (СМО) з обмеженим часом очікування в черзі зростає зі збільшенням величини відношення інтенсивності вхідного потоку до інтенсивності обслуговування A , яка пропорційна утилізації сервера [11]:

$$P = \frac{e^{-(V-A)\mu\tau_{крит}}}{\frac{1}{E_v(A)} + \frac{A}{V-A}(1 - e^{-(V-A)\mu\tau_{крит}})}, \quad (3.3)$$

де V - кількість каналів СМО; μ – інтенсивність обслуговування; $E_v(A)$ –

перша формула Ерланга; $\tau_{крит}$ - критичний час очікування у черзі, після якого виклик втрачається.

Отже, для покращення параметрів QoS на рівні послуг необхідним є зменшення ймовірності перевантаження серверів. Для вирішення даної задачі авторами проекту запропоновані засоби, які підвищують ефективність роботи кластерів серверів (див. п. 3.4).

3.3.2 Математична модель роботи кластера серверів

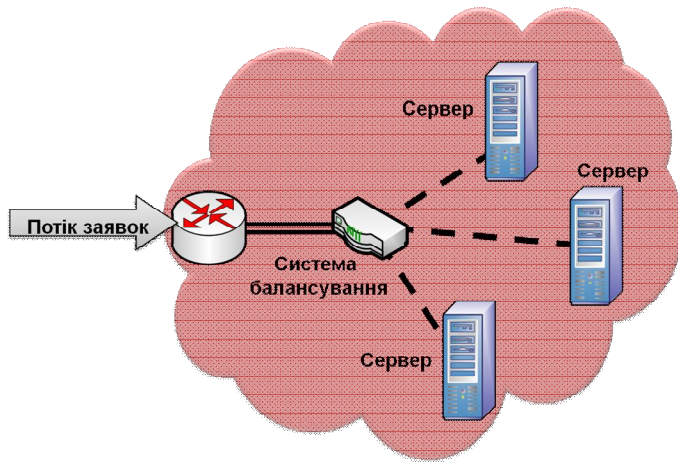


Рис. 3.4. Схема організації кластера

Серверні ресурси рівня послуг та керування організуються у кластери за узагальненою схемою, яку представлено на рисунку 3.4.

Кожен запит на обслуговування характеризується навантаженням, яке він створює на сервер (ємність серверних ресурсів, необхідних для обслуговування цієї заявки) та тривалістю обслуговування (час,

що витрачає сервер на обслуговування заявки). Величина навантаження від потоку заявок характеризується імовірнісним законом розподілу. Сервери кластера виконують обробку заявок та характеризуються ємністю та продуктивністю ресурсів. Система балансування виконує динамічний розподіл потоку заявок, що надходить, по серверах.

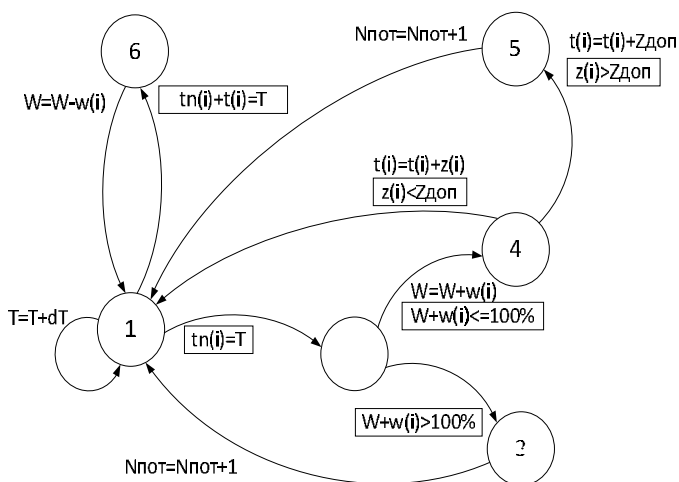


Рис. 3.5. Граф станів сервера

Математична модель роботи сервера представлена графом станів (рис. 3.5). Модель сервера має такі стани: 1 — початковий стан; 2 — надходження нового запиту; 3 — запит не отримав обслуговування через перевантаження процесора; 4 — запит надіслано на обробку; 5 — запит відкинуто через перевищення припустимого часу обробки; 6 — завершення роботи з запитом.

3.4 Синтез структурно-алгоритмічних рішень для підвищення якості надання послуги Triple Play

Погіршення якості обслуговування заявок у кластері може бути викликано ситуацією, коли ресурсів одного з серверів недостатньо для обслуговування усіх запитів, що спрямовано до нього; натомість ресурси іншого сервера можуть бути незадіяними. Це призводить до збільшення затримки та відмов на перевантажених серверах.

Існують такі способи підвищення якості обслуговування у кластері:

- 1) збільшення обсягу та продуктивності серверних ресурсів, що потребує від оператора значних фінансових витрат;
- 2) підвищення ефективності роботи наявних серверних ресурсів, що дозволяє досягти результату без значних фінансових витрат.

Для підвищення ефективності роботи наявних серверних ресурсів використовують засоби статичного та динамічного балансування навантаження. Так як максимальна якість обслуговування при умові суттєвої утилізації ресурсів досягається, коли навантаження розподіляється між існуючими серверами рівномірно, то метою засобів балансування є досягнення якомога меншого розбалансу утилізації серверів кластера (рівномірне завантаження).

3.4.1 Статичне балансування навантаження на сервери

Методика статичного балансування навантаження полягає у розміщенні на серверах одиниць контенту з урахуванням їх популярності. Статичне балансування доцільно застосовувати для контент-серверів послуг VoD та файл-серверів. Кожна одиниця контенту на цих серверах користується певною популярністю у абонентів, а тому характеризується частотою звернень до неї. Досліджено, що частота звернень до одиниці контенту є сталою величиною у відносно великому проміжку часу. Це зумовлює залежність частоти звернення, а, отже, ймовірності звернення до кожного з серверів від популярності контенту, розміщеного на ньому.

Розміщення контенту на серверах кластера без урахування попиту може призвести до різного завантаження серверів. Згідно з принципом статичного балансування, контент розміщується таким чином, щоб ймовірність звернення до кожного з серверів була однаковою.

Якщо у компанії-оператора є вільні економічні ресурси, то доцільно застосовувати механізм реплікації одного елементу контенту на декількох серверах. Це дозволить підвищити відмовостійкість системи.

3.4.2 Динамічне балансування навантаження на сервери

Системи динамічного балансування застосовують для контент-серверів та серверів сигналізації, коли сервери кластера виконують однакові функції. У такій ситуації постає задача оптимального розподілу заявок, що надійшли, між серверами.

Робота системи балансування повністю визначається закладеним алгоритмом. Найбільш відомі алгоритми балансування здійснюють вибір сервера для спрямування чергової заявки на основі

- порядкового номера чергової заявки та сервера (наприклад, алгоритм Round Robin);
- інформації про завантаженість серверів (наприклад, алгоритм Least Connection, Least Load).

Можливість приймати рішення на основі інформації про навантаження від потоку заявок може значно підвищити ефективність балансування. Але у зв'язку із наявністю багатьох факторів обчислити навантаження від кожної з заявок має можливість лише сервер, тому система балансування не має змоги безпосередньо врахувати властивості навантаження потоку заявок.

Проте у мережі NGN з наданням послуг Triple Play характерна наявність прихованих залежностей між навантаженням, що створюють серії заявок. Це зумовлене властивостями циклічності коливання та самоподібності трафіку. Завдяки цьому можливим є прогнозування навантаження для наступних m заявок на основі реальних даних про попередні p заявок. У нашому проекті пропонується використовувати нейронну мережу для прогнозування навантаження потоків заявок та використання цієї інформації алгоритмом балансування.

3.4.3 Використання нейромережі для прогнозування навантаження

Для задачі прогнозування навантаження використовується фокусована рекурентна нейронна мережа [12], побудована на основі багатошарового перцептронну (рис. 3.6). Вона є універсальною структурою, що дозволяє виконувати операцію відображення «вхід-вихід». Як функцію активації обрано логістичну сигмоїдальну функцію, використання якої робить систему ВІВО-стійкою. Можливість такого прогнозування обґрунтовано теоремою Такенса про вкладні затримки [12]. Використання нейромережі передбачає проведення заходів щодо її попереднього навчання.

Багатокрокове прогнозування відбувається за концепцією ітераційного

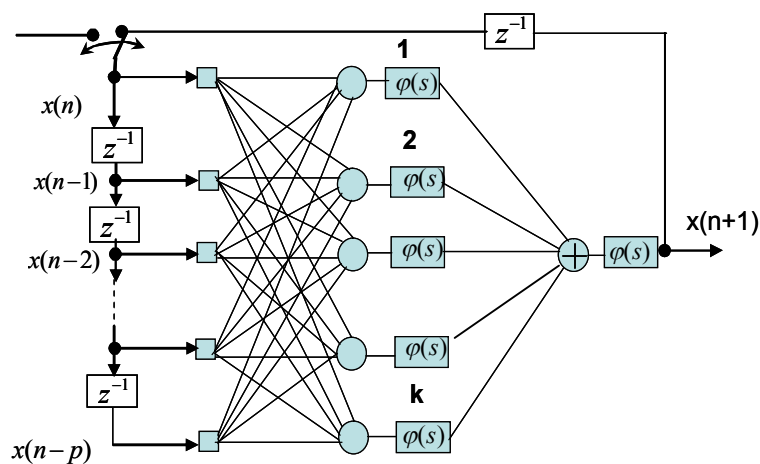


Рис. 3.6. Структура нейронної мережі

прогнозування [12]. Після того, як вхідний шар було ініціалізовано значеннями навантаження попередніх сесій, що зберігаються у пам'яті лінійної затримки порядку p , обчислюється значення вихідного нейрона, що відображає прогнозне навантаження наступної $(n+1)$ сесії:

$$x^*(n+1) = f(x(n), x(n-1), x(n-2), \dots, x(n-p)).$$

З точки зору алгоритму прямого розповсюдження сигнал на виході нейронної мережі можна формалізувати таким чином:

$$x^*(n+1) = \varphi\left(\sum_{j=1}^k \left[\omega_j \varphi\left(\sum_{l=0}^p \omega_j(l)x(n-l) + b_j\right) + b_0\right]\right).$$

Це значення приймається за дійсне та додається до вхідного вектора на наступному кроці для того щоб обчислити прогнозне значення $x(n+2)$.

Процедура прогнозу повторюється ітераційно до досягнення горизонту. Похибка прогнозування має властивість зростати з кожною ітерацією, тому горизонт прогнозу обирається за критерієм помилки на кроці m або за середньою помилкою на всіх кроках.

Слід зауважити, що вибір порядку мережі p , є компромісним між обчислювальною складністю (складність та тривалість процесів функціонування та тренування) нейронної мережі і якості прогнозу. Для систем багатокрокового прогнозування бажано, щоб виконувалася умова $p > 2m$, де m — кількість кроків прогнозування.

Дана математична модель застосована у структурі системи балансування навантаження, яку описано далі.

3.4.4 Розробка структури та алгоритму роботи системи динамічного балансування навантаження на сервери

Система динамічного балансування складається з двох підсистем та елементу керування (рис. 3.7). Підсистема 1 є основною та реалізує прогностичне балансування з використанням нейронної мережі, друга є допоміжною та реалізує балансувальник з динамічним зважуванням

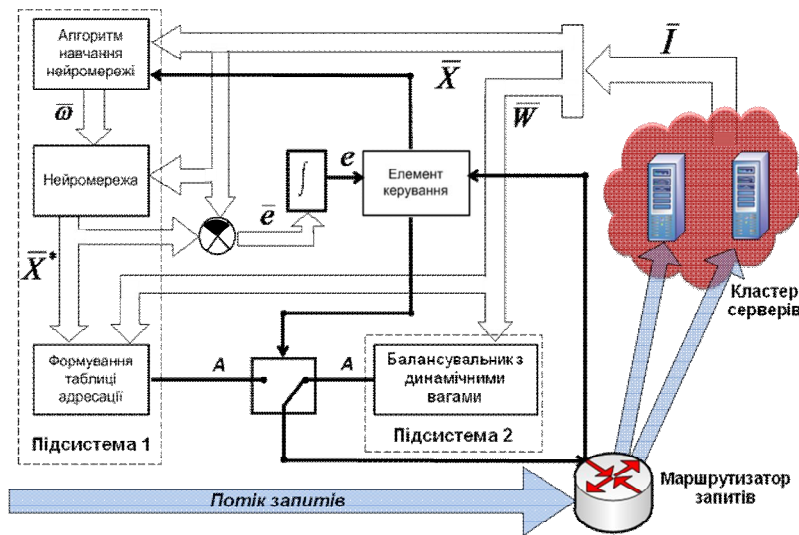


Рис. 3.7. Схема системи балансування

$\bar{X} = \{x_n, x_{n-1}, \dots, x_{n-p}\}$. Результатами роботи даних підсистем є адреса сервера (A), на який буде надіслано наступний запит.

На початковому етапі для балансування навантаження використовується модифікований Round Robin з динамічним зважуванням (підсистема 2). Паралельно проводиться навчання нейромережі. Результатом навчання є вектор вагових коефіцієнтів синаптичних зв'язків $\bar{\omega}$, що передається нейронній мережі. Після закінчення навчання проводиться оцінка достовірності прогнозу. Якщо помилка e не перевищує граничного значення e_p , керування передається підсистемі з прогностичним балансуванням навантаження.

На відміну від стандартного зваженого Round Robin, вагові коефіцієнти якого є статичними та задаються ззовні на основі даних про різну обчислювальну потужність серверів, модифікований зважений Round Robin реалізований в підсистемі 2 здійснює періодичну зміну ваг. Ваги розраховуються в залежності від поточного завантаження серверів:

$$K_i = \frac{k_i}{\min_{j=0..S}(k_j)}, \quad (3.4)$$

де k_i — частка вільних ресурсів i -го сервера, K_i — значення вагового коефіцієнта i -го сервера. перераховує вагові коефіцієнти з періодом T в залежності від поточного навантаження на сервери.

Алгоритм прогностичного балансування, що використовується в підсистемі 1, базується на m -кроковому прогнозі навантаження сесій

(модифікований Round Robin).

Інформація $\bar{I} = \{\bar{W}, \bar{X}\}$ про поточний стан кластера надходить до системи з інтервалом T . Стан серверів описується векторами утилізації $\bar{W} = \{w_1, \dots, w_S\}$ та характеристик запитів

$\bar{X}^* = \{x_{n+1}^*, x_{n+2}^*, \dots, x_{n+m}^*\}$, який формується нейронною мережею. Ця інформація використовується для складення таблиці адресації за допомогою методів алгоритму Least Load, а саме ведеться пошук на кожному кроці мінімально завантаженого сервера з урахуванням прогнозованого навантаження сесій, що надходять. У випадку, коли горизонт прогнозу вичерпано, тобто таблиця розподілення закінчилася, керування передається підсистемі 2.

Система балансування постійно веде моніторинг середньої помилки прогнозування e , та при досягненні нею граничного значення приймає рішення про передачу керування підсистемі 2 та перенавчання нейронної мережі. На підставі описаного підходу було складено комплексний алгоритм балансування навантаження на сервери.

3.5 Імітаційне моделювання системи балансування навантаження

Для імітаційного моделювання роботи системи балансування на базі запропонованих алгоритмів було розроблено відповідне програмне забезпечення у пакеті CodeGear RAD Studio 2007 мовою C++.

Моделювалися системи балансування з використанням поширених існуючих алгоритмів Round Robin (RR), Exclude, Weights, Cycles, а також з використанням запропонованого алгоритму балансування з прогнозуванням навантаження за допомогою нейромережі (NN).

Для кожного з алгоритмів оцінювалися параметри: середня затримка обробки пакету, СКВ затримки, а також кількість втрачених пакетів. Ці

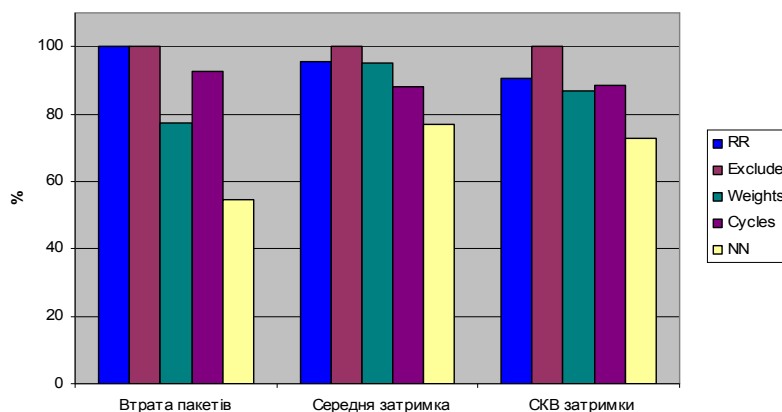


Рис. 3.8. Результати імітаційного моделювання

величини безпосередньо впливають на такі параметри QoS як затримка, джиттер та вірогідність втрат. Порівняння результатів роботи різних алгоритмів балансування показано на рис.3.8. Тут за 100% взяті найгірші з точки зору

якості обслуговування значення параметрів, що отримано в ході моделювання.

Запропонований в даній роботі алгоритм балансування з використанням нейронної мережі для прогнозування навантаження показує найкращі

результати серед усіх алгоритмів, що моделювались. Порівняно зі стандартним алгоритмом балансування Round Robin затримка і джиттер знижуються на 22%, а ймовірність втрати пакетів майже вдвічі.

Використовуючи критерій оцінки ефективності описаний в п. 3.2 та результати імітаційного моделювання встановлено, що покращення якості обслуговування на рівні послуг завдяки використанню розроблених структурно-алгоритмічних рішень дозволяє оператору зв'язку збільшити дохід на 16%.

4 ПОТРЕБИ ФІНАНСОВИХ ВИТРАТ

- Придбання та встановлення обладнання для впровадження концепції IMS.
- Придбання та встановлення обладнання для впровадження концепції SDP.
- Придбання вузлів RAN-gateway для впровадження фемтостільників.
- Інтеграція системи динамічного балансування в мережу.
- Придбання відповідних ліцензій.

РЕЗУЛЬТАТИ ВИКОНАННЯ ПРОЕКТУ

В результаті виконання проекту запропоновані науково-технічні рішення щодо формування та ефективного використання технологічних ресурсів різних рівнів конвергентної мережі з метою надання високоякісних послуг Triple Play мобільним оператором зв'язку.

Аналіз архітектури NGN та принципів її функціонування дозволив виявити проблемні ділянки на різних рівнях мережі при наданні послуги Triple Play. На рівні доступу основною проблемою є недостатня пропускна здатність каналів. Її можна вирішити шляхом застосування технологій фіксованого доступу технології (GPON) та радіодоступу (UMTS/HSPA+MIMO, LTE, WiMAX). Розширення покриття радіомереж та підвищення якості їх роботи в приміщеннях може бути здійснено завдяки встановленню фемтостільників. На рівні керування запропоновано використовувати архітектуру SO-IMS із застосуванням системи балансування навантаження серверів P-CSCF та S-CSCF. Це значно підвищує стабільність сесій різноманітних послуг. На рівні послуг запропоновано впровадження платформи SDP, що дозволяє здійснювати легке керування життєвим циклом послуг, надаючи можливість оператору оперативно реагувати на зміни у попиті користувачів.

Для оцінки ефективності роботи конвергентної мережі було запропоновано інтегральний критерій, який базується на функціях корисності послуг Triple Play для абонента в залежності від параметрів QoS мережі.

Проведено аналіз впливу завантаженості серверів рівня послуг та керування на параметри QoS на основі розробленої математичної моделі кластера серверів. Удосконалено методи динамічного балансування навантаження на сервери рівня послуг та керування, завдяки їх адаптації до характеристик потоку запитів та впровадження алгоритму прогностичного керування з використанням нейромережі.

На основі запропонованих математичних моделей та алгоритмів була розроблена система динамічного балансування навантаження на сервери рівня послуг та керування. Проведено імітаційне моделювання роботи даної системи за допомогою програмної моделі та апаратного макету. Аналіз результатів моделювання показав, що при використанні даної системи затримка і джиттер на рівні послуг та керування знижуються на 22%, а втрати пакетів на 47% порівняно зі стандартною системою балансування.

Комплексне використання результатів даного проекту дозволить мобільному оператору підвищити якість надання послуг Triple Play та збільшити показник економічної ефективності більш ніж на 16%.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ITU-T Rec. Y.2001 “General overview of NGN”.
2. ITU-T Rec. Y.1541 “Network performance objectives for IP-based services”.
3. Jie Zhang, Guillaume de la Roche. Femtocells: technologies and deployment. Wiley. - 2010.
4. Poikselka, Miikka. The IMS: IP Multimedia concepts and services. Wiley, 2009.
5. C. Makaya, A. Dutta. Service Continuity Support in Self-Organizing IMS Networks. 2010.
6. P.Tirana, D. Medhi, Distributed approach to S-CSCF selection in an IMS Network. Proceedings of IEEE/IFIP NOMS 2010, Osaka, Japan, April 2010.
7. Barać. Service Creation Environment. Ericsson Nikola Tesla d.d., Zagreb, Croatia.
8. Utility-based Bandwidth Allocation for Triple-Play Services. Changbin Liu, Lei Shi, Bin Liu. Department of Computer Science and Technology, Tsinghua University, 2009.
9. S. Shenker, “Fundamental Design Issues for the Future Internet”, IEEE Journal on Selected Areas in Communications, vol. 13, no. 7, pp. 1176–1188, Sept. 1995.
10. Соловійов М.С., Воропаєва В.Я. Забезпечення показників якості в конвергентних мережах GSM/Wi-Fi при впровадженні нових сервісів // Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація. Випуск 16 (147). - Донецьк-2009. – 248 с., С. 22-28
11. Ершов В.А., Кузнецов Н.А. Мультисервисные телекоммуникационные сети – М. - 2003.
11. Хайкин Саймон Нейронные сети: полный курс.- М: ВИЛЬЯМС, - 2006. – 1104 с.