ВСТУП

Актуальність. У сучасних умовах стрімкого розвитку інформаційних технологій та зростання обсягів передавання даних мобільні системи зв’язку стають ключовою інфраструктурою цифрового суспільства. Від якості їх роботи безпосередньо залежить ефективність функціонування державних, комерційних і соціальних сервісів. Перехід від традиційних мереж 2G/3G до сучасних технологій 4G та 5G, а також підготовка до впровадження 6G зумовлюють значне ускладнення процесів планування й контролю, оскільки виникає потреба у гнучкому управлінні ресурсами, динамічному балансуванні навантаження та забезпеченні гарантованих показників якості обслуговування (QoS, QoE).В умовах зростання кількості мобільних пристроїв, розвитку Інтернету речей (IoT), сервісів реального часу та технологій доповненої й віртуальної реальності постає проблема ефективного використання обмежених радіочастотних і обчислювальних ресурсів. Традиційні методи планування й контролю вже не відповідають сучасним вимогам, оскільки не забезпечують достатньої швидкодії та адаптивності при роботі з динамічними потоками даних. Оптимізація процесів планування й контролю дозволяє підвищити ефективність використання ресурсів мобільної мережі, знизити рівень перевантажень, покращити якість обслуговування користувачів та забезпечити стійкість функціонування телекомунікаційної інфраструктури. Застосування сучасних методів оптимізації, зокрема математичного моделювання, евристичних алгоритмів та методів машинного навчання, відкриває нові можливості для автоматизації управління мобільними мережами та забезпечення їх масштабованості.

Таким чином, дослідження питань оптимізації планування й контролю є актуальним як з наукової, так і з практичної точки зору, оскільки сприяє підвищенню ефективності мобільних систем зв’язку, відповідає світовим тенденціям розвитку телекомунікацій та має важливе значення для подальшого впровадження інноваційних технологій зв’язку в Україні та світі.

Мета дослідження: підвищення ефективності використання ресурсів мережі, підвищення якості обслуговування користувачів та зменшення ймовірності перевантажень за допомогою розробленого методу оптимізації процесів планування й контролю в системах мобільного зв'язку на основі математичного моделювання, алгоритмів оптимізації та методів машинного навчання.

Завдання дослідження:

1. Проаналізувати сучасні підходи до планування та контролю в мобільних мережах різних поколінь та визначити їхні обмеження.
2. Дослідити існуючі методи оптимізації та обґрунтувати вибір найбільш доцільних для дослідження.
3. Розробити математичну модель задачі оптимізації процесів планування й контролю з урахуванням обмеженості ресурсів і вимог до показників QoS та QoE.
4. Створити метод оптимізації на основі математичного моделювання, алгоритмів оптимізації та методів машинного навчання.
5. Розробити програмний або імітаційний прототип для перевірки роботи запропонованої метод.
6. Провести експериментальні дослідження на основі змодельованих сценаріїв функціонування мобільних мереж, оцінити точність, швидкодію та ефективність розробленого методу.
7. Сформувати практичні рекомендації щодо впровадження результатів у системи мобільного зв'язку операторів.

Об'єкт дослідження: процеси планування та контролю функціонування систем мобільного зв'язку.

Предмет дослідження: моделі, методи та алгоритми оптимізації процесів планування й контролю в мобільних мережах з урахуванням обмежень ресурсів та вимог до якості обслуговування.

Для досягнення мети та розв'язання поставлених завдань у роботі застосовано комплекс наукових методів:

1. Аналіз і систематизація наукових джерел – для дослідження сучасного стану та тенденцій розвитку мобільних мереж, методів планування й контролю, а також підходів до їх оптимізації.

2. Методи математичного моделювання – для формалізації процесів планування та контролю в мобільних мережах, побудови моделей розподілу ресурсів і навантаження.

3. Алгоритмічні методи оптимізації (лінійне та нелінійне програмування, генетичні алгоритми, евристичні методи) – для пошуку оптимальних рішень у задачах розподілу обмежених ресурсів.

4. Методи машинного навчання (класифікація, регресія, прогнозування) – для підвищення точності оцінки навантаження та адаптивного управління мережевими ресурсами.

5. Імітаційне моделювання – для відтворення процесів функціонування мобільних мереж та перевірки ефективності запропонованого методу оптимізації.

6. Методи статистичного аналізу – для обробки результатів моделювання, оцінки ефективності та порівняння із традиційними підходами.

Наукова новизна роботи полягає у розробці методу оптимізації процесів планування й контролю в системах мобільного зв’язку, яка поєднує математичне моделювання, алгоритмічні методи оптимізації та підходи машинного навчання для забезпечення адаптивного розподілу ресурсів. Запропоновано математичну модель, що враховує обмеженість радіочастотних та обчислювальних ресурсів, а також показники якості обслуговування (QoS, QoE). Розроблено алгоритм оптимізації, який дозволяє знизити ймовірність перевантажень та підвищити ефективність використання мережевих ресурсів, а також удосконалено підхід до прогнозування навантаження в мобільних мережах на основі методів машинного навчання.

Практичне значення отриманих результатів полягає в можливості впровадження розробленої методу оптимізації процесів планування й контролю у сучасні системи мобільного зв’язку, що дозволяє підвищити ефективність використання мережевих ресурсів, зменшити ймовірність перевантажень та

забезпечити стабільну якість обслуговування користувачів. Запропоновані моделі та алгоритми можуть бути інтегровані до систем управління мобільними мережами операторів зв’язку, а також використані для подальшого розвитку технологій 5G і перспективних мереж 6G.

1 ТЕОРЕТИЧНІ ОСНОВИ ПЛАНУВАННЯ ТА КОНТРОЛЮ В СИСТЕМАХ МОБІЛЬНОГО ЗВ'ЯЗКУ

1.1 Еволюція систем мобільного зв'язку: від 2G до 5G/6G

Еволюція мобільних систем зв'язку – це послідовний перехід від суто голосових сервісів і поканальної комутації до універсальної, програмно-керованої платформи для даних, відео, масового IoT і критично важливих застосунків. Стартовою точкою стала 2G (GSM), що в 1990-х уперше оцифрувала радіоінтерфейс: голос передавався з кращою якістю та захистом, з'явилися SMS і міжнародний роумінг. Дані були додатковими й повільними: спочатку десятки кбіт/с, з GPRS – приблизно до сотні, з EDGE – до кількох сотень кбіт/с у піку. Планування радіоресурсу ґрунтувалося на статичних частотних планах і суворій координації сот, а контроль якості зводився до підтримання прийнятного рівня завад і показників доступності голосу.

Перехід до 3G на початку 2000-х (UMTS/WCDMA, а також CDMA2000) відкрив еру мобільного Інтернету. Радіодоступ на широкій смузі з кодовим розділенням дозволив підвищити спектральну ефективність і швидкості передачі даних: від сотень кбіт/с до одиниць і десятків Мбіт/с з HSDPA/HSUPA. Мережі стали динамічнішими: з'явилися складніші алгоритми керування потужністю, м'які хендовери, адаптивна модуляція й кодування. Планування вже враховувало неоднорідні навантаження даних, а контроль якості охоплював не тільки доступність голосу, а й затримку та пропускну здатність для пакетних сервісів.

Справжню «перепрошивку» архітектури принесла 4G LTE у 2010-х. Радіо перейшло на OFDMA в даунлінку та SC-FDMA в аплінку, повсюдно застосовано MIMO та агрегацію несучих, а ядро перетворилося на all-IP EPC. Очікувані користувацькі швидкості сягнули десятків і сотень Мбіт/с, знизилися затримки, з'явився VoLTE для голосу поверх IP. Планування ресурсів стало багатовимірним: частота, час, простір (бімформінг), багатосекторність і багатошаровість (макро- та малі стільники) об'єдналися з політиками QoS. Контроль мережі перейшов до

телеметрії в реальному часі, SON-функцій (self-optimizing networks) і аналітики, що дозволило автоматично вирівнювати навантаження, калібрувати параметри сусідства та мінімізувати завади.

Покоління 5G (NR) з кінця 2010-х додало гнучку нумерологію, роботу як у традиційних діапазонах sub-6 GHz, так і у мм-хвилях, масивний MIMO з формуванням вузьких променів і сервісно-орієнтоване ядро 5GC. 5G вперше офіційно розділило сценарії на eMBB (масовий широкосмуговий доступ), URLLC (ультранадійні з наднизькою затримкою) та mMTC (масовий машинний трафік). У керуванні мережею з'явилися network slicing для ізольованих «логічних мереж» під різні SLA, вбудована хмарність і крайові обчислення (MEC) для близькості сервісів до користувача, SDN/NFV для програмної конфігурованості. Для планування це означає одночасний облік спектра, щільної малосотової топології, променів антенних решіток, політик зрізів і розміщення функцій на краю. Контроль якості став аналітично керованим: телеметрія потоку, предиктивні моделі навантаження та автоматизоване виявлення аномалій дозволяють превентивно змінювати конфігурацію та маршрутизацію трафіка.

Паралельно зростала роль безпеки та ідентифікації: від базових алгоритмів шифрування й процедур аутентифікації 2G/3G до розширених AKA-механізмів LTE та захисту ідентифікаторів у 5G. Це критично для мережевих зрізів, промислового IoT і приватних мереж підприємств, де втрата керування чи QoS напряму впливає на виробництво й безпеку.

Горизонт 6G – це не просто ще більші швидкості, а «AI-native» мережа, що інтегрує зв'язок, обчислення й сенсинг. Дослідження спрямовані на суб-ТГц і терагерцові діапазони, cell-free massive MIMO, реконфігуровні інтелектуальні поверхні, інтеграцію наземних і не наземних сегментів (NTN із супутниками та БПЛА), а також на об'єднання зв'язку з радіолокаційними функціями для позиціювання й сприйняття оточення. Очікується нативне використання машинного навчання в RAN і ядрі, семантичні підходи до передавання даних, жорсткі енергетичні бюджети та «вбудована» кіберстійкість. Планування в такій мережі стане контекстно-адаптивним і передбачувальним: ресурси, промені, зрізи,

обчислення на краю та в хмарі розподілятимуться в реальному часі під цільові метрики застосунків, а контроль спиратиметься на безперервне навчання моделей і спільний аналіз радіо-, транспортного та сервісного рівнів.

Таблиця 1.1

Порівняння характеристик поколінь мобільного зв’язку

Покоління	Період впровадження	Основні технології доступу	Швидкість передачі даних	Затримка (латентність)	Основні сервіси	Частотний діапазон
2G	1990-ті	GSM, CDMA	до 64 Кбіт/с	~500 мс	Голос, SMS	до 1 ГГц
3G	2000-ні	UMTS, CDMA2000	до 2 Мбіт/с	100–200 мс	Інтернет, мультимедіа	1–2 ГГц
4G	2010-ті	LTE, OFDMA	100 Мбіт/с – 1 Гбіт/с	20–50 мс	Відео, VoIP, хмарні сервіси	2–8 ГГц
5G	2020-ті	Massive MIMO, mmWave	до 10 Гбіт/с	1–10 мс	IoT, AR/VR, автономні авто	<6 ГГц, 24–100 ГГц
6G	очікується з 2030-х	AI-native, THz-комунікації	>1 Тбіт/с	<1 мс	Голографічна комунікація, повсюдний IoT	100 ГГц – 1 ТГц

Еволюція систем мобільного зв’язку є яскравим прикладом стрімкого розвитку телекомунікаційних технологій, які змінюють не лише способи комунікації, а й створюють нові соціально-економічні умови. Початок було покладено з 2G, яка запровадила цифрову передачу голосу та SMS, що дало поштовх до масового поширення мобільних послуг. Перехід до 3G означав впровадження мобільного інтернету з підтримкою мультимедійних сервісів, що суттєво вплинуло на розвиток мобільних додатків і стало основою для формування інформаційного суспільства. 4G принесла високошвидкісний доступ до інтернету,

що дозволило розвивати потокове відео, хмарні сервіси, мобільні ігри та інтеграцію мобільного зв'язку з бізнес-процесами. Найсучасніший стандарт 5G не лише забезпечує в рази більшу швидкість і наднизьку затримку, але й відкриває можливості для масового підключення пристроїв Інтернету речей, підтримки автономного транспорту, розумних міст і критично важливих додатків. Нині активно тривають дослідження та розробки щодо 6G, яка розглядається як інтеграційна платформа для штучного інтелекту, голографічної комунікації, повсюдного обчислювального середовища та мереж з пропускною здатністю у терабітному діапазоні.

Таблиця 1.2

Виклики та проблеми планування і контролю

Покоління	Основні виклики у плануванні мереж	Проблеми контролю
2G	Покриття сільських територій, розміщення базових станцій	Обмеження ємності, якість голосу
3G	Потреба у вищій щільності БС, інтеграція з 2G	Контроль завантаженості каналів, QoS
4G	Планування високошвидкісного мобільного інтернету, робота з широкими спектрами	Управління трафіком, затримки при зростанні навантаження
5G	Щільність мережі (small cells), використання mmWave, енергоефективність	Забезпечення наднизької затримки, контроль масових підключень IoT
6G	Планування у терагерцовому діапазоні, глобальна інтеграція з супутниковими системами	Кібербезпека, адаптивний контроль на основі ШІ, масштабованість

Важливо відзначити, що еволюція від 2G до 6G – це не лише вдосконалення швидкості чи ємності, а й зрушення у концептуальному баченні мобільних мереж: від забезпечення базових потреб користувачів до створення інтелектуального

середовища, яке здатне забезпечувати безперервну інтеграцію фізичного, цифрового та віртуального світів.

Таблиця 1.3

Виклики та обмеження кожного покоління у плануванні й контролі мережі

Покоління	Виклики планування	Обмеження контролю
2G	Оптимізація розміщення базових станцій для покриття великих територій; обмежений спектр частот	Контроль якості голосових дзвінків та навантаження каналів; відсутність механізмів управління даними
3G	Планування щільнішої мережі для підтримки мобільного інтернету; інтеграція з 2G	Обмежені можливості управління QoS; труднощі з балансуванням навантаження між голосовими та мультимедійними сервісами
4G	Планування високошвидкісного доступу в умовах зростання кількості користувачів; необхідність роботи в різних діапазонах частот	Контроль затримок та пропускної здатності при масовому використанні відео та хмарних сервісів; підвищені вимоги до моніторингу мережі
5G	Планування щільних мереж з малими комірками; забезпечення покриття у mmWave-діапазоні; енергоефективність	Контроль наднизької затримки (<10 мс), управління масовими підключеннями IoT; забезпечення стабільної роботи критично важливих сервісів
6G	Планування мереж у терагерцовому діапазоні з глобальною інтеграцією (наземні + супутникові системи); адаптивне використання ШІ	Обмеження у масштабованості, потреба в автоматизованому контролі на основі ШІ; високі вимоги до кіберзахисту та стійкості мережі

Еволюція технологій модуляції, методів доступу та архітектурних рішень у мобільних системах від 2G до 6G є ключовим чинником зростання їхньої пропускної здатності, зниження затримок і підвищення ефективності використання радіочастотного спектра. У системах 2G застосовувалися прості цифрові методи модуляції, такі як GMSK у GSM, що дозволяло забезпечити якісну передачу голосу при відносно низькій швидкості даних. Доступ до каналу здійснювався на основі TDMA або CDMA, що стало першим кроком до більш гнучкого управління

ресурсами. Перехід до 3G був ознаменований впровадженням технологій WCDMA та CDMA2000, які використовували розширення спектра і дозволяли обслуговувати більшу кількість користувачів, а також забезпечувати передачу мультимедійних даних. В архітектурному плані з’явилися більш складні ядра мережі, здатні інтегрувати голос і дані.

Подальший розвиток у 4G відзначався переходом до OFDMA (ортогональний частотний поділ каналів), що забезпечило значне підвищення швидкості й спектральної ефективності, а також впровадження технологій MIMO, які дозволили збільшити пропускну здатність за рахунок використання багатьох антен. Архітектура стала більш IP-орієнтованою, що забезпечило інтеграцію з інтернет-сервісами і хмарними технологіями. У 5G ці принципи отримали подальший розвиток завдяки використанню massive MIMO, mmWave-комунікацій, beamforming і network slicing. Це дозволило створювати віртуалізовані сегменти мережі для різних типів сервісів – від IoT до автономного транспорту. Архітектура набула модульного характеру з активним використанням хмарних і програмно-визначуваних мереж (SDN, NFV).

У майбутньому 6G очікується інтеграція штучного інтелекту в управління модуляцією та доступом, застосування адаптивних схем модуляції в терагерцовому діапазоні, голографічного MIMO та навіть комунікацій із залученням квантових принципів. Методи доступу можуть базуватися на принципах NOMA (нерівноправного множинного доступу) з динамічним розподілом ресурсів, що дозволить обслуговувати мільярди підключених пристроїв одночасно. Архітектурні рішення спрямовуються на глобальну інтеграцію наземних, супутникових і повітряних мереж у єдине інтелектуальне середовище з самокерованими системами управління.

Таблиця 1.4

Еволюція модуляції, методів доступу та архітектурних рішень у мобільних системах

Покоління	Методи модуляції	Методи доступу	Архітектурні рішення
2G	GMSK (GSM), QPSK (CDMA)	TDMA, CDMA	Централізована архітектура, орієнтація на голосові послуги, обмежена передача даних
3G	QPSK, 16-QAM (для HSPA)	WCDMA, CDMA2000	Інтеграція голосу й даних, багаторівнева архітектура з розвиненим ядром мережі
4G	QPSK, 16-QAM, 64-QAM	OFDMA (downlink), SC-FDMA (uplink)	IP-орієнтована архітектура, підтримка MIMO, повна інтеграція з інтернетом і хмарними сервісами
5G	QPSK, 16-QAM, 64-QAM, 256-QAM	OFDMA, Massive MIMO, beamforming, network slicing	Хмарно-орієнтована, модульна архітектура; використання SDN/NFV; підтримка IoT та критично важливих сервісів
6G	Адаптивні схеми модуляції (вище 256-QAM), THz-модуляції, голографічне MIMO	NOMA, динамічний доступ, AI-driven управління	Глобальна інтеграція наземних, супутникових і повітряних мереж; інтелектуальні самокеровані системи; тісна інтеграція з ШІ

Таблиця 1.4 демонструє, що розвиток модуляції й методів доступу прямо впливає на архітектуру, яка поступово переходить від жорстко централізованих структур до інтелектуальних, адаптивних і гнучких мереж із самокеруванням.

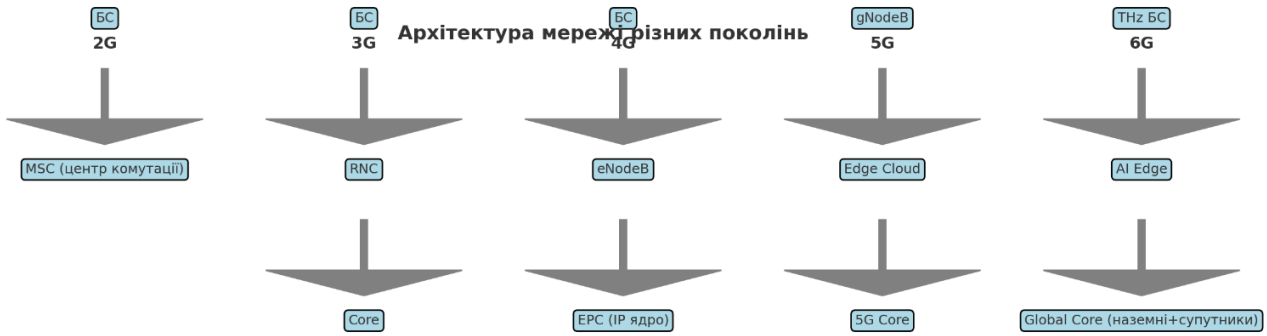


Рисунок1.1 Архітектура мереж різних поколінь

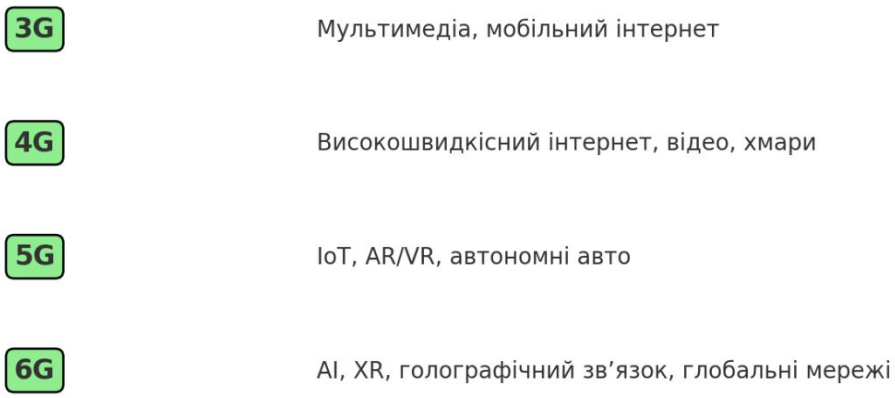


Рисунок 1.2 Схема застосування мобільних мереж

1.2 Основні принципи планування мереж мобільного зв’язку

Планування мереж мобільного зв’язку є багаторівневим процесом, який поєднує технічні, організаційні та економічні аспекти. Його головна мета полягає у створенні такої архітектури системи, яка здатна забезпечити високу якість обслуговування абонентів, ефективне використання обмеженого частотного ресурсу та оптимальні витрати на будівництво й експлуатацію інфраструктури. Основні принципи планування формуються на основі взаємозв’язку між характеристиками радіоінтерфейсу, особливостями розташування абонентів, прогнозованим навантаженням, вимогами до пропускної здатності та стандартами якості послуг.

Одним із ключових принципів є забезпечення повного та якісного покриття території. Планування починається з визначення зони обслуговування та аналізу географічних, кліматичних і демографічних характеристик регіону. Рельєф місцевості, щільність забудови, наявність транспортних магістралей і скупчень людей суттєво впливають на вибір місць розташування базових станцій та конфігурацію стільникових комірок. Головним завданням у цьому контексті є мінімізація «білих плям» у покритті та досягнення стабільного рівня сигналу у більшості місць перебування абонентів.

Другим принципом є забезпечення достатньої ємності мережі. Оскільки кількість користувачів та обсяг трафіку зростають, проектування має враховувати як поточне, так і прогнозоване навантаження. Це передбачає визначення необхідної кількості каналів, оптимізацію смуги пропускання та розрахунок кількості базових станцій для уникнення перевантаження. При цьому широко застосовуються математичні моделі та методи теорії масового обслуговування для оцінки ймовірності блокування викликів, затримок у передачі даних та інших показників якості сервісу.

Важливим принципом виступає раціональне використання частотного спектра, який є обмеженим і дорогим ресурсом. Планування базових станцій здійснюється з урахуванням повторного використання частот і мінімізації інтерференції між сусідніми комірками. У сучасних системах LTE та 5G застосовуються додаткові механізми – динамічний розподіл частот, використання когнітивного радіо, а також алгоритми самоналаштування (SON), що підвищують ефективність управління спектром.

Не менш важливим є принцип забезпечення якості обслуговування (QoS, QoE). Під час планування враховуються параметри рівня сигналу, ймовірність з’єднання, швидкість передачі даних, затримки та надійність з’єднання. Для цього проводяться симуляції покриття та навантаження, радіообстеження, а також використовується балансування трафіку між комірками.

Сучасні підходи до планування також включають принцип багатошаровості мережі. Мережа складається з макростільників, які забезпечують базове покриття, та дрібніших комірок (мікро-, піко-, фемто-), що підвищують ємність у місцях високої концентрації абонентів. Це дозволяє досягти безшовного переходу користувачів між різними рівнями мережі та підвищити ефективність її роботи.

Таблиця 1.5

Основні принципи планування мобільних мереж

Принцип	Сутність	Очікуваний результат
Забезпечення покриття	Аналіз території, рельєфу, забудови та абонентів для оптимального розташування базових станцій	Мінімізація зон без сигналу, стабільний рівень покриття
Забезпечення ємності	Розрахунок кількості каналів і смуги пропускання для уникнення перевантажень	Обслуговування прогнозованої кількості абонентів без втрати якості
Ефективне використання частот	Планування повторного використання частот, мінімізація інтерференції	Раціональне використання спектра, висока якість сигналу
Якість обслуговування	Використання показників QoS та QoE, балансування навантаження	Високий рівень задоволеності абонентів
Масштабованість і гнучкість	Можливість розвитку мережі та впровадження нових технологій	Підтримка переходу до нових поколінь зв’язку
Економічна доцільність	Оптимізація витрат на будівництво та експлуатацію	Зменшення собівартості послуг та підвищення прибутковості

Окрему роль відіграє принцип енергоефективності. Зважаючи на значні витрати енергії в системах мобільного зв’язку, проектування мереж здійснюється з урахуванням оптимізації потужності передавачів, застосування алгоритмів динамічного керування трафіком та використання «зелених» технологій, що відповідають концепції сталого розвитку.

Принцип надійності та стійкості забезпечує безперервність роботи мережі навіть у випадках часткових відмов обладнання. Це досягається резервуванням каналів і базових станцій, використанням дублюючих елементів інфраструктури та алгоритмів автоматичного відновлення після збоїв.

Необхідно також підкреслити принцип відповідності міжнародним стандартам. Планування мобільних мереж має відповідати вимогам таких організацій, як ITU, 3GPP, ETSI, що гарантує сумісність обладнання різних

виробників і забезпечує можливість інтеграції локальних систем у глобальну телекомунікаційну інфраструктуру.

Узагальнені характеристики основних принципів планування наведені в таблиці 1.5.

З таблиці видно, що всі принципи взаємопов’язані та спрямовані на досягнення балансу між якістю обслуговування і витратами на інфраструктуру.

Таблиця 1.6

Вплив середовища на планування мобільних мереж

Фактор середовища	Вплив на планування	Приклади заходів
Рельєф місцевості	Затінення сигналу горами, пагорбами, лісами	Використання більшої кількості базових станцій, ретрансляторів
Урбанізація	Високі будівлі створюють «тіні» та відбиття сигналу	Використання мікро-, піко- та фемтостільників
Щільність населення	Концентрація великої кількості абонентів у певних зонах	Збільшення ємності комірок, балансування навантаження
Транспортні магістралі	Потреба у стабільному зв’язку під час руху	Оптимізація розташування базових станцій уздовж доріг
Кліматичні умови	Атмосферні явища можуть погіршувати сигнал	Резервування потужності та використання надійних антенних систем

Суттєвий вплив на процес планування здійснюють фактори зовнішнього середовища, які часто визначають специфіку розташування базових станцій та архітектуру мережі загалом. Це показано у таблиці 1.6.

Як видно, планування не може бути універсальним: для різних територій необхідні різні підходи та технічні рішення.

Завершуючи, варто відзначити, що принципи планування значно змінюються залежно від покоління мобільних систем. Якщо для GSM першочерговим завданням було забезпечення голосового зв’язку та максимального покриття, то в епоху LTE і 5G на перший план виходять вимоги до швидкості передачі даних, низької затримки, підтримки масових IoT-пристроїв та масштабованості мережі. Це порівняння наведене в таблиці 1.7.

Таблиця 1.7

Порівняння принципів планування у різних поколіннях мобільного зв’язку

Покоління	Основні вимоги до планування	Особливості
2G (GSM)	Покриття території, голосові послуги	Частотне планування з високим коефіцієнтом повторного використання
3G (UMTS)	Підвищення ємності, підтримка передачі даних	Планування з урахуванням інтерференції та зростання трафіку
4G (LTE)	Висока пропускна здатність, QoS для мультимедіа	Використання OFDMA, гнучкий розподіл частот
5G	Масовий IoT, ультранизькі затримки, масштабованість	Використання малих комірок, beamforming, мережеві зрізи (network slicing)

Як свідчать дані, кожне покоління мобільного зв’язку вимагає нових підходів до планування, однак незмінним залишається базовий комплекс принципів – покриття, ємність, ефективність спектра, якість, гнучкість та економічність.

На рисунку 1.3 представлено графік, який демонструє еволюцію мобільних систем від 2G до 5G. Якщо в мережах другого покоління швидкість передачі даних становила лише сотні кбіт/с, то в LTE вона досягла сотень Мбіт/с, а в 5G – гігабітного рівня. Логарифмічна шкала підкреслює стрімке зростання продуктивності, що зумовлює необхідність нових принципів планування.

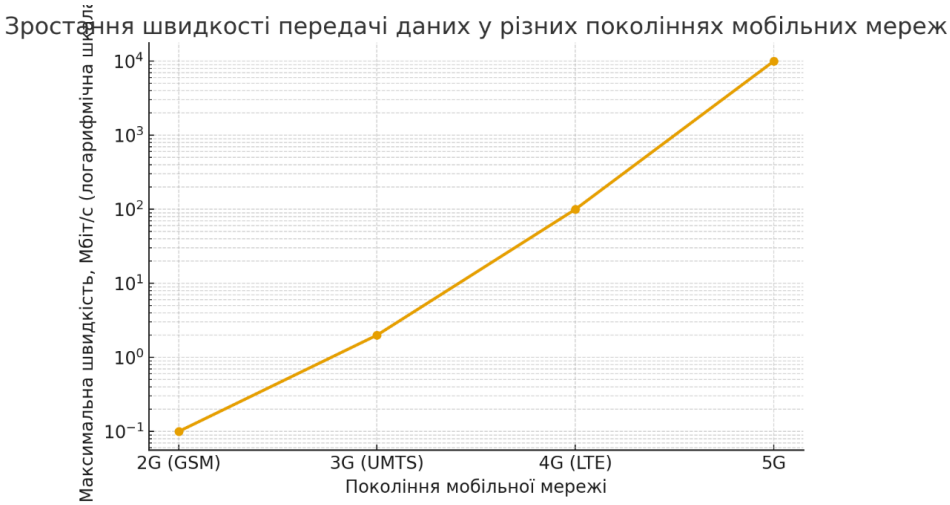


Рисунок 1.3 Зростання швидкості передачі даних у різних покоління мобільних мереж

На рисунку 1.4 представлено графік, який ілюструє залежність навантаження на мережу від кількості користувачів у зоні покриття. Зростання кількості абонентів призводить до нелінійного збільшення потреби в ємності, що вимагає застосування методів оптимізації, балансування трафіку та використання дрібних стільників.

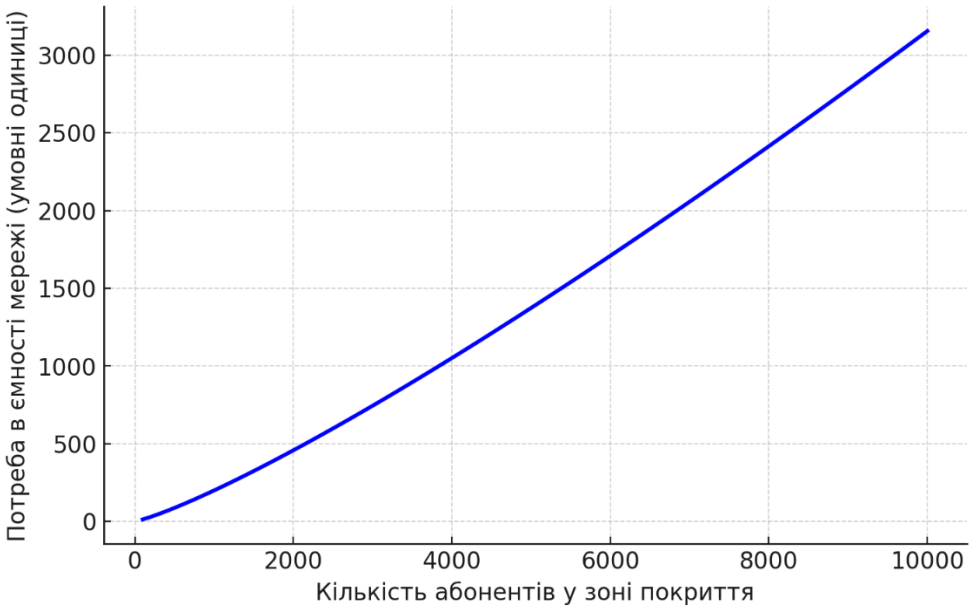


Рисунок 1.4 Взаємозв’язок між кількістю абонентів та потребою в ємності мережі

Отже, основні принципи планування мобільних мереж формуються на стику технічних, організаційних та економічних вимог. Їх дотримання дозволяє створювати гнучкі та ефективні мережі, здатні адаптуватися до динамічних умов телекомунікаційного ринку і відповідати потребам сучасного суспільства.

1.3 Механізми контролю та управління якістю обслуговування (QoS, QoE)

Якість обслуговування у мобільних мережах визначається не лише технічними параметрами радіоінтерфейсу, а й тим, як користувач суб’єктивно сприймає сервіс. Тому керування здійснюється у двох взаємопов’язаних площинах: QoS (об’єктивні мережеві гарантії затримки, пропускної здатності, втрат,

пріоритезації) та QoE (суб’єктивний досвід користувача, що вимірюється на рівні сервісу: відео, голос, ігри, веб). Ефективна система контролю об’єднує вимірювання, аналіз і замкнений контур впливу – від мілісекундних рішень у радіодоступі до політик у ядрі й оптимізації в транспортній мережі.

Базовою інституційною основою для керування QoS є параметризовані класи обслуговування, що визначають бюджет затримки, цільову імовірність втрат і пріоритет доступу. У 4G (LTE) це реалізовано через індекс QCI, у 5G – через 5QI; для трафіка виділяють гарантовані та негарантовані бітрейти, правила прийому в мережу, пріоритети доступу та атрибути передавання. На прикладному рівні це відображається у створенні мережевих носіїв/потоків для конкретної послуги (голос по IMS, відео, дані з найкращим зусиллям), що має власні політики допуску й обмеження. У 5G додатково застосовуються мережеві зрізи (network slicing), які виділяють логічні мережі з власними SLO для eMBB, URLLC чи mMTC, а також механізми reflective QoS і політики PCF/UPF для гнучкого маркування та форсування.

Контроль якості на радіорівні базується на адаптації до умов каналу. Планувальники MAC застосовують алгоритми розподілу ресурсів (round-robin, proportional fair, M-LWDF та ін.), що балансують між пропускною здатністю та справедливістю, а адаптація модуляції та кодування (AMC) разом із HARQ забезпечують необхідну надійність кадрів за мінімально можливої затримки. DRX/DTX режим знижує енергоспоживання терміналів без значної втрати QoS у фонових сервісах. Управління мобільністю (пороги тригерів, гістерезис, таймери) і оптимізація стійкості до інтерференції (ICIC/eICIC/feICIC, координація потужності, beamforming, MU-MIMO, масивні антени у 5G) прямо впливають на KPI з’єднань на межі стільників, і це є критичним для зниження показників обривів та коливань швидкості. У багатошарових мережах важливими залишаються балансування навантаження між шарами macro/micro/pico/femto, оптимізація offload на Wi-Fi і механізми SON (ANR, MRO, MLB, CCO), що автоматизують налаштування на основі телеметрії.

На рівні ядра й політик контроль реалізує PCC-архітектура: вузол політик (PCRF/PCF) застосовує правила до абонентських сесій, а елементи примусового виконання (PCEF/UPF) здійснюють формацію та шейпінг трафіку, маркування, пріоритезацію, лімітування чи прискорення, у т.ч. залежно від плану обслуговування, сервісу, часу доби чи стану мережі. Для синхронізації з білінгом використовується онлайн-й офлайн-таргетинг; це дозволяє гарантувати SLA для корпоративних клієнтів і чесний розподіл ресурсу для масового сегмента. В транспортній мережі застосовують DiffServ/DSCP, класові черги (WFQ/CBWFQ/PQ), керування чергами й активне керування буферами (RED/CoDel/PIE), а також інженерію трафіку MPLS/Segment Routing і резервування ресурсів. Така багатоступенева ієрархія QoS забезпечує енд-ту-енд характеристики – від інтерфейсу радіодоступу до міжоператорських стиків.

Моніторинг і контроль QoE спираються на перетворення KPI у KQI для конкретних сервісів. До мережевих KPI належать RSRP/RSRQ/SINR, успішність встановлення й утримання з'єднань (CSSR, DCR), швидкість, затримка, джиттер, втрати, ефективність хендоверу. Їх агрегують у KQI: MOS для голосу, частота буферизацій, час старту, тривалість зависань і якість зображення (PSNR/SSIM/VMAF) для відео, TTFB/PLT для веб, латентність/стабільність для ігор. Для збору даних використовують пасивні зонди у ключових точках мережі, активні вимірювання (синтетичні транзакції, пінг/трасування, HTTP-/DNS-тести), термінальні телеметрії (MDT, crowdsourcing) та DPI для ідентифікації сервісів. Отримані ряди даних обробляються системами аналітики з виявленням аномалій, кореляцією інцидентів і локалізацією вузьких місць; дедалі частіше – із застосуванням ML для прогнозу деградацій і рекомендацій дій.

Замкнений контур керування реалізується на різних часових масштабах. На рівні мілісекунд відпрацьовують планування, HARQ, AMC; на рівні секунд-хвилин спрацьовують перерозподіл ресурсів між носіями, параметри мобільності, обмеження швидкостей, динамічні політики PCF; на рівні годин-днів – перемаршрутизація транспортного трафіку, зміни профілів QoS, ретюнінг радіопараметрів, а також оптимізація контентної доставки (кеші/CDN, MEC). В 5G

важливу роль відіграє обчислення на межі (MEC), яке зменшує кінець-у-кінець затримки для URLLC-сценаріїв та покращує QoE відео-, ігрових і AR/VR-сервісів; інтеграція з ABR-плеєрами дає змогу динамічно підбирати бітрейт відповідно до миттєвої пропускної здатності каналу.

Забезпечення якості неможливе без чітко визначених SLA/SLO та процедур Service Assurance. Оператор формує каталоги послуг з цільовими KQI, порогами тривог і правилами ескалації, після чого будує дашборди реального часу та звітність для технічних і бізнес-команд. На етапі введення в експлуатацію проводяться польові випробування та приймальні тести (drive-/walk-тести, N-часові навантажувальні випробування, A/B-перевірки QoE), а у режимі експлуатації – безперервна оптимізація за результатами спостережень. Питання безпеки та регуляторики (наприклад, вимоги нейтральності мережі та захисту персональних даних) інтегруються в політики QoS, аби уникнути дискримінації трафіку чи небажаного DPI у чутливих сценаріях.

Окремим напрямом є керування перевантаженнями та стабілізація TCP. Щоб уникнути «bufferbloat», у вузлах увімкнено AQM; у радіо- та транспортних сегментах налаштовують вікна TCP, PEP-механізми, сучасні контролери (наприклад, BBR у дата-центрових стиках). Для масових подій застосовують тимчасові політики throttling/boost, динамічне виділення додаткових носіїв і транспортних маршрутів, а також переносні базові станції (COW/COL).

Інтелектуалізація керування відбувається через SON і AIOps. Системи автоматично коригують параметри сусідства, пороги хендоверу, потужність і tilt антен, перерозподіляють користувачів між шарами, прогнозують «гарячі точки», планують включення/вимкнення елементів для енергозбереження. У 5G це доповнюється policy-driven slicing: для ургентних сервісів встановлюються жорсткі бюджети затримки та високий пріоритет ARP, тоді як для best-effort трафіку – м’які вимоги і гнучкі черги.

У практичному вимірі QoS/QoE – це постійний компроміс між ресурсами й очікуваннями користувача. Надмірна пріоритезація одиничних потоків може погіршити спільний добробут, тоді як суто середні метрики приховують

індивідуальні деградації. Тому сучасні підходи комбінують перцентильне оцінювання (P95/P99), контекст (тип пристрою, радіоумови, рухливість) і причинно-наслідковий аналіз, а рішення приймаються на основі end-to-end телеметрії з радіо, IP-транспорту, ядра та прикладного рівня.

Таблиця 1.8

Параметри QoS у мобільних мережах

Параметр	Позначення	Опис	Вплив на якість
Пропускна здатність	Throughput	Максимальна швидкість передачі даних для користувача	Визначає можливість використання ресурсоємних сервісів (відео, ігри)
Затримка	Latency	Час доставки пакета від джерела до отримувача	Критично важливо для реального часу (голос, відеоконференції, IoT)
Джиттер	Jitter	Варіації затримки між пакетами	Впливає на якість потокового відео та VoIP
Імовірність втрат пакетів	PLR	Частка втрачених пакетів у загальному потоці	Знижує якість передачі мультимедіа, особливо для голосу та відео
Надійність	Reliability	Ймовірність успішної доставки повідомлення	Особливо важливо для критичних сервісів (URLLC)

Таблиця 1.9

Основні класи QoS у LTE/5G (QCI / 5QI)

QCI / 5QI	Тип трафіку	Затримка (мс)	Втрати пакетів (%)	Приклади сервісів
1 / 80	Гарантований бітрейт (GBR)	≤ 100	$\leq 10^{-2}$	VoIP, відеодзвінки
2 / 82	GBR	≤ 150	$\leq 10^{-3}$	Інтерактивне відео
5 / 85	Non-GBR	≤ 100	$\leq 10^{-6}$	IMS signaling
7 / 7	Non-GBR	≤ 300	$\leq 10^{-3}$	Потокове відео
9 / 9	Non-GBR	≤ 300	$\leq 10^{-6}$	Інтернет-браузинг, best-effort
URLLC (5QI=85/86)	GBR	≤ 1	$\leq 10^{-5}$	Дистанційне керування, IoT критичного рівня

Підсумовуючи, механізми контролю та управління якістю обслуговування у мобільних мережах – це багаторівнева система, яка поєднує стандартизовані класи QoS, алгоритми розподілу радіоресурсів, політики в ядрі та транспорті, проактивний моніторинг і перетворення KPI у KQI для конкретних сервісів. У поєднанні з автоматизацією SON/AIOps, MEC, network slicing і гнучкими політиками PCF така система дозволяє одночасно дотримуватися технічних цілей (затримка, надійність, пропускна здатність) і забезпечувати високу суб'єктивну якість досвіду для кінцевого користувача.

Таблиця 1.10

KPI та KQI для оцінки QoS і QoE

Показники	Опис	Приклади вимірювань
KPI (Key Performance Indicators)	Технічні показники роботи мережі	RSRP, RSRQ, SINR, CSSR, DCR, Throughput, Latency, Jitter
KQI (Key Quality Indicators)	Показники досвіду користувача (QoE)	MOS для голосу, час старту відео, частота буферизацій, середня тривалість зависань, PLT (Page Load Time)

Таблиця 1.11

Методи забезпечення QoS на різних рівнях мережі

Рівень	Механізми	Приклади
Радіомережа	Планування ресурсів, адаптація модуляції та кодування (AMC), HARQ, ICIC/eICIC, beamforming	Розподіл ресурсів між користувачами
Ядро мережі	Політики PCC, пріоритезація трафіку, формування черг, шейпінг	PCRF/PCF керування
Транспортна мережа	DiffServ/DSCP, MPLS, інженерія трафіку, AQM (RED, CoDel)	QoS у магістральних каналах
Сервісний рівень	Оптимізація контенту, кешування, адаптивна потокова передача (ABR)	QoE для відео та веб

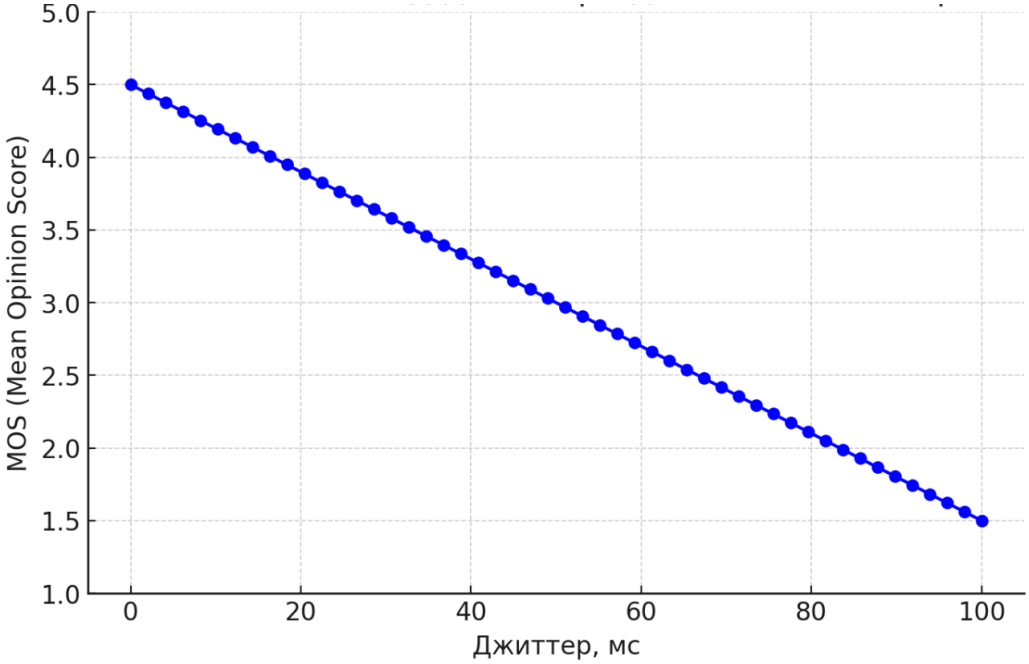


Рисунок 1.5 Залежність MOS від джиттера для голосових сервісів

На графіку зображено, як збільшення джиттера (варіацій затримки) негативно впливає на якість голосового зв'язку, що вимірюється за шкалою MOS (Mean Opinion Score). При низьких значеннях джиттера (до 20 мс) якість оцінюється як «відмінна» (MOS ≈ 4,5), проте вже при джиттері понад 50 мс вона істотно знижується, а при 100 мс наближається до незадовільного рівня. Це підкреслює критичну важливість мінімізації коливань затримки для забезпечення прийнятної якості VoIP та відеодзвінків.

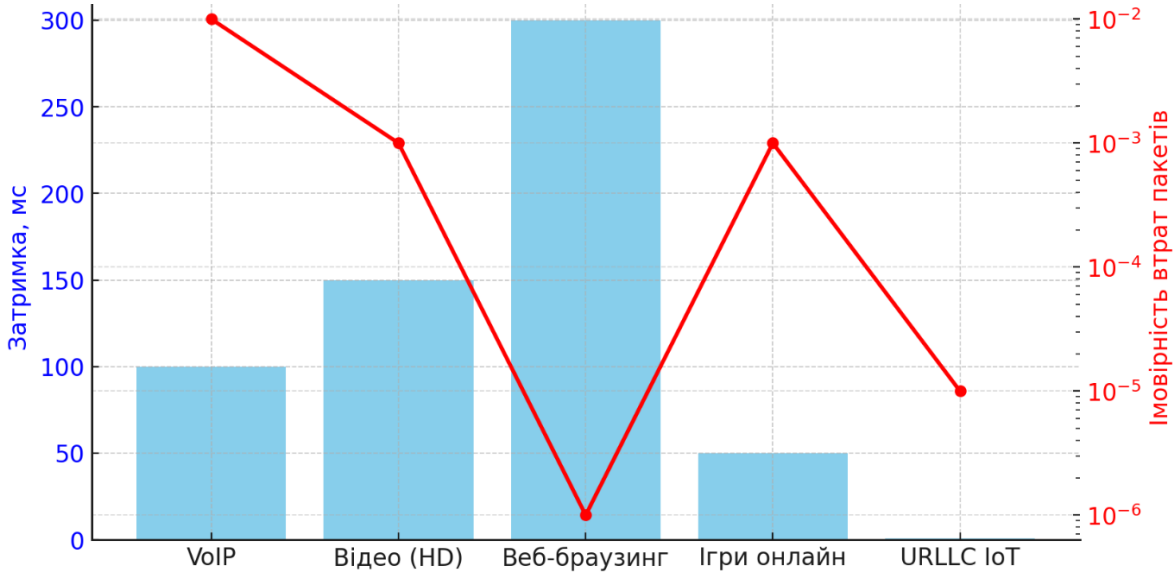


Рисунок 1.6. QoS-вимоги для різних типів сервісів

Графік порівнює ключові параметри якості обслуговування – затримку та ймовірність втрат пакетів – для п’яти основних категорій сервісів. Для голосових викликів (VoIP) допустимою є затримка не більше 100 мс при втраті до 1%. Відео високої чіткості (HD) допускає затримку до 150 мс, проте чутливе до втрат понад 0,1%. Веб-браузинг має відносно м’які вимоги (затримка до 300 мс, втрати 10^{-6}), тоді як онлайн-ігри вимагають низької затримки (до 50 мс). Найсуворіші умови спостерігаються для критичних сервісів URLLC у 5G, де затримка має бути менше 1 мс, а втрати пакетів – менше 10^{-5}.

У таблиці 1.12 наведено співвідношення між ключовими показниками продуктивності мережі (KPI) та показниками якості досвіду користувача (QoE). Як видно, технічні параметри, такі як RSRP та SINR, визначають базові характеристики радіоканалу і безпосередньо впливають на стабільність з’єднання та швидкість передачі даних.

Показники вищих рівнів – пропускна здатність, затримка, джиттер і втрати пакетів – відображають ефективність функціонування транспортної та мережевої інфраструктури, формуючи користувацьке сприйняття відео- та голосових сервісів. Водночас індикатори на кшталт DCR чи PLT відображають безпосередній вплив на кінцевий досвід абонента, оскільки вони пов’язані з надійністю викликів та зручністю доступу до веб-ресурсів.

Таблиця 1.12

Відповідність KPI та показників QoE

KPI (Key Performance Indicator)	Опис показника	Типове значення	Вплив на QoE (Key Quality Indicator)
RSRP (Reference Signal Received Power)	Сила прийнятого сигналу	від –140 до –80 дБм	Визначає стабільність з’єднання; слабкий сигнал → низька швидкість, зростання ймовірності обривів
SINR (Signal-to-Interference plus Noise Ratio)	Відношення сигнал/шум + завади	від –10 до +30 дБ	Високий SINR → краща якість відео та голосу; низький SINR → артефакти, спотворення
Throughput (Пропускна здатність)	Швидкість передачі даних користувача	1–1000 Мбіт/с	Впливає на якість потокового відео, швидкість завантаження сторінок, ігор; низька швидкість → буферизації

Latency (Затримка)	Час доставки пакета	1–300 мс	Критично важливо для ігор, VoIP та IoT; велика затримка → поганий MOS, лаги в іграх
Jitter (Джиттер)	Коливання затримки	< 20 мс (VoIP)	Високий джиттер → переривання голосу, погіршення відео
PLR (Packet Loss Ratio)	Втрати пакетів	0.001–1%	Високі втрати → «фрізи» у відео, спотворення звуку
DCR (Dropped Call Rate)	Частка перерваних викликів	< 1%	Визначає надійність голосового зв'язку; високий DCR → низький рівень задоволеності абонентів
Page Load Time (PLT)	Час завантаження веб-сторінки	< 2–3 с	Прямо впливає на зручність веб-серфінгу; великі затримки → низьке QoE
MOS (Mean Opinion Score)	Суб'єктивна оцінка якості голосу (1–5)	3.5–4.5 бажано	Інтегральний показник сприйняття якості абонентом

Таким чином, QoE є інтегральним відображенням сукупності KPI, і саме тому оператори повинні розглядати управління якістю не лише як технічне завдання, а й як ключовий бізнес-фактор, що визначає лояльність і задоволеність користувачів.

1.4 Огляд сучасних методів оптимізації в телекомунікаційних системах

Оптимізація телекомунікаційних систем є одним із ключових завдань сучасної інженерії зв'язку, оскільки зростання кількості користувачів, різноманіття сервісів і підвищені вимоги до якості обслуговування потребують ефективного використання обмежених ресурсів. До основних ресурсів мобільних мереж відносяться радіочастотний спектр, пропускна здатність каналів, потужність передавачів і обчислювальні можливості мережевих вузлів. Оптимізація передбачає пошук такого розподілу ресурсів і конфігурації мережі, які забезпечать максимальну якість обслуговування абонентів при мінімальних витратах для оператора.

Серед сучасних методів оптимізації важливе місце займають алгоритми динамічного розподілу радіоресурсів (Radio Resource Management, RRM). Вони

забезпечують баланс між ефективністю використання спектра та якістю зв’язку для кожного абонента. Використання адаптивної модуляції та кодування (AMC), багатопотокових технологій (MIMO) і beamforming дозволяє значно підвищити спектральну ефективність. Оптимізація в цьому випадку полягає у виборі таких параметрів передачі, які враховують поточний стан радіоканалу та вимоги сервісу.

Важливим напрямом є оптимізація мережевої інфраструктури через планування розташування базових станцій та використання гетерогенних мереж (HetNet), що поєднують макростільники, мікростільники, піко- та фемтосоти. Такий підхід дає змогу підвищити ємність мережі й покращити покриття в зонах з високою щільністю користувачів. Оптимізаційні моделі визначають оптимальну кількість та місце розташування малих стільників з урахуванням економічних і технічних факторів.

У сучасних мобільних системах значна увага приділяється оптимізації маршрутизації та управління трафіком. Традиційні протоколи доповнюються механізмами SDN (Software Defined Networking) і NFV (Network Function Virtualization), що дозволяють централізовано керувати потоками даних, динамічно змінювати конфігурацію мережі та забезпечувати пріоритезацію трафіку відповідно до вимог QoS. Така оптимізація знижує затримки й покращує рівень QoE, особливо для критично важливих сервісів (онлайн-ігри, відеоконференції, IoT).

В останні роки активно розвивається напрям енергоефективної оптимізації. Оскільки базові станції споживають до 70% енергії мобільної мережі, застосовуються алгоритми динамічного вимкнення неактивних вузлів, оптимізація потужності передавачів та використання «зелених технологій» (відновлюваних джерел енергії, energy harvesting). Це дозволяє знизити експлуатаційні витрати та зменшити вуглецевий слід телекомунікаційної інфраструктури.

Таблиця 1.13

Класифікація методів оптимізації в телекомунікаційних системах

Група методів	Основні підходи	Приклади застосування
Оптимізація спектра і радіоресурсів	Динамічний розподіл ресурсів (RRM), адаптивна модуляція та кодування (AMC), MIMO, beamforming	Підвищення пропускної здатності, мінімізація інтерференцій
Інфраструктурна оптимізація	Гетерогенні мережі (HetNet), планування розташування базових станцій, small cells	Покращення покриття та ємності в зонах з високим навантаженням
Оптимізація управління трафіком	SDN/NFV, маршрутизація на основі QoS, балансування навантаження	Пріоритезація критичного трафіку, зменшення затримок
Енергоефективна оптимізація	Вимкнення неактивних вузлів, оптимізація потужності, «зелені» технології	Зменшення енергоспоживання та OPEX
Інтелектуальні методи (AI/ML)	Прогнозування навантаження, reinforcement learning, anomaly detection	Автоматизація управління мережею, самоналаштування
Оптимізація на основі QoE	Моніторинг KQI, прогнозування MOS, адаптивне відео	Підвищення задоволеності користувачів

Особливу роль у сучасній оптимізації відіграє штучний інтелект і машинне навчання. Методи глибинного навчання та прогнозної аналітики застосовуються для передбачення навантаження в мережі, виявлення аномалій і прийняття рішень щодо перерозподілу ресурсів. Алгоритми reinforcement learning дозволяють мережі «навчатися» на основі історичних даних і поступово удосконалювати власні стратегії керування. У 5G і майбутніх 6G саме інтелектуальна оптимізація стане основним механізмом досягнення високої ефективності.

Не менш важливим напрямом є оптимізація на рівні користувацького досвіду (QoE-oriented optimization). Якщо раніше головним завданням було забезпечення певних технічних показників (пропускна здатність, затримка), то сьогодні акцент робиться на максимізації задоволеності абонента. Для цього використовуються системи моніторингу реального трафіку, суб’єктивні оцінки якості (MOS), а також алгоритми прогнозування якості відео- та голосових сервісів.

Таблиця 1.14

Порівняння традиційних та AI-орієнтованих методів оптимізації

Критерій	Традиційні методи	Методи з використанням AI/ML
Гнучкість	Обмежена, правила задаються вручну	Висока, алгоритми адаптуються до змінних умов
Масштабованість	Складна при великій кількості користувачів	Легко масштабується завдяки автоматизації
Прогнозування	Базується на статистичних моделях	Використовує глибинне навчання і big data
Реакція на зміни	Реактивна (усунення проблем після виникнення)	Превентивна (прогноз і запобігання деградаціям)
Приклади застосування	Планування частот, балансування навантаження	Прогноз трафіку, динамічне керування QoS, self-healing

Важливим трендом у сучасних дослідженнях є оптимізація з урахуванням мультисервісності. Оскільки мережі 5G обслуговують різнорідні сервіси (mMTC, URLLC, eMBB), оптимізаційні моделі повинні враховувати суперечливі вимоги: високу пропускну здатність, мінімальні затримки та енергоефективність одночасно. Для цього застосовуються підходи багатокритеріальної оптимізації, де рішення приймається з урахуванням компромісу між кількома показниками.

Таким чином, сучасні методи оптимізації в телекомунікаційних системах можна умовно поділити на кілька груп: оптимізація спектра і радіоресурсів, інфраструктурна оптимізація, управління трафіком і маршрутизацією, енергоефективна оптимізація та інтелектуальні методи на основі AI/ML. Їхнє поєднання дозволяє створювати гнучкі, масштабовані й ефективні мережі, здатні задовольняти зростаючі вимоги користувачів і забезпечувати стабільний розвиток мобільного зв’язку.

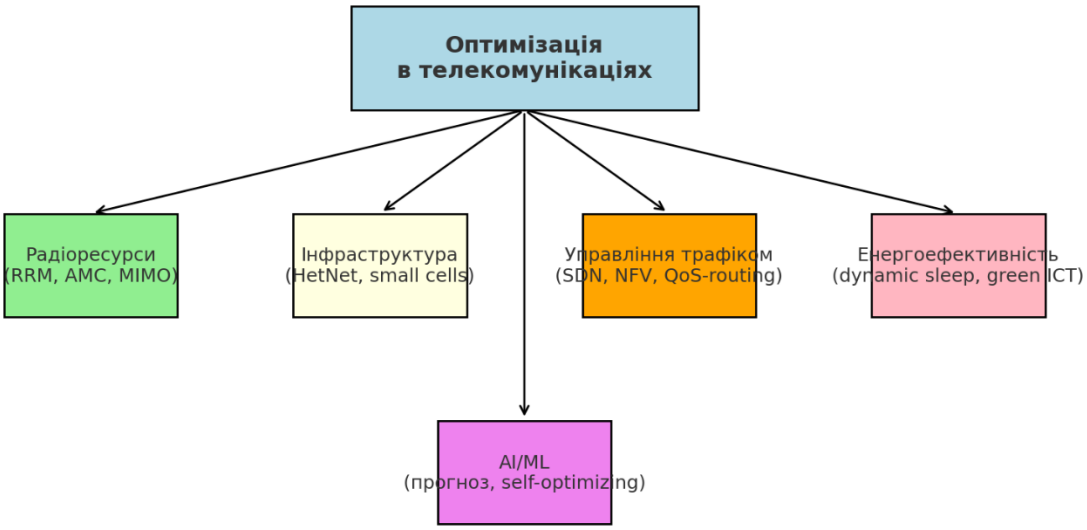


Рисунок 1.7 Основні напрямки оптимізації в телекомунікаційних системах

Рисунок 1.7 ілюструє головні напрями оптимізації в сучасних телекомунікаційних системах. Центральним елементом виступає концепція оптимізації, яка охоплює різні рівні функціонування мережі. Зокрема, оптимізація радіоресурсів (RRM, AMC, MIMO, beamforming) забезпечує ефективне використання спектра. Інфраструктурна оптимізація передбачає використання гетерогенних мереж і малих стільників для підвищення ємності. Управління трафіком реалізується через SDN/NFV та інтелектуальні алгоритми маршрутизації. Енергоефективна оптимізація спрямована на зменшення енергоспоживання шляхом динамічного вимкнення вузлів. Нарешті, інтеграція штучного інтелекту та машинного навчання забезпечує самонавчання мереж і адаптивне управління ресурсами.

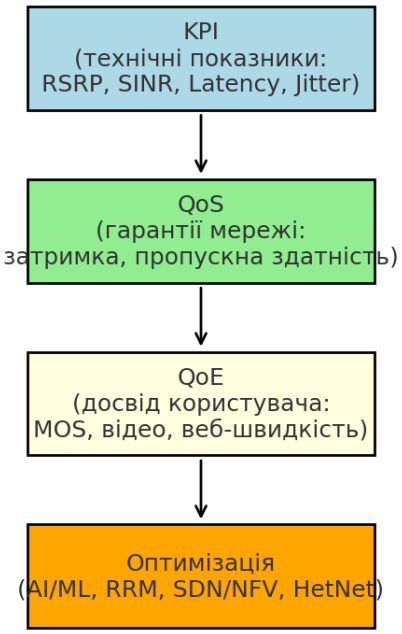


Рисунок1.8. Взаємозв’язок між KPI, QoS, QoE та оптимізацією

На рисунку 1.8 відображено логічний зв’язок між технічними показниками (KPI), параметрами якості обслуговування (QoS), якістю досвіду користувача (QoE) та процесами оптимізації. На нижньому рівні знаходяться KPI, що відображають вимірювані параметри мережі (затримка, SINR, RSRP). Ці параметри формують рівень QoS – об’єктивні гарантії якості (наприклад, забезпечення певної пропускної здатності чи максимальної затримки). QoS, у свою чергу, впливає на QoE, тобто суб’єктивне сприйняття користувачем якості сервісів (якість голосу, відео, швидкість доступу). Оптимізація виступає замикальною ланкою, яка на основі аналізу KPI та QoS коригує конфігурацію мережі для підвищення QoE.

2 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ТА МОДЕЛЕЙ ОПТИМІЗАЦІЇ ПРОЦЕСІВ ПЛАНУВАННЯ Й КОНТРОЛЮ

2.1 Методи моделювання мобільних мереж

Моделювання мобільних мереж є одним з ключових інструментів у дослідженні та оптимізації телекомунікаційних систем, оскільки дозволяє аналізувати складні процеси, прогнозувати поведінку системи за різних сценаріїв навантаження та обирати оптимальні стратегії управління ресурсами. Вибір методу моделювання залежить від цілей дослідження, рівня деталізації та наявних вхідних даних.

Одним із фундаментальних напрямів є системна динаміка, яка використовується для аналізу еволюційних процесів у мобільних мережах. Вона ґрунтується на побудові моделей у вигляді диференціальних рівнянь або причинно-наслідкових діаграм, що описують зміни параметрів мережі в часі. Типовим прикладом є рівняння стану:

$$\frac{dx(t)}{dt} = f(x(t), u(t), t), \tag{2.1}$$

де $x(t)$ – вектор стану системи (наприклад, кількість активних користувачів у мережі), $u(t)$ – вхідні параметри (трафік, ресурси), а f – функція, що описує динаміку змін. Використання системної динаміки дозволяє виявити довгострокові тенденції розвитку, взаємозв’язки між елементами системи та оцінити вплив стратегічних рішень (наприклад, введення нових технологій чи зміни політики тарифікації).

Іншим потужним інструментом є імітаційне моделювання, яке застосовується для відтворення детальних сценаріїв функціонування мережі з урахуванням випадкових факторів. Імітаційні моделі будуються на основі дискретних подій, що відбуваються у певні моменти часу:

$$T_{k+1} = T_k + \Delta t, \tag{2.2}$$

де T_k – час настання k-ої події, а Δt – випадкова величина, яка задає інтервал між подіями й визначається певним статистичним розподілом. Цей метод дозволяє досліджувати роботу мережі в умовах змінної інтенсивності трафіку, різних типів сервісів (голос, відео, IoT), а також при виникненні відмов чи перевантажень. Використання імітаційного моделювання дає змогу отримати статистику про затримки, пропускну здатність, коефіцієнт втрат пакетів, час встановлення з'єднання тощо.

Важливе місце у дослідженні мобільних мереж займають статистичні підходи, що базуються на аналізі емпіричних даних. Вони включають методи регресійного аналізу, теорії черг, марковських процесів і моделювання трафіку. Однією з найпоширеніших моделей є система масового обслуговування типу M/M/1, для якої середні характеристики можна подати у вигляді:

$$L = \frac{\rho}{1-\rho}, W = \frac{1}{\mu-\lambda}, \tag{2.3}$$

де L – середня довжина черги, W – середній час очікування, λ – інтенсивність вхідного потоку, μ – швидкість обслуговування, $\rho = \lambda/\mu$ – коефіцієнт завантаження системи. Такі моделі дозволяють оцінити продуктивність мережі, знайти «вузькі місця» та сформувати рекомендації для оптимізації.

Сучасні дослідження часто поєднують різні методи моделювання. Наприклад, статистичні моделі використовуються для генерації вхідних даних (трафіку), імітаційне моделювання – для детального відтворення роботи мережі, а системна динаміка – для прогнозування довгострокових ефектів від упровадження оптимізаційних рішень. Така інтеграція дозволяє отримати комплексне уявлення про функціонування мобільних мереж та підвищити точність результатів.

Отже, методи моделювання мобільних мереж можна умовно розділити на три основні групи: системно-динамічні, імітаційні та статистичні. Кожен із них має свої переваги й обмеження, а їхнє комбіноване використання забезпечує найбільш повне і практично значуще відображення реальних процесів у телекомунікаційних системах.

Уявімо задачу планування розгортання фрагментованої (sliced) 5G-мережі, де необхідно забезпечити якість обслуговування для різних категорій користувачів:

- мобільний широкосмуговий доступ (eMBB) – відео у високій роздільності, онлайн-ігри;
- масове підключення IoT (mMTC) – сенсорні пристрої та «розумні міста»;
- ультранадійні з'єднання з малою затримкою (URLLC) – автономний транспорт, телемедицина.

Таблиця 2.1

Порівняння методів моделювання мобільних мереж

Метод моделювання	Переваги	Недоліки	Типові сфери застосування
Системна динаміка	- Дозволяє аналізувати довгострокові тенденції - Виявляє причинно-наслідкові зв'язки - Придатна для стратегічного планування	- Низька деталізація - Не підходить для моделювання короткострокових подій	Прогнозування розвитку мереж, стратегічне планування, оцінка політики тарифікації
Імітаційне моделювання	- Висока точність і деталізація - Урахування випадкових факторів - Можливість перевірки алгоритмів перед практичним упровадженням	- Високі обчислювальні витрати - Складність побудови моделей - Трудомісткий аналіз результатів	Тестування протоколів, оптимізація алгоритмів розподілу ресурсів, аналіз QoS
Статистичні підходи	- Відносна простота реалізації - Використання реальних даних - Можливість створення аналітичних моделей	- Обмежена точність - Залежність від якості даних - Складність у врахуванні багатьох факторів	Оцінка навантаження, моделювання трафіку, теорія черг, аналіз продуктивності

Для вирішення задачі застосовується комбінована модель:

1. Системна динаміка використовується для прогнозування зростання навантаження у довгостроковій перспективі (наприклад, очікуване збільшення

кількості IoT-пристроїв у місті протягом 5 років). Це дозволяє зрозуміти, які сегменти мережі можуть стати «вузькими місцями».

2. Статистичні підходи застосовуються для моделювання характеристик трафіку, зокрема для опису розподілу пакетів у mMTC та затримок у URLLC, що дозволяє враховувати стохастичну природу навантаження.

3. Імітаційне моделювання використовується для перевірки роботи мережі у конкретних сценаріях – наприклад, під час масового заходу (концерту чи спортивного матчу), коли кількість активних користувачів зростає у кілька разів.

Таблиця 2.2

Приклади формалізацій методів моделювання мобільних мереж

Метод	Формула	Пояснення
Системна динаміка	$\frac{dx(t)}{dt} = f(x(t), u(t), t)$	Описує зміну стану системи $x(t)$у часі під впливом вхідних параметрів $u(t)$. Використовується для прогнозування довгострокових тенденцій.
Імітаційне моделювання	$T_{k+1} = T_k + \Delta t$	Дискретна подія настає у випадковий момент часу, інтервал Δtвизначається статистичним розподілом. Дозволяє відтворювати сценарії роботи мережі.
Статистичні підходи (M/M/1)	$L = \frac{\rho}{1-\rho}, W = \frac{1}{\mu - \lambda}$	Характеристики черги: середня довжина Lта середній час очікування W. Використовується для оцінки продуктивності й виявлення «вузьких місць».

Завдяки поєднанню цих методів вдається отримати комплексний прогноз:

- на стратегічному рівні – визначити оптимальні інвестиції в інфраструктуру;
- на тактичному рівні – протестувати роботу алгоритмів динамічного розподілу ресурсів;
- на операційному рівні – оцінити реальний рівень QoS/QoE для кінцевих користувачів.

Інтеграція методів моделювання створює багаторівневу аналітичну систему, яка дозволяє одночасно враховувати довгострокові тренди, стохастичні процеси та

поведінку мережі у реальних сценаріях. Це робить комбінований підхід незамінним у плануванні та оптимізації мобільних мереж наступних поколінь.

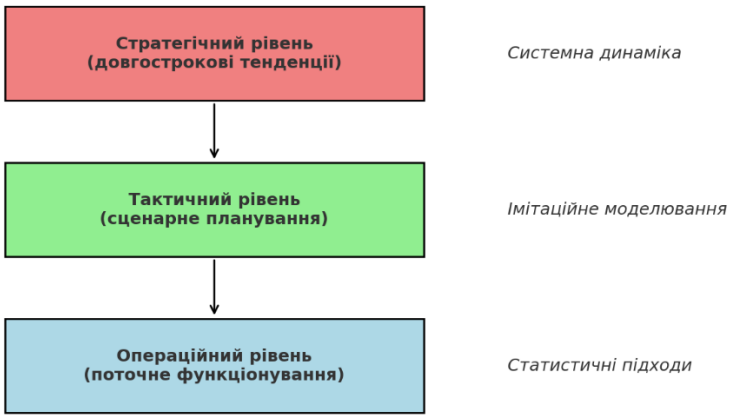


Рисунок 2.1 Рівні моделювання мобільних мереж і відповідні методи

На рисунку 2.1 наведено співвідношення між рівнями аналізу мобільних мереж та методами моделювання, які найбільш ефективно застосовуються на кожному з них. Системна динаміка орієнтована на стратегічний рівень, де важливо прогнозувати довгострокові тенденції розвитку мережі. Імітаційне моделювання використовується на тактичному рівні для відтворення сценаріїв функціонування та тестування алгоритмів управління ресурсами. Статистичні підходи доцільні на операційному рівні, де необхідна оцінка поточного навантаження, продуктивності та якості обслуговування.

2.2 Алгоритми планування ресурсів радіомережі

Планування ресурсів у радіодоступі визначає, кому, коли і на якій частоті видати радіоканал, яку модуляцію/код (MCS) застосувати, яку потужність і просторовий режим (MIMO/beamforming) обрати, аби за обмеженого спектра та змінних радіоумов забезпечити потрібну пропускну здатність і QoS/QoE. У LTE/NR ресурси подаються як часо-частотна сітка OFDMA: смуга ділиться на піднесучі з кроком $\Delta f = 15 \cdot 2^{\mu}$кГц (нумерологія μ), а базовою одиницею планування є PRB (Physical Resource Block) –12 піднесучих на інтервал (слот/мініслот). Планувальник

(scheduler) у кожному TTI приймає рішення для кожного UE за метриками каналу (CQI/PMI/RI), буфером трафіку (BSR/HoL delay), сигналами керування потужністю (PHR) і політиками сервісу (GBR/non-GBR, 5QI). Миттєва корисна швидкість оцінюється як сума по виділених PRB із відповідною спектральною ефективністю η_{MCS}, скоригованою на протоколі накладні (PDCCH/PDCCH, референс-сигнали, HARQ).

Цілі планувальника багатокритеріальні: максимізувати пропускну здатність осередку, гарантувати затримку/надійність для URLLC, зберігати справедливість між UE, уникати перепланування у граничних зонах покриття та контролювати енергоспоживання. На практиці використовують ієрархію: спершу –прийом/допуск і міжслайсове виділення (Admission/Resource Partitioning), далі –міжклітинна координація (ICIC/eICIC/FeICIC/CoMP) для зменшення міжсекторних завад, і вже потім внутрішньоклітинне рознесення PRB у часі/частоті/просторі. Для багатосмугових конфігурацій NR застосовують Carrier Aggregation та Bandwidth Parts: активні BWP перемикаються, щоб зменшити енергоспоживання UE, а крос-кар’єрне планування дозволяє керувати PDCCH на «якірному» несучому й дані на допоміжних.

Алгоритми планування умовно можна поділити на кілька класів: базові бенчмарки, пропорційно-справедливі, QoS-орієнтовані для реального часу, стохастична оптимізація, методи опуклої оптимізації, ML/DRL-підходи, а також спеціалізовані алгоритми для mMTC і масового доступу.

Основна мета кожного класу алгоритмів полягає у пошуку компромісу між пропускною здатністю, справедливістю розподілу ресурсів, затримкою, енергоефективністю та стабільністю мережі. У таблиці 2.3 наведено порівняльну характеристику основних груп алгоритмів.

Таблиця 2.3

Порівняння моделей/алгоритмів планування ресурсів у мобільних мережах

Клас підходу	Репрезентативні алгоритми	Мета/метрика	Теоретичні гарантії	Переваги	Обмеження	Типові кейси
Базові бенчмарки	Round-Robin (RR), Max-C/I	RR: справедливість; Max-C/I: пропускна здатність	Немає	Простота, базові орієнтири	Низький рівень QoS-контролю	Тестування, порівняння
Пропорційно-справедливі	PF, Weighted PF	Компроміс throughput–fairness	Доведена PF-оптимальність	Баланс продуктивності та справедливості	Не контролює затримки	eMBB, LTE/5G
QoS-орієнтовані RT	M-LWDF, EXP/PF, EDF	Мінімізація затримки, PLR	Контроль черг	Підтримка RT-трафіку	Вища складність	VoIP, відео, URLLC
Стохастична оптимізація	Lyapunov DPP	Баланс стабільності/ефективності	Доведена стабільність	Універсальність	Параметрична складність	HetNet, енерго-адаптивні системи
Опукла оптимізація	Water-filling, Lagrange dual	Максимізація спектральної ефективності	Глобальний/локальний оптимум	Висока ефективність	Вимоги до CSI	MIMO/OFDM, beamforming
ML/DRL	DQN, PPO, Multi-agent RL	Адаптивне навчання політики	Емпіричні	Самоорганізація, прогнозування	Висока обчислювальна вартість	RAN-slicing, V2X
mMTC/масовий доступ	Grant-free RA, NOMA + Lyapunov	Масштабованість, стабільність	Стабільність черг	Низькі накладні витрати	Компроміси по BER	IoT, сенсори

Класичні алгоритми формують базову лінійку. Round-Robin рівномірно роздає час усім UE –максимально проста, але ігнорує стан каналу та дає низьку спектральну ефективність. Max-C/I завжди обирає UE з найкращим миттєвим CQI –максимізує пропускну здатність, проте погіршує справедливість і QoS слабких користувачів. Пропорційно-справедливий (PF) зводить компроміс «throughput–fairness»: для кожного PRB обирає UE з найбільшим відношенням $R_{i,k}/\bar{r}_i$, де $R_{i,k}$– досяжна швидкість на PRB, $\bar{r}_i$–ковзне середнє UE; так слабші користувачі періодично «підтягуються» до середнього. Для сервісів реального часу застосовують QoS-орієнтовані метрики: M-LWDF/EXP-PF вагують користувача затримкою/чергою та класом обслуговування, надаючи пріоритет потокам із дедлайнами навіть за гіршого CQI. У NR для URLLC можливе пріоритезоване планування мініслотами з «pre-emption» –короткі URLLC-передачі можуть витісняти eMBB, після чого для eMBB тригеряться механізми відновлення (retransmission/partial retransmission). Для періодичних малих пакетів (VoIP, URLLC-контроль) використовують напівпостійне планування (SPS) або configured grant (особливо в UL), щоб мінімізувати сигналізацію і затримку доступу.

Пропорційно-справедливе (PF) планування

Оптимізаційна постановка (довга перспектива):

$$\max_{\{r_i\}} \ \sum_{i=1}^{N} \log_2(\bar{r}_i) \text{s.t.} r(t) \in \mathcal{R}(t), \tag{2.4}$$

де $\bar{r}_i$–«повільне» середнє пропускної здатності користувача i, $\mathcal{R}(t)$–множина досяжних векторів швидкостей у слоті t.

Онлайн-метрика (TTI-рівень):

$$\text{PF}_i(t) \ = \ \frac{r_i(t)}{\bar{r}_i(t)}, \bar{r}_i(t+1) = \left(1-\frac{1}{T}\right)\bar{r}_i(t) + \frac{1}{T}\, r_i(t), \tag{2.5}$$

де $r_i(t)$–миттєва (CQI/MCS-залежна) швидкість, T–стала «пам'яті» ковзного середнього.

OFDMA (RB-гранулярність):

$$\max_{\{x_{i,k}\}} \ \sum_k \sum_i \ x_{i,k} \ \frac{R_{i,k}(t)}{\bar{r}_i(t)} \text{s.t.} \ \sum_i x_{i,k} \le 1, \ x_{i,k} \in \{0,1\}, \tag{2.6}$$

де $R_{i,k}(t)$–швидкість UE iна RB k(за CQI/MCS). Правило вибору RB: признач RB kтому i, для кого $R_{i,k}/\bar{r}_i$найбільше.

Стохастична оптимізація на основі Ляпунова (drift-plus-penalty) додає керування чергами: рішення на слот максимізує $\sum_i Q_i\,\mu_i - V \cdot \text{cost}$ за обмежень потужності/PRB/інтерференції, де Q_i–черга або HoL-затримка потоку, μ_i–обслуговування (кількість бітів), V–регулятор компромісу зі споживанням енергії, штрафами SLA тощо. Цей підхід дає гарантії стабільності черг і добре працює у змішаних сценаріях (eMBB+URLLC+mMTC). Для задач розподілу потужності й просторового мультиплексування у MIMO/OFDM застосовують методи опуклої оптимізації: лагранжева двоїстість, water-filling, спільний вибір користувачів і променів із урахуванням PMI/RI та обмежень PAPR/CSI-точності. У FR2 (мм-хвилі) додається вузькопроменеве планування: кожний UE обслуговується «найкращим» бімом/панеллю; планувальник поєднує вибір бімів, часових інтервалів і MCS з урахуванням блокувань і швидких федингів; у TDD-режимах важлива відповідність DL/UL конфігурації кадру й мінімізація перехресних завад (CLI).

Смугові особливості впливають на стратегію. У низьких діапазонах (FR1 <1 ГГц) краща далекобійність і проникнення –доречно більше уваги приділяти справедливості та покриттю. У середніх (2–7 ГГц) –баланс пропускної здатності й покриття, масове застосування CA і багатопотокового MIMO; тут PF/M-LWDF із урахуванням beamforming та ICIC дає найкращий загальний ефект. У FR2 (24–71 ГГц) –великі смуги (400–800 МГц), але висока чутливість до блокувань; доцільні короткі TTI/мініслоти, щільне переобчислення бімів, швидка переалокація PRB при втраті лінії видимості, а також подвійна підключеність (DC) з «якорем» у FR1, щоб утримувати керування і безшовно падати на нижчий діапазон за деградації каналу.

В uplink планувальник спирається на BSR/PHR і обмеження UE по потужності: призначення PRB і MCS має гарантувати, що потрібна енергетика досяжна; для масових IoT (mMTC) використовують grant-free доступ, групове кодування, NOMA або реплікації, а стабільність і колізії контролюють Lyapunov-подібними правилами і адаптацією конфігурацій RACH/пілотів. У downlink

ключові аспекти –вибір просторових шарів і користувачів у MU-MIMO так, щоб кореляція каналів мінімізувала взаємні завади (user pairing/precoding), а також координація сусідніх сот через ICIC/eICIC (ресурсне «м'ютинг», майже порожні субкадри для піко-UE) або CoMP (спільна передача/динамічний пункт сервісу) у важких інтерференційних топологіях.

Стохастична оптимізація: Lyapunov Drift-Plus-Penalty (DPP)

Динаміка черг (класична модель):

$$Q_i(t+1) = \max\{Q_i(t) - \mu_i(t), 0\} + A_i(t), \qquad (2.7)$$

де $A_i(t)$–прибуття трафіку, $\mu_i(t)$–обслуговування (залежить від рішення планувальника).

Функція Ляпунова та дрейф:

$$L(\mathrm{Q}(t)) = \frac{1}{2}\sum_i Q_i^2(t), \Delta(t) = \mathbb{E}\left[L(\mathrm{Q}(t+1)) - L(\mathrm{Q}(t)) \mid \mathrm{Q}(t)\right]. \qquad (2.8)$$

Критерій DPP (стандартний вигляд):

$$\min\ \Delta(t) + V\,\mathbb{E}\left[g(t) \mid \mathrm{Q}(t)\right], \qquad (2.9)$$

де $g(t)$ – «штраф/вартість» (наприклад, енергія, порушення SLA, негативна корисність), $V > 0$–параметр компромісу «затримка/стабільність ↔ штраф».

Типове пер-слотове правило прийняття рішення (випливає з верхньої оцінки дрейфу):

$$\max_{\text{feasible}}\ \sum_i Q_i(t)\,\mu_i(t)\ -\ V\,c(t), \qquad (2.10)$$

де $c(t)$–миттєва «ціна» дії (наприклад, сумарна потужність), а «feasible» задає радіо-обмеження (PRB, потужність, інтерференція, MCS). *Сенс:* високі черги ⇒агресивніше обслуговувати відповідних UE; великий V ⇒сильніше мінімізувати витрати/штраф.

Приклад для OFDMA з енергетичною ціною:

$$\max_{\{x_{i,k}, P_{i,k}\}}\ \sum_k \sum_i Q_i(t)\,R_{i,k}(P_{i,k})\,x_{i,k}\ -\ V\sum_k \sum_i P_{i,k}$$

$$\sum_i x_{i,k} \le 1,\ x_{i,k} \in \{0,1\},\ \sum_{k,i} P_{i,k} \le P_{\max},\ R_{i,k}(P) = B\log_2(1+\gamma_{i,k}(P)). \qquad (2.11)$$

Так отримуємо «черго-зважене» енерго-ефективне планування з гарантіями стабільності.

Таблиця 2.4

Таблиця рішень для планування ресурсів у радіомережі

Умова / Тип трафіку	URLLC (критичні затримки)	eMBB (високі швидкості)	mMTC (масовий IoT)
B1. Наявність пакету	Якщо дедлайн < 1 TTI → активується pre-emption	Планується лише за наявності вільних PRB після URLLC	Планується лише якщо ресурси залишились
B2. Перевірка пулу PRB	Виділяється мініслот навіть за рахунок інших UE	Використовує основний пул ресурсів	Використовує мінімальний пул (NB-IoT/BWP)
B3. Метрика	Deadline-based (EDF, M-LWDF)	PF / Max-C/I	Random Access Priority + проста метрика
B4. Вибір UE	Найвищий пріоритет, можливе витіснення eMBB	Вибирається UE з найкращим співвідношенням метрики	Вибір груповий (batch allocation)
B5. MCS/шари	Дуже консервативний MCS, low-BLER target (≤1%)	Агресивний MCS, максимально можливий	Дуже низький MCS, 1 шар
B6. Обмеження ICIC/CA	Дозволяється мульти-сайт CoMP для надійності	Використовується CA для піку швидкості	Може вимикатися частина несучих для енергозбереження
B7. DCI формування	DCI з мінімальною затримкою, пріоритетний CCE	Стандартний DCI, агрегація можливих UE	Спрощений DCI, об’єднання в групові grants
B8. Зворотний зв’язок	Негайний HARQ, швидка переадаптація	ACK/NACK з використанням HARQ Chase/IR	Можливі групові ACK для зменшення сигналізації

З погляду реалізації, сучасні планувальники змішують кілька метрик: у «базі» – PF або його зважений варіант для eMBB; поверх –клас-специфічні ваги за 5QI (GBR/Non-GBR), HoL-затримкою, пріоритетами користувача/слайсу; для URLLC –мініслоти та пріемпція; для mMTC –конфігурований беззапитний UL і контроль колізій. Для міжслайсового поділу спектра застосовують статичні квоти,

«вирізання» PRB-вікон (RRM-ізоляція) або динамічне спільне використання з верхньорівневими політиками (наприклад, ваги/кепи на рівні слайсу та admission control на основі token/leaky-bucket). Оцінювання ефективності ведуть за набором метрик: пропускна здатність на UE/сотку, затримка й джиттер, PLR/BLER, справедливість (індекс Джайна), енергоефективність (біт/Дж), а також стабільність черг і виконання SLA для слайсів.

Узагальнюючи, алгоритми планування радіоресурсів –це багаторівнева комбінація правил на часо-частотній сітці, просторі й потужності з урахуванням сервісних вимог і інтерференційного оточення. У типовій конфігурації для FR1/FR2 використовують гібрид: PF/weighted-PF як «хребет» для eMBB, QoS-орієнтовані метрики (M-LWDF/EXP-PF/EDF) для потоків із дедлайнами, Lyapunov-орієнтований контроль черг для змішаних навантажень та енергообмежень, а також спеціалізовані механізми URLLC (мініслоти, pre-emption) і mMTC (grant-free, колізійний контроль). Вибір параметрів (вікно $\bar{r}$, ваги QoS, Vу DPP, правила пріоритетів, розклад TDD, політика CA/BWP) визначає конкретний компроміс між пропускною здатністю, затримкою, справедливістю й енергоспоживанням у даній смузі частот і топології мережі.

Отже, методи моделювання мобільних мереж можна умовно розділити на три основні групи: системно-динамічні, імітаційні та статистичні. Кожен з них має свої переваги та обмеження, а їхнє комбіноване використання забезпечує найбільш повне і практично значуще відображення реальних процесів у телекомунікаційних системах.

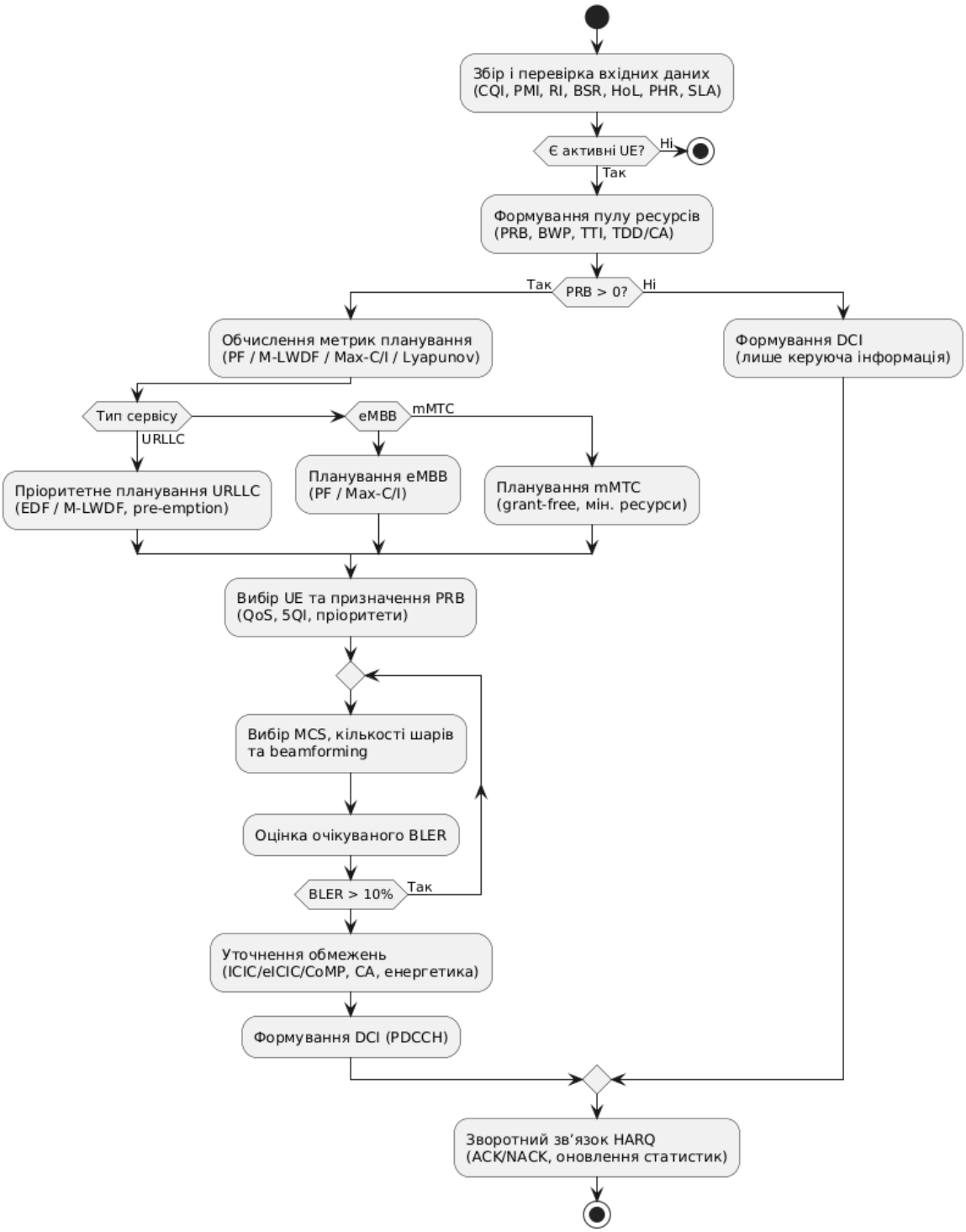


Рисунок 2.2 Блок-схема планування ресурсів у радіомережі

Таблиця 2.4

Класифікація алгоритмів планування ресурсів у радіомережі

Клас алгоритму	Принцип роботи	Сильні сторони	Обмеження / Недоліки	Приклад застосування
Round Robin (RR)	Ресурси розподіляються рівномірно між усіма UE	Простота, справедливість	Ігнорує стан каналу, низька ефективність спектру	LTE/5G у низькому навантаженні
Maximum C/I	Призначення ресурсів UE з найкращим CQI	Максимізація пропускної здатності	Високий ризик несправедливості, слабка підтримка QoS	eMBB користувачі з добрим SINR
Proportional Fair (PF)	Баланс між поточною швидкістю та середнім бітрейтом	Компроміс ефективність/справедливість	Може не гарантувати URLLC дедлайни	Основний планувальник у LTE/NR
M-LWDF / EXP-PF	Врахування QoS (затримки, пріоритетів) у формулі метрики	Підтримка GBR та критичних сервісів	Вищі обчислювальні витрати	VoIP, URLLC
Lyapunov / Drift-Plus-Penalty	Мінімізація черг та порушень QoS через оптимізацію Lyapunov функції	Формальні гарантії стабільності	Складна реалізація у реальному часі	Системи з великою кількістю UE
Machine Learning (DRL, RL)	Навчання оптимальних політик планування	Адаптивність до складних сценаріїв	Вимагає навчання, ризик нестабільності	5G/6G, Network Slicing

У таблиці 2.4 наведено основні класи алгоритмів планування, що застосовуються в мобільних мережах. Вона демонструє різницю між простими підходами (наприклад, *Round Robin*), які забезпечують справедливий розподіл ресурсів, але не враховують стан каналу, та більш складними методами, що оптимізують продуктивність мережі або підтримують якість обслуговування. Кожен клас алгоритмів має свої переваги й недоліки: наприклад, *Maximum C/I* досягає максимальної пропускної здатності, проте ігнорує справедливість, тоді як *M-LWDF* або *Lyapunov* гарантують QoS і стабільність, але вимагають більших

обчислювальних ресурсів. Окрему увагу приділено сучасним алгоритмам на базі машинного навчання, які розглядаються як основа для 5G/6G систем.

Таблиця 2.5

Відповідність алгоритмів типам сервісів

Тип сервісу (5QI)	Критерії QoS	Оптимальні алгоритми	Додаткові механізми
URLLC (Ultra Reliable Low Latency)	Мін. затримка (<1 мс), висока надійність	EDF, M-LWDF, EXP-PF	Pre-emption, mini-slots, HARQ з низьким таймінгом
eMBB (Enhanced Mobile Broadband)	Висока швидкість, пропускна здатність	PF, Max-C/I	CA, MU-MIMO, Beamforming
mMTC (Massive Machine Type Communication)	Масовість, низька швидкість	RR, Grant-Free, прості чергові	NB-IoT, груповий доступ, агрегація ACK

Таблиця 2.5 ілюструє, як різні алгоритми планування відповідають специфічним вимогам сервісів у сучасних мобільних мережах. Для *URLLC* ключовим є зниження затримки та забезпечення надійності, тому використовуються алгоритми EDF та M-LWDF з підтримкою механізмів pre-emption. Для *eMBB* пріоритетом є максимальна пропускна здатність, тому ефективними є PF та Max-C/I у поєднанні з технологіями CA, MU-MIMO та beamforming. Для *mMTC* важливим є масштабований доступ для великої кількості пристроїв, що реалізується через прості планувальники (RR) та механізми grant-free доступу. Таблиця демонструє, що жоден алгоритм не є універсальним, і вибір залежить від профілю сервісу та вимог QoS.

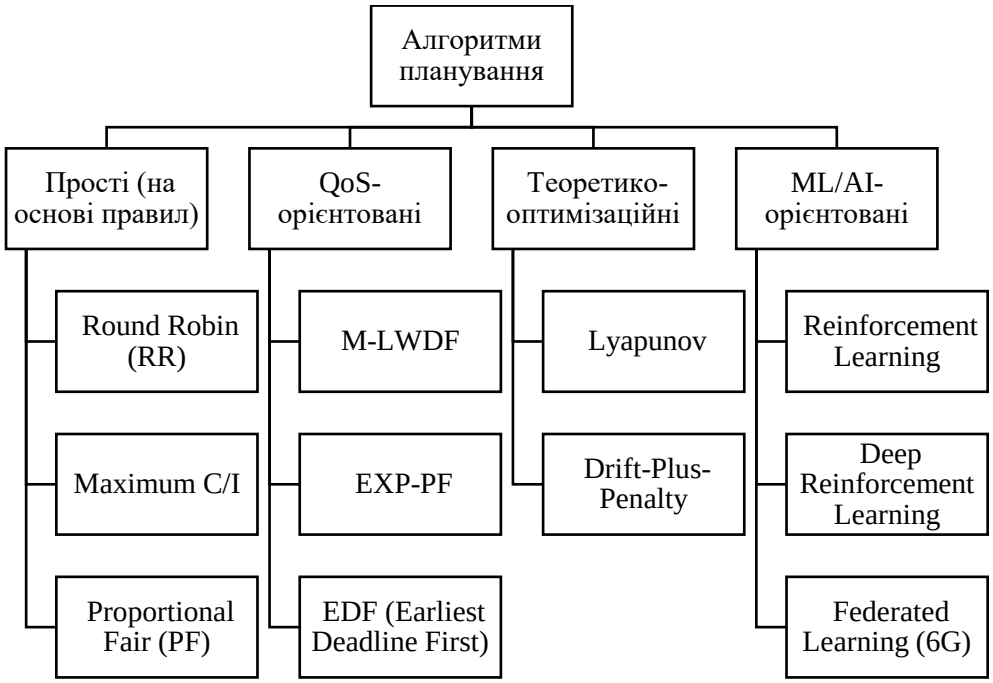


Рисунок 2.3 Загальна класифікація алгоритмів планування

2.3 Методи контролю та прогнозування навантаження в мобільних мережах

У сучасних мобільних мережах методи контролю та прогнозування навантаження все більше розвиваються у напрямку адаптивності та інтелектуалізації. Якщо раніше ключова роль відводилася традиційним механізмам, таким як контроль доступу (Admission Control) чи балансування навантаження (Load Balancing), то сьогодні їх ефективність доповнюється сучасними технологіями когнітивного радіо, віртуалізації мережевих функцій (NFV) та програмно-визначених мереж (SDN). Контроль навантаження вже не обмежується простим моніторингом параметрів, а передбачає автоматичне регулювання потужності, динамічне переналаштування параметрів handover та зміну пріоритетів сервісів у залежності від ситуації. Це дозволяє мінімізувати ймовірність блокування викликів і підвищувати якість обслуговування навіть у пікові години.

Особливе місце займають когнітивні методи, що базуються на концепції «розумного радіо». Вони дозволяють здійснювати інтелектуальний пошук вільних смуг частот, аналізувати ефективність їх використання та оптимізувати розподіл

спектра в реальному часі. Такі методи відкривають шлях до ефективнішого використання ресурсів і динамічного реагування на перевантаження. Паралельно розвивається інтеграція контролю навантаження з NFV та SDN, що забезпечує централізоване й гнучке управління мережевими ресурсами. Завдяки цьому оператор може в реальному часі масштабувати віртуальні мережеві функції, перенаправляти трафік між сегментами мережі та підтримувати стабільність обслуговування навіть за умов нерівномірного розподілу навантаження.

Ще одним напрямком еволюції є використання технологій Big Data та штучного інтелекту. Аналітика великих даних дозволяє не лише прогнозувати рівень завантаженості мережі, але й передбачати поведінку користувачів, наприклад, збільшення трафіку під час масових заходів чи у транспортних вузлах. Методи глибинного навчання, зокрема рекурентні нейронні мережі (LSTM, GRU), забезпечують точне прогнозування часових рядів навантаження та дозволяють враховувати довготривалі залежності, які складно описати традиційними статистичними моделями. Водночас сучасні підходи враховують не лише часову, а й просторову динаміку, що дає змогу будувати моделі просторово-часового прогнозування. Використання геостатистики та просторових нейронних мереж підвищує точність оцінки завантаженості конкретних секторів, що особливо важливо для щільних міських сценаріїв 5G/6G.

Окремий напрям розвитку пов’язаний з прогнозуванням навантаження з урахуванням типу сервісу. Адже URLLC, eMBB та mMTC мають різні профілі трафіку та вимоги до QoS. Для URLLC критичною є затримка та джиттер, для eMBB –середня пропускна здатність і пікові швидкості, тоді як для mMTC – кількість одночасних з’єднань і стабільність обслуговування масових підключень. Це означає, що алгоритми прогнозування повинні враховувати специфіку кожного класу трафіку, що дозволяє оптимізувати мережу точніше і гнучкіше.

Таблиця 2.6

Порівняння методів прогнозування навантаження в мобільних мережах

Метод	Переваги	Недоліки	Приклади застосування
Статистичні методи	Простота реалізації; низькі обчислювальні витрати; добре працюють для стабільних трендів	Низька точність при різких змінах трафіку; не враховують нелінійні залежності	Короткострокове прогнозування навантаження в годинах «пік»
Моделі часових рядів (ARIMA, SARIMA)	Враховують сезонність та довгострокові тренди; добре описують циклічні процеси	Потребують ретельної попередньої обробки даних; слабкі при наявності шуму	Аналіз добових/тижневих змін у використанні трафіку
Машинне навчання (SVM, дерева рішень, ANN)	Можливість моделювати складні залежності; висока точність при достатньому навчанні	Високі обчислювальні витрати; потреба у великих навчальних вибірках	Прогноз пікового навантаження, виявлення перевантажених секторів
Глибинне навчання (LSTM, GRU)	Висока точність прогнозування часових рядів; здатність працювати з довгими залежностями	«Чорна скринька»; складність інтерпретації результатів; велика потреба у даних	Передбачення мобільності користувачів і майбутнього завантаження сот
Гібридні підходи	Поєднання сильних сторін статистики та ML; адаптивність до змінних умов	Складність реалізації; інтеграційні витрати	Інтелектуальне керування ресурсами у мережах 5G/6G

Традиційні методи контролю навантаження базувалися переважно на реактивному управлінні ресурсами, коли система приймала рішення після виникнення перевантаження. У сучасних мережах 4G/5G акцент робиться на проактивних та інтелектуальних методах, здатних передбачати поведінку користувачів і трафіку за допомогою математичних моделей та алгоритмів штучного інтелекту.

Таблиця 2.7

Порівняння традиційних і сучасних методів контролю навантаження в мобільних мережах

Підхід	Характеристика	Переваги	Недоліки	Приклади застосування
Традиційні методи (AC, LB)	Реагування на перевантаження після його виникнення; регулювання доступу та балансування між сотами	Простота реалізації; низькі обчислювальні витрати	Реактивність; низька гнучкість; слабка ефективність у випадку різнорідного трафіку	2G/3G мережі, базові LTE-системи
Адаптивний контроль	Автоматичне регулювання потужності, параметрів handover та QoS залежно від стану мережі	Гнучкість; швидке реагування на зміни; зниження блокування викликів	Потребує складніших алгоритмів; можливі похибки в налаштуванні	LTE-Advanced, початкові 5G-сценарії
Когнітивне радіо	Аналіз спектра та пошук вільних смуг у реальному часі	Висока ефективність використання спектра; самонавчання	Складність апаратної та програмної реалізації	Дослідні мережі 5G, IoT-сценарії
SDN/NFV управління	Централізоване керування ресурсами, масштабування віртуальних функцій	Масштабованість; централізація; висока адаптивність	Залежність від контролера; складність інтеграції	5G Core, хмарні мобільні мережі
Big Data та AI	Аналіз великих обсягів даних і прогнозування навантаження за допомогою ML/AI	Висока точність прогнозу; можливість передбачення поведінки користувачів	Великі вимоги до даних і обчислювальних ресурсів	Прогнозування пікового навантаження, інтелектуальні SON
Просторово-часові методи	Прогнозування з урахуванням як часових, так і географічних факторів	Висока точність; врахування просторової динаміки	Складність моделей; потреба у великих вибірках	Щільні міські сценарії 5G/6G

Для кількісної оцінки завантаженості мережі використовуються базові показники. Один із ключових –коефіцієнт завантаженості ресурсу:

$$U(t) = \frac{R_{\text{used}}(t)}{R_{\text{total}}}, \tag{2.12}$$

де $R_{\text{used}}(t)$–кількість ресурсних блоків, задіяних у момент часу t, а R_{total} – їх загальна кількість. Значення $U(t)$, що прямує до 1, свідчить про критичне навантаження та потенційні збої в обслуговуванні.

Іншою фундаментальною характеристикою є ймовірність блокування викликів, яка визначається класичною формулою Ерланга (Erlang B):

$$P_b = \frac{\frac{A^N}{N!}}{\sum_{k=0}^{N} \frac{A^k}{k!}}, \tag{2.13}$$

де A – інтенсивність навантаження в Ерлангах, N–кількість каналів. Ця формула дозволяє оцінити, яка частина викликів буде відхилена в умовах високого навантаження.

Якщо контроль забезпечує поточну стабільність системи, то прогнозування дозволяє здійснити її адаптацію до майбутніх умов. Одним із поширених підходів є використання моделей часових рядів, зокрема ARIMA, яка описує значення навантаження в момент часу tчерез попередні значення та випадкові збурення:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{2.14}$$

де ϕ_iта θ_j–параметри авторегресії та ковзного середнього, ε_t–випадковий шум. Ця модель добре працює при прогнозуванні короткострокових змін навантаження, наприклад у годинах пік.

Сучасні підходи до прогнозування базуються також на використанні нейронних мереж, зокрема LSTM-моделей, здатних враховувати довгострокові залежності в трафіку:

$$h_t = f(W \cdot x_t + U \cdot h_{t-1} + b), \tag{2.15}$$

де h_t–прихований стан у момент часу t, x_t–вхідний сигнал (наприклад, дані про кількість користувачів або спожиту пропускну здатність), W, U, b–параметри навченої моделі, а $f(\cdot)$–активаційна функція.

Застосування математичних формул у контролі та прогнозуванні навантаження дозволяє перейти від якісних оцінок до кількісних показників, що створює основу для розробки алгоритмів автоматизованого управління ресурсами.

Якщо класичні моделі (Erlang B, коефіцієнт завантаженості) забезпечують базові можливості аналізу, то методи на основі ARIMA та LSTM відкривають шлях до проактивного управління мережею, коли система здатна не лише реагувати на поточні перевантаження, але й передбачати їх виникнення та завчасно переналаштовувати ресурси.

Застосування методів контролю та прогнозування навантаження в мобільних мережах передбачає послідовне поєднання аналітичних та інтелектуальних підходів, що дозволяють не лише оцінювати поточний стан системи, але й передбачати її розвиток у часі. На етапі оперативного управління ключову роль відіграє контроль коефіцієнта завантаженості ресурсів, що розраховується як відношення використаних блоків до загальної кількості доступних. У випадку перевищення певного порогу, наприклад 70–80 %, мережа автоматично ініціює алгоритми балансування трафіку між сусідніми сотами або змінює параметри доступу, щоб уникнути перевантаження. Цей підхід використовується у режимі реального часу та забезпечує стабільність роботи навіть у пікові години.

Паралельно із цим застосовується ймовірнісний аналіз, який базується на формулі Ерланга B. Він дозволяє ще на етапі планування визначати оптимальну кількість каналів і розраховувати ймовірність відмови при різних рівнях навантаження. Такий підхід є основою стратегічного управління ресурсами, оскільки дає змогу заздалегідь закласти у проєкт мережі достатній запас ємності, щоб у майбутньому не виникали масові відмови.

Коли йдеться про передбачення динаміки трафіку, на перший план виходять методи прогнозування. ARIMA-моделі забезпечують можливість короткострокового прогнозу на основі історичних даних, враховуючи регулярні цикли активності користувачів. Це особливо ефективно у міських умовах, де навантаження повторюється вранці, в обідній час та ввечері. Використання таких моделей дозволяє операторам заздалегідь оптимізувати розподіл ресурсів і зменшити ризик перевантаження під час прогнозованих пікових періодів.

У випадках, коли трафік формується складними процесами, зокрема у мережах 5G з великою кількістю IoT-пристроїв, доцільним є застосування

нейронних мереж типу LSTM. Вони дають змогу не лише виявляти короткострокові закономірності, але й будувати довгострокові прогнози з урахуванням сезонних чи контекстних факторів. Наприклад, така модель може передбачати збільшення навантаження у транспортних вузлах у певні дні тижня чи прогнозувати зростання трафіку під час масових заходів.

Сучасні методи контролю та прогнозування навантаження поступово переходять від реактивного підходу, який лише реагував на перевантаження, до проактивного та інтелектуального управління. Їх інтеграція із системами штучного інтелекту, SDN/NFV та концепцією когнітивних мереж забезпечує формування самоорганізованих мереж нового покоління (SON), де навантаження контролюється автоматично і з високим рівнем адаптивності до зовнішніх умов. Це відкриває можливості для підвищення ефективності використання ресурсів, забезпечення високої якості обслуговування користувачів та побудови надійних телекомунікаційних систем 5G і майбутніх 6G.

2.4 Порівняльний аналіз інструментів оптимізації

Оптимізація є ключовим інструментом у підвищенні ефективності мобільних мереж, оскільки від правильного вибору алгоритмів і методів залежить якість обслуговування користувачів, швидкість доступу до послуг та раціональність використання обмежених радіоресурсів. У сучасних дослідженнях можна виокремити три основні групи підходів до оптимізації: класичні методи, евристичні алгоритми та методи машинного навчання.

Класичні методи оптимізації базуються на строгих математичних моделях і часто формулюються як задачі лінійного, нелінійного чи цілочислового програмування. Вони дозволяють отримати оптимальні або наближені до оптимальних рішення для задач розподілу каналів, планування частот і управління пропускною здатністю. Їхня сильна сторона полягає у математичній обґрунтованості та можливості точної перевірки оптимальності рішення. Однак із зростанням складності системи, кількості користувачів і різнорідності сервісів такі

методи втрачають ефективність, оскільки час обчислень значно збільшується, що унеможливлює застосування в режимі реального часу.

Таблиця 2.8

Порівняльна характеристика методів оптимізації в мобільних мережах

Метод	Характеристика	Переваги	Недоліки	Типові сфери застосування
Класичні методи (лінійне, нелінійне, цілочислове програмування)	Використовують строгі математичні моделі; забезпечують оптимальні рішення при невеликих масштабах	Висока точність; формальна обґрунтованість; можливість перевірки оптимальності	Висока обчислювальна складність; погана масштабованість; непридатність для реального часу	Планування частот, оптимізація топології невеликих мереж
Евристичні методи (генетичні алгоритми, мурашині колонії, рой частинок, локальний пошук)	Орієнтовані на пошук «достатньо хороших» рішень у складних багатовимірних задачах	Висока швидкість пошуку рішень; здатність працювати в умовах складних обмежень; гнучкість	Відсутність гарантії оптимального рішення; залежність від параметрів алгоритму; можливі локальні мінімуми	Балансування навантаження, розподіл радіоканалів, оптимізація QoS
Методи машинного навчання (нейронні мережі, підкріплювальне навчання, глибинне навчання)	Використовують дані для навчання моделей, здатних до прогнозування та адаптації	Самонавчання; здатність працювати з Big Data; адаптивність у динамічних умовах; висока точність прогнозів	Потреба у великих обчислювальних ресурсах; складність реалізації; проблема «чорної скриньки»	Прогнозування навантаження, динамічний розподіл ресурсів, управління 5G/6G

Евристичні методи стали відповіддю на обмеження класичних підходів. Вони передбачають використання алгоритмів, що не гарантують пошуку абсолютно оптимального рішення, проте дозволяють отримати прийнятний результат за обмежений час. Прикладами є алгоритми генетичної оптимізації, мурашині колонії, рой частинок або локального пошуку. Їх перевага полягає у здатності знаходити хороші рішення у складних, багатовимірних просторах оптимізації, де класичні методи працюють надто повільно. Водночас евристики можуть страждати від

нестабільності результатів і залежності від параметрів налаштування, що вимагає додаткових експериментів для їх адаптації під конкретну задачу.

Третім і найбільш перспективним напрямом виступають методи машинного навчання, зокрема глибинного навчання та підкріплювального навчання. Вони дозволяють не лише розв'язувати задачі оптимізації, але й прогнозувати динаміку системи, виявляти приховані залежності та адаптуватися до змін у реальному часі. Наприклад, нейронні мережі можуть прогнозувати навантаження на базові станції та завчасно перерозподіляти ресурси, тоді як алгоритми підкріплювального навчання здатні навчатися оптимальним стратегіям управління мережею через взаємодію з середовищем. Серед переваг ML-методів –висока гнучкість, здатність працювати з великими даними та самоудосконалюватися. Проте слід враховувати їхню складність у реалізації, потребу в значних обчислювальних ресурсах і ризик «чорної скриньки», коли оператор не завжди може пояснити, чому алгоритм прийняв те чи інше рішення.

Класичні методи зазвичай формулюються як задача оптимізації з обмеженнями. Загальний вигляд можна записати так:

$$\min f(x), x \in X, g_i(x) \leq 0,\ i = 1, \dots, m, \tag{2.16}$$

де $f(x)$–цільова функція (наприклад, мінімізація затримки чи енергоспоживання), X–допустима область рішень, $g_i(x)$–обмеження (наприклад, обмеження на смуги частот чи кількість користувачів у соті). У випадку планування частот це може бути задача мінімізації інтерференції між сусідніми сотами:

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} I_{ij} \cdot \delta(f_i, f_j), \tag{2.17}$$

де I_{ij}–коефіцієнт взаємних перешкод між сотами iта j, а $\delta(f_i, f_j)$–функція, що показує, чи використовується однакова частота в обох сотах.

Евристичні методи зазвичай імітують природні або еволюційні процеси. Наприклад, у генетичних алгоритмах нове покоління рішень створюється за допомогою формули:

$$x^{(t+1)} = \text{crossover}(x_i^{(t)}, x_j^{(t)}) + \text{mutation}(x^{(t)}), \tag{2.18}$$

де $x^{(t)}$–рішення в поколінні t, а функції crossoveri mutationвідповідають за рекомбінацію та випадкові зміни. Це дозволяє знаходити рішення у складних багатовимірних просторах, де класичні методи не працюють ефективно.

Таблиця 2.9

Порівняння методів оптимізації за ключовими критеріями

Критерій	Класичні методи	Евристичні методи	Машинне навчання
Точність	Висока (оптимальне рішення)	Середня (наближене рішення)	Висока (залежить від якості даних)
Швидкість	Низька для великих задач	Висока	Середня (тренування довге, прогноз швидкий)
Масштабованість	Обмежена	Висока	Висока
Адаптивність	Низька	Середня	Висока
Обчислювальні ресурси	Помірні	Помірні	Високі
Придатність для реального часу	Обмежена	Добра	Висока
Типові застосування	Планування частот, невеликі задачі	Розподіл каналів, QoS	Прогнозування навантаження, 5G/6G оптимізація

Методи машинного навчання формулюють оптимізацію як задачу мінімізації функції втрат. Наприклад, у нейронних мережах для прогнозування навантаження використовується функція середньоквадратичної похибки (MSE):

$$L = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{2.19}$$

де y_i – реальне значення трафіку, $\hat{y}_i$–прогноз моделі, n–кількість спостережень. Оптимізація відбувається за допомогою алгоритму зворотного поширення похибки (backpropagation), який змінює ваги нейронів для мінімізації L.

У підкріплювальному навчанні задача оптимізації визначається як максимізація очікуваної винагороди:

$$\max_{\pi} \ \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t \, r_t \Big], \tag{2.20}$$

де π – стратегія прийняття рішень, r_t–винагорода у момент часу t, γ–коефіцієнт дисконтування. У контексті мобільних мереж ця формула описує пошук політики керування ресурсами, яка максимізує ефективність мережі (наприклад, мінімізує блокування або затримки).

Таблиця 2.10

Приклади застосування методів оптимізації в мобільних мережах

Завдання	Класичні методи	Евристичні методи	Машинне навчання
Планування частот	Лінійне програмування	Генетичні алгоритми	Нейронні мережі для прогнозування інтерференції
Балансування навантаження	Цілочислове програмування	Рой частинок, мурашині алгоритми	Підкріплювальне навчання
Розподіл пропускної здатності	Нелінійне програмування	Локальний пошук	Глибинне навчання (DNN, LSTM)
Прогнозування трафіку	Статистичні моделі (ARIMA)	Евристики для трендів	Нейронні мережі, CNN, RNN
Оптимізація QoS/QoE	Формальні обмеження	Евристики для адаптивності	Гібридні ML-алгоритми

Таблиця 2.11

Сильні та слабкі сторони методів

Метод	Сильні сторони	Слабкі сторони
Класичні методи	Математична точність, строгі гарантії	Повільні, погано масштабуються
Евристики	Гнучкі, швидкі, придатні для складних задач	Немає гарантії оптимальності, чутливі до параметрів
Машинне навчання	Адаптивність, прогнозування, робота з Big Data	Високі вимоги до ресурсів, складність пояснення результатів

Інструменти оптимізації в телекомунікаційних системах використовуються для забезпечення ефективного використання обмежених радіоресурсів, підтримки необхідного рівня якості обслуговування (QoS) та підвищення задоволеності користувачів (QoE). Основна мета полягає в тому, щоб узгодити технічні можливості мережі з потребами користувачів і динамічними умовами середовища.

Класичні методи оптимізації застосовуються у випадках, коли система може бути формалізована через строгі математичні обмеження. Наприклад, задача планування частот у стільникових мережах часто формулюється як задача цілочислового програмування, де цільова функція полягає у мінімізації інтерференції між сусідніми сотами. Такі методи також використовуються для розрахунку оптимальної потужності передавачів, мінімізації затримки пакетів або визначення найкращого маршруту для трафіку. Попри високу точність, ці підходи обмежені масштабованістю: вони ефективні для невеликих мереж, але малопридатні для систем 5G/6G з великою кількістю користувачів і сервісів.

Евристичні методи дозволяють ефективно працювати у великих мережах із високою складністю та невизначеністю. Вони застосовуються для балансування навантаження між базовими станціями, коли класичні методи надто повільні. Наприклад, генетичні алгоритми можуть адаптивно перерозподіляти користувачів між сотами залежно від рівня сигналу й завантаженості. Алгоритми рою частинок або мурашині алгоритми можуть шукати оптимальні маршрути для трафіку в транспортних мережах, враховуючи змінність каналів і динаміку користувацької активності. Завдяки швидкому зближенню до прийнятного рішення, евристики особливо корисні для сценаріїв, де потрібна висока швидкість реакції, але не завжди обов’язково отримати математично оптимальний результат.

Методи машинного навчання відкривають новий етап оптимізації мобільних мереж, оскільки вони здатні не тільки реагувати на поточний стан, а й прогнозувати майбутні зміни. Наприклад, використання рекурентних нейронних мереж (RNN, LSTM) дає змогу передбачати трафік у різних географічних регіонах і завчасно виділяти додаткові ресурси для пікових навантажень. Алгоритми підкріплювального навчання застосовуються для самонавчального керування мережею: система взаємодіє із середовищем, отримує винагороди за правильні дії (наприклад, зменшення ймовірності блокування виклику) і поступово формує оптимальну політику управління. Крім того, глибинне навчання використовується для прогнозування рівня інтерференції, що допомагає динамічно змінювати розподіл частот.

На практиці найкращі результати забезпечує гібридне застосування інструментів оптимізації. Наприклад, класичні методи можуть використовуватися для довгострокового стратегічного планування частот і топології мережі, евристики –для середньострокових адаптивних рішень (балансування навантаження в регіоні), а машинне навчання –для оперативного прогнозування та динамічного управління у режимі реального часу. Такий підхід дозволяє поєднати точність, швидкість і адаптивність, що є критично важливим для сучасних мобільних мереж, орієнтованих на підтримку сервісів з різними вимогами до затримки, пропускної здатності та надійності.

Порівняльний аналіз показує, що класичні методи доцільно використовувати для невеликих або формалізованих задач, де критично важлива математична точність; евристики –для середніх за складністю сценаріїв, де потрібно швидко знайти прийнятне рішення; а методи машинного навчання –для великих і динамічних мереж, які вимагають високого рівня адаптивності та прогнозування. Найбільш ефективною стратегією у сучасних телекомунікаційних системах є гібридний підхід, коли класичні методи застосовуються на рівні глобального планування, евристики –для швидкої локальної оптимізації, а машинне навчання – для адаптивного управління та прогнозування в реальному часі.

3 РОЗРОБКА ТА РЕАЛІЗАЦІЯ МЕТОДУ ОПТИМІЗАЦІЇ ПРОЦЕСІВ ПЛАНУВАННЯ Й КОНТРОЛЮ В СИСТЕМАХ МОБІЛЬНОГО ЗВ'ЯЗКУ

3.1 Формалізація задачі оптимізації процесів планування та контролю на основі критерію QoE

Оптимізація планування та контролю в системах мобільного зв'язку формулюється як багатопараметрична задача, у якій потрібно максимізувати якість користувацького досвіду при обмеженості ресурсів радіодоступу, затримках, рівні завад та обмеженій пропускній здатності. Нехай у секторі базової станції одночасно обслуговується множина користувачів $U = \{1,2, \dots, N\}$, а в кожному планувальному інтервалі часу tоператор має розподілити між ними скінченну кількість ресурсних блоків R. Розподіл позначається змінними $x_i(t)$, тоді загальне ресурсне обмеження має вигляд

$$\sum_{i=1}^{N} x_i\,(t) \le R. \tag{3.1}$$

Символ $x_i(t)$позначає кількість ресурсних блоків (RB), виділених користувачу iу момент часу t; R– загальна кількість доступних RB у цьому інтервалі; N– кількість активних користувачів. Формула виражає фундаментальну властивість радіоканалу: ресурс подільний і обмежений, а отже оптимізація – це завжди гра з нульовою сумою, де виграш одного абонента означає зменшення потенційної корисності для інших. Ця «фішка» математично примушує алгоритм уникати тривіального рішення «дати все найкращому».

Згідно з моделлю LTE/5G швидкість користувача визначається добутком кількості призначених ресурсів і спектральної ефективності:

$$T_i(t) = x_i(t) \cdot S_i(t), \tag{3.2}$$

де $T_i(t)$– миттєва пропускна здатність абонента i, а $S_i(t)$– спектральна ефективність у біт/с/Гц, визначена рівнем модуляції та кодовим коефіцієнтом. Щоб формалізувати залежність $S_i(t)$від умов радіоканалу, вводимо співвідношення сигнал–завада–шум:

$$SINR_i(t) = \frac{P_i}{I_i + N_0}, \quad (3.3)$$

де $SINR_i(t)$– індекс якості каналу для абонента i; P_i– потужність корисного сигналу; I_i– сумарна потужність інтерференції; N_0– рівень шуму. Підвищення $SINR$збільшує можливий рівень модуляції. Це відображено в залежності спектральної ефективності:

$$S_i(t) = \log_2(1 + SINR_i(t)). \quad (3.4)$$

Символ $\log_2$ визначає бітову ємність каналу відповідно до формули Шеннона; додаток «1 +» гарантує додатність аргументу; $S_i(t)$виражає максимально можимий бітовий приріст на герц частоти. «Фішка» полягає в тому, що приріст пропускної здатності має спадну маржинальну корисність: приріст SINR на низьких рівнях дає величезний виграш, а на високих – майже не впливає, що робить систему чутливою до абонентів «на межі соти» та змушує алгоритм балансувати між ефективністю та справедливістю.

Оскільки затримка в пакетних системах визначається накопиченням черг, застосовується стандартна M/M/1-модель:

$$D_i(t) = \frac{1}{\mu_i(t) - \lambda_i(t)}, \quad (3.6)$$

де $D_i(t)$– середня затримка для користувача i; $\lambda_i(t)$– інтенсивність надходження пакетів; $\mu_i(t)$– інтенсивність обслуговування, яку визначає $T_i(t)$. Якщо $\lambda_i(t)$наближається до $\mu_i(t)$, знаменник прямує до нуля і затримка стає вибуховою. Саме тут ключовий сенс: навіть достатня пропускна здатність не гарантує низького часу доставки, якщо користувач перевантажує свою «чергу». У 5G цей ефект формує основну проблему low-latency.

Джиттер як варіативність затримки формально подається як

$$J_i(t) = | D_i(t) - \bar{D}_\iota |, \quad (3.7)$$

де $J_i(t)$– джиттер користувача i; $D_i(t)$– миттєва затримка; $\bar{D}_\iota$– середня затримка на інтервалі спостереження. Формула вимірює абсолютну різницю між поточним і середнім значенням, демонструючи коливання часу доставки. «Фішка» в тому, що навіть мала затримка може знищити QoE, якщо коливання великі: відео починає «заїдати», а голос – рватися.

Ймовірність втрат пов’язана з відношенням сигнал/завада і залежить від імовірності появи бітової помилки:

$$L_i(t) = 1 - (1 - BER_i(t))^k, \quad (3.8)$$

де $L_i(t)$– імовірність втрат пакета; $BER_i(t)$– імовірність бітової помилки при заданому MCS; k– кількість бітів у пакеті. Функція показує, що навіть незначний BER за умов великого розміру пакетів породжує суттєві втрати. Тому втрати – це експоненційний руйнівник QoE, а не лінійна величина.

Інтегральну залежність QoE подамо у вигляді утилітарної функції:

$$QoE_i(t) = \alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t), \quad (3.9)$$

де $QoE_i(t)$– рівень користувацького задоволення сервісом; $\ln(T_i(t))$– логарифмічний внесок пропускної здатності; $D_i(t)$, $J_i(t)$, $L_i(t)$– штрафні складові за затримку, варіативність затримки та втрати; $\alpha, \beta, \gamma, \delta$– вагові коефіцієнти, що визначають чутливість моделі до кожного фактору. Логарифмічний характер корисності від пропускної здатності пояснює, чому приріст швидкості понад рівень комфортного HD-відео майже не збільшує QoE, тоді як додаткова затримка чи 1–2% втрат можуть його повністю зруйнувати. У цьому й основна концептуальна «фішка»: модель прив’язує оптимізацію до людського сприйняття, а не до технічних мегабітів.

Оптимізаційна постановка полягає у максимізації сумарної QoE:

$$\max_{x(t)} \sum_{i=1}^{N} Qo\, E_i(t), \quad (3.10)$$

де $\sum_{i=1}^{N}$ узагальнює оптимізаційний ефект для всіх користувачів, а не для окремого «улюбленця» з ідеальним каналом. Це гарантує, що система не буде вигравати у throughput для одного за рахунок катастрофи QoE у всіх інших.

Щоб уникнути диспропорцій «з’їдання ресурсу» користувачами з найкращим SINR, вводиться показник справедливості Джейна:

$$F = \frac{\left(\sum_{i=1}^{N} T_i\right)^2}{N \sum_{i=1}^{N} T_i^2}. \quad (3.11)$$

У виразі $\sum_{i=1}^{N} T_i$у чисельнику стоїть квадрат суми пропускних здатностей, а у знаменнику – сума квадратів. Якщо всі користувачі отримують подібні швидкості,

знаменник і чисельник зростають однаково, і $F \to 1$. Якщо хтось монополізує ресурс, знаменник різко збільшується, і $F \to 0$. Це створює математичний бар'єр проти ресурсної нерівності – той самий «антиолігархічний механізм» у радіомережі.

Завершальна задача формулюється як багатомірна оптимізація з обмеженнями:

$$\max_{x(t)} \sum_{i=1}^{N} [\alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t)] \quad (3.12)$$

при умовах

$$\sum_{i=1}^{N} x_i\,(t) \le R, D_i(t) \le D_i^{max}, L_i(t) \le L_i^{max}, x_i(t) \ge 0. \quad (3.13)$$

У цій сукупності D_i^{max}– максимально допустима затримка для сервісу користувача i, L_i^{max}– гранично допустимі втрати, а $x_i(t) \ge 0$– умова невід'ємності радіоресурсів. Сенс цих формул у тому, що вони комбінують суб'єктивну якість (QoE), об'єктивні параметри (QoS), фізичні характеристики каналу (SINR), статистику трафіка (інтенсивність надходження), обмеження радіодоступу (R) і гуманітарний критерій рівності (F). Іншими словами, «фішка» моделі: вона перетворює хаотичну радіосереду і людські очікування на формалізовану інженерну ціль.

Фінальна постановка задачі:

$$\max_{x(t)} \sum_{i=1}^{N} [\alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t)]$$

за умов:

$$\sum_{i=1}^{N} x_i\,(t) \le R$$

$$D_i(t) \le D_i^{max},\ J_i(t) \le J_i^{max},\ L_i(t) \le L_i^{max}$$

$$T_i(t) \ge T_i^{min},\ x_i(t) \ge 0$$

Це – мультикритеріальна оптимізація, приведена до єдиної мети через QoE.

3.2 Математична модель радіоресурсів і навантаження в мобільній мережі

У процесі оптимізації планування радіоресурсів необхідно формально врахувати взаємозв'язок між кількістю доступних ресурсних блоків, спектральною ефективністю, статистикою трафіка, рівнем завад оточення та можливим перевантаженням комірки. Відмінною рисою радіомереж мобільного зв'язку є те, що навантаження формується випадково-стохастичним чином, а потреба в радіоресурсах змінюється під впливом кількості активних користувачів, їх мобільності та характеру сервісу. Тому математична модель планування повинна включати два ядра: модель пропускної здатності каналу та модель трафіка/навантаження.

Нехай доступна смуга частот сектора базової станції дорівнює B. Тоді максимальна теоретична пропускна здатність радіоінтерфейсу, згідно з Шенноном, дорівнює

$$C = B\log_2(1 + SINR), \tag{3.16}$$

де C– пікова пропускна здатність каналу в біт/с; B– виділена смуга частот; $SINR$– відношення сигнал/завади/шум. Цей вираз не описує реальні кінцеві швидкості абонента, однак створює верхню межу, яку жоден алгоритм планування не зможе перевищити.

Функціонально формула демонструє характерну властивість радіоканалу: збільшення смуги дає лінійний приріст, тоді як приріст потужності сигналу дає логарифмічний ефект, що робить боротьбу за спектр критичнішою, ніж за збільшення потужності.

Реальне навантаження на комірку визначається інтенсивністю λнадходження заявок на обслуговування і середнім часом обслуговування $1/\mu$. У моделі масового обслуговування система представлена як черга M/M/1 із коефіцієнтом завантаження

$$\rho = \frac{\lambda}{\mu}, \tag{3.17}$$

де ρ– частка використання ресурсу; λ– інтенсивність надходження пакетів; μ– інтенсивність обробки пакетів. Якщо $\rho \to 1$, система входить у режим насичення, а пропускна здатність стає нестабільною.

Сенс цієї залежності полягає в тому, що навіть за високої теоретичної ємності каналу перевантаження трафіком робить систему некерованою, і саме тут оптимізатор має втручатися: не для нарощування швидкості, а для стабілізації поведінки.

Ймовірність блокування нових з'єднань визначається класичною формулою Ерланга B:

$$P_{block} = \frac{\frac{A^n}{n!}}{\sum_{k=0}^{n} \frac{A^k}{k!}}, \tag{3.18}$$

де P_{block}– імовірність блокування звернень абонентів; $A = \lambda/\mu$– навантаження в ерлангах; n– кількість доступних каналів еквівалентного обслуговування. Формула демонструє фундаментальну властивість обмежених ресурсів: блокування зростає нелінійно при перевищенні навантаження, а отже навіть помірне збільшення кількості активних користувачів може різко підвищити імовірність відмови в обслуговуванні. «Фішка» полягає у тому, що оператор не контролює попит, але може оптимізувати ресурс так, щоб стримати ріст блокувань.

Для трафіка передачі даних у багатокористувацькому OFDMA-секторі показником нормованого навантаження є співвідношення між реально зайнятою пропускною здатністю й максимально можимою:

$$\eta = \frac{\sum_{i=1}^{N} T_i}{C}, \tag{3.19}$$

де η– коефіцієнт завантаження ресурсу; $\sum_{i=1}^{N} T_i$– сумарна пропускна здатність виділена всім користувачам; C– ємність каналу за формулою Шеннона. Якщо $\eta \approx 1$, сектор працює на межі свого фізичного потенціалу, і будь-яке зростання попиту неминуче переросте в затримки, джиттер чи втрати. Якщо $\eta \ll 1$, це вказує на

недовикористання спектра, що також є небажаним –тобто оптимізація має балансувати між надмірним і недостатнім завантаженням.

Щоб задати залежність між кількістю користувачів і навантаженням у каналі, введемо модель середньої вимоги в ресурсах для одного користувача:

$$x_{avg} = \frac{\bar{\lambda}}{\bar{\mu} S_{avg}}, \tag{3.20}$$

де x_{avg}– середня кількість RB, необхідних для обслуговування середнього трафіка одного абонента; $\bar{\lambda}$– середня інтенсивність трафіка; $\bar{\mu}$– середня швидкість його обробки; S_{avg}– середня спектральна ефективність. Збільшення споживання трафіка або зниження ефективності модуляції призводить до пропорційного зростання вимог до RB. «Фішка» формули проста: вона показує ціну «одного абонента» для ресурсу соти, і це безпосередньо пов’язує комерційну поведінку користувачів з радіоінженерними параметрами.

Оскільки планувальник приймає рішення кожну мілісекунду, необхідно описати модель доступного ресурсу в одному Transmission Time Interval. Нехай один RB має ширину Δf, а тривалість символу становить T_s. Тоді кількість бітів, яку здатен перенести один RB з урахуванням модуляції M, становить:

$$b_{RB} = M \cdot \Delta f \cdot T_s, \tag{3.21}$$

де b_{RB}– кількість бітів на ресурсний блок; M– кількість бітів на символ (1 для BPSK, 2 для QPSK, 4 для 16-QAM, 6 для 64-QAM, 8 для 256-QAM); Δf– частотна ширина RB; T_s– тривалість OFDM-символу. Вираз демонструє цілком матеріальну природу «високої швидкості»: це не просто мегабіти, а кількість символів і бітів, здатних пройти через одного оператора спектру у фізичних параметрах часу і частоти.

Для мультимедійних сервісів із гарантованою якістю введемо вимогу мінімальної пропускної здатності:

$$T_i \geq T_i^{min}, \tag{3.22}$$

де T_i^{min}– мінімальна швидкість для підтримки класу сервісу користувача i. Відео FullHD потребує 5–8 Мбіт/с, 4K –понад 20 Мбіт/с, відеоконференції –мінімальних

затримок. Отже, якщо система не здатна гарантувати навіть мінімум –QoE зривається.

Для сервісів ultra-reliable low-latency передбачено вимогу до надійності:

$$L_i \leq 10^{-5}, \tag{3.23}$$

де L_i– імовірність втрати пакетів для користувача i. Цей поріг демонструє не статистичну, а практично безпомилкову необхідність доставки: URLLC не допускає «імовірності помилки», він допускає лише її нульову наближеність.

Зведена система рівнянь та залежностей:

$$\begin{cases} C = B\log_2(1 + SINR), \\ \sum_{i=1}^{N} x_i \leq R, \\ T_i = x_i \cdot S_i, S_i = \log_2(1 + SINR_i), \\ \eta = \frac{\sum_{i=1}^{N} T_i}{C}, \\ \rho = \frac{\lambda}{\mu}, P_{block} = \frac{\frac{A^n}{n!}}{\sum_{k=0}^{n} \frac{A^k}{k!}}, \\ x_{avg} = \frac{\bar{\lambda}}{\bar{\mu} S_{avg}} \end{cases} \tag{3.24}$$

утворює загальну математичну модель радіоресурсів і навантаження в секторі мобільної мережі.

У цій системі кожна формула виконує свою роль: перша задає фізичну межу ємності, друга – обмеження ресурсу, третя – індивідуальні пропускні здатності абонентів, четверта – сумарний рівень завантаження спектра, п’ята – поведінку черг та імовірність блокування, шоста – середню «вартість» одного користувача в одиницях ресурсу.

Особливість такої моделі полягає в тому, що вона одночасно враховує фізику радіоканалу, статистику трафіка і вимоги до обслуговування, створюючи формальний базис для подальшої розробки оптимізаційних алгоритмів, які вже будуть орієнтовані на максимізацію QoE в межах цих об’єктивних обмежень.

Отже, математична модель радіоресурсів і навантаження задає три ключові інженерні сенси.

Перший сенс: планування не може йти всупереч фізиці, бо Shannon-межа встановлює стелю, що не пробивається оптимізацією. Другий сенс: планування не може ігнорувати статистику, тому що черги, навантаження й блокування визначають реальний користувацький досвід. Третій сенс: планувальник має бути сервіс-обізнаним, тобто не розподіляти біт/с механічно, а адаптувати їх під вимоги різних класів трафіка –від м'якого відео до жорсткого URLLC. Саме тому модель навантаження є фундаментом для усіх алгоритмів, які будуть розроблені у наступних підрозділах.

3.3 Оптимізаційні алгоритми планування ресурсів у мобільній мережі

На основі побудованої в попередніх підрозділах математичної моделі процеси планування й контролю радіоресурсів у мобільній мережі розглядаються як задача оптимізації, в якій необхідно для кожного планувального інтервалу часу вибрати такий розподіл ресурсних блоків між користувачами, що забезпечує максимальний рівень якості користувацького досвіду за умов обмеженості ресурсу та вимог до показників якості обслуговування. Для цього інтегральну оцінку якості досвіду $QoE_i(t)$, визначену як функцію від пропускної здатності, затримки, джиттера і втрат, доцільно інтерпретувати як корисність користувача. Позначимо корисність у вигляді

$$U_i(t) = QoE_i(t) = \alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t), \qquad (3.25)$$

де $U_i(t)$– корисність (utility) для i-го користувача в момент часу t; $\alpha, \beta, \gamma, \delta$– вагові коефіцієнти, що задають відносну важливість пропускної здатності, затримки, джиттера та втрат; $T_i(t)$– миттєва пропускна здатність, яка визначається через кількість виділених ресурсних блоків та спектральну ефективність; $D_i(t)$, $J_i(t)$і $L_i(t)$– відповідно затримка, варіативність затримки та ймовірність втрати пакетів для користувача i. Таким чином, ми явно ототожнюємо корисність з QoE, тобто алгоритм планування працює не з «сирими» технічними параметрами, а з інтегральним показником, що відображає сприйняття сервісу користувачем.

З урахуванням введеного вектора розподілу ресурсів $x(t) = \{x_1(t), x_2(t), \dots, x_N(t)\}$, де $x_i(t)$– кількість ресурсних блоків, виділених користувачу iв інтервалі t, задача планування формулюється як

$$x^*(t) = \arg\max_{x(t)} \sum_{i=1}^{N} U_i\,(t), \tag{3.26}$$

де $x^*(t)$– оптимальний вектор призначення ресурсів, а $\sum_{i=1}^{N} U_i\,(t)$– сумарна корисність усіх користувачів у секторі. Ця формула відображає ключову ідею: в кожен момент часу планувальник шукає такий розподіл ресурсів, який максимізує сумарний QoE, причому рішення є динамічним і адаптивним, оскільки перераховується для кожного планувального інтервалу. Саме ця динамічність і є «фішкою» моделі: оптимум не є статичним, він постійно слідує за змінами каналу, навантаження та типу сервісів.

При цьому розподіл ресурсів обмежений фізичною кількістю доступних ресурсних блоків і вимогами до якості обслуговування. Система обмежень має вигляд

$$\sum_{i=1}^{N} x_i\,(t) \le R, D_i(t) \le D_i^{max}, L_i(t) \le L_i^{max}, x_i(t) \ge 0, \tag{3.27}$$

де R– загальна кількість ресурсних блоків у секторі; D_i^{max}– максимально допустима затримка для класу сервісу i-го користувача; L_i^{max}– гранично допустима ймовірність втрати пакетів; умова $x_i(t) \ge 0$гарантує невід’ємність виділеного ресурсу. Обмеження означають, що навіть якщо збільшення ресурсу для окремого користувача формально підвищує сумарну корисність, таке рішення є неприйнятним, якщо при цьому порушуються сервісні вимоги щодо затримки та надійності.

Враховуючи, що залежності $T_i(t)$, $D_i(t)$, $J_i(t)$, $L_i(t)$від $x_i(t)$є нелінійними та стохастичними, аналітичне розв’язання цієї задачі в реальному часі є практично неможливим. Тому для її розв’язання застосовуються числові та евристичні алгоритми, серед яких базовим є планувальник типу Proportional Fair. У класичному вигляді пропорційно-справедливий алгоритм для кожного користувача обчислює показник

$$PF_i(t) = \frac{T_i(t)}{\bar{T}_i(t)}, \tag{3.28}$$

де $PF_i(t)$– значення пропорційно-справедливого критерію для користувача i; $T_i(t)$– поточна пропускна здатність; $\bar{T_i}(t)$– усереднена за часом пропускна здатність, яку користувач отримував на попередніх інтервалах. Планувальник у кожний TTI віддає пріоритет користувачу з максимальним значенням $PF_i(t)$. Якщо користувач тривалий час отримував недостатньо ресурсу, його середня швидкість $\bar{T_i}(t)$мала, і тому навіть невелике зростання $T_i(t)$різко збільшує $PF_i(t)$. Такий механізм реалізує ефект компенсації: алгоритм «підтягує» відстаючих користувачів і не дозволяє одному абоненту постійно домінувати за рахунок кращого стану каналу.

Однак класичний PF враховує лише пропускну здатність і не оперує затримкою, джиттером і втратами, тобто оптимізує QoS лише частково. Для переходу до QoE-орієнтованого планування в роботі пропонується модифікований показник, який включає штраф за погіршення якості сервісу. Такий показник можна задати у вигляді

$$QPF_i(t) = \frac{T_i(t)}{\bar{T_i}(t)} \cdot \frac{1}{1+\beta D_i(t)+\gamma J_i(t)+\delta L_i(t)}, \quad (3.29)$$

де $QPF_i(t)$– QoE-орієнтований пропорційно-справедливий критерій; множник $\frac{1}{1+\beta D_i(t)+\gamma J_i(t)+\delta L_i(t)}$зменшує значення показника для користувачів із великою затримкою, великим джиттером або підвищеними втратами. При цьому коефіцієнти β, γ, δті самі, що використовуються у функції корисності $U_i(t)$. Сенс такої модифікації полягає в тому, що планувальник не лише заохочує користувачів із добрими миттєвими можливостями каналу та низькою історичною швидкістю, але й додатково враховує, наскільки поточний сервіс для цього користувача є стабільним і надійним. Іншими словами, пріоритет отримує не просто «спритний» канал, а канал, де додатковий ресурс реально покращує суб'єктивний досвід.

Важливою перевагою QoE-орієнтованого переписування PF-критерію є узгодженість із загальною цільовою функцією. Оскільки $U_i(t) = QoE_i(t)$та $QPF_i(t)$побудовані на тих самих змінних $T_i(t), D_i(t), J_i(t), L_i(t)$, вибір користувача з максимальним $QPF_i(t)$на кожному кроці є локальним наближенням до глобальної задачі максимізації $\sum_{i=1}^{N} U_i\ (t)$. Таким чином, модифікований PF виступає швидким

евристичним алгоритмом, який реалізує принципи QoE-оптимізації без необхідності розв’язувати повну нелінійну задачу на кожному кроці.

Ще одним підходом до розв’язання задачі оптимального планування є використання ітераційних градієнтних методів, у яких вектор розподілу ресурсів оновлюється за правилом

$$x_i(t+1) = x_i(t) + \eta \cdot \frac{\partial U_i(t)}{\partial x_i(t)}, \tag{3.30}$$

де $x_i(t+1)$– нове значення ресурсу для користувача iу наступному планувальному інтервалі; η– коефіцієнт кроку (швидкість навчання); $\frac{\partial U_i(t)}{\partial x_i(t)}$– частинна похідна корисності за виділеним ресурсом. Якщо приріст корисності від додаткового ресурсного блоку є суттєвим (похідна додатна і велика за модулем), алгоритм збільшує $x_i(t)$; якщо приріст малий або зменшення ресурсу дає вищу сумарну корисність, виділений ресурс зменшується. Такий підхід відображає класичну економічну ідею граничної корисності, коли ресурс «інвестується» у ті напрямки, де він дає максимальний приріст QoE.

У випадках, коли цільовий простір є складним, а функція корисності може мати декілька локальних максимумів, доцільно застосовувати еволюційні алгоритми, наприклад генетичні. У цьому разі один потенційний розподіл $x(t)$розглядається як «особина» з певним генотипом, а якість рішення оцінюється фітнес-функцією

$$F(x) = \sum_{i=1}^{N} U_i, \tag{3.31}$$

де $F(x)$– числова оцінка якості конкретного варіанта розподілу ресурсів; $\sum_{i=1}^{N} U_i$– сумарна корисність, що є аналогом загального QoE у секторі. Процес оптимізації включає формування початкової популяції розподілів, відбір найкращих за значенням $F(x)$, операції схрещування та мутації, які дозволяють досліджувати простір рішень у ширшому діапазоні, ніж це роблять локальні методи. Сильна сторона такого підходу полягає в здатності знаходити близькі до глобального оптимуми рішення без строгих вимог до опуклості функції корисності, що є корисним у складних сценаріях навантаження.

Оптимізаційні алгоритми планування, що базуються на узгодженій моделі $U_i(t) = QoE_i(t)$, забезпечують трирівневий ефект. На базовому рівні класичний PF гарантує пропорційну справедливість між користувачами. На вдосконаленому рівні модифікований QoE-орієнтований показник $QPF_i(t)$дозволяє явно враховувати затримку, джиттер і втрати, тобто переводить планування на рівень суб'єктивної якості досвіду. На поглибленому рівні градієнтні та еволюційні методи дозволяють досліджувати простір розподілів ресурсів у пошуку глобально кращих рішень, які можна використовувати як еталон чи орієнтир для побудови спрощених онлайн-алгоритмів. У сукупності це створює формалізований і водночас практично реалізований механізм оптимізації процесів планування й контролю радіоресурсів, який безпосередньо орієнтований на максимізацію QoE в сучасних мережах мобільного зв'язку.

На відміну від класичних підходів, де у якості utility використовують тільки $\ln(T_i(t))$і оптимізують суто пропускну здатність, у роботі utility ототожнено з QoE через розширений вираз

$$U_i^{(QoE)}(t) = \alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t), \qquad (3.32)$$

тобто вперше на рівні алгоритму планування вводиться корисність, яка одночасно враховує швидкість, затримку, джиттер і втрати, а не лише throughput.

Запропоновано QoE-орієнтований варіант пропорційно-справедливого планувальника у вигляді показника $QPF_i(t)$, де до класичного PF-дробу $\frac{T_i(t)}{\bar{T}_i(t)}$ додано нормувальний штрафовий множник $\frac{1}{1+\beta D_i(t)+\gamma J_i(t)+\delta L_i(t)}$. Це дозволяє алгоритму PF не лише "вирівнювати" користувачів за пропускною здатністю, а й віддавати пріоритет тим, у кого додатковий ресурс реально покращує суб'єктивну якість сервісу (зменшує затримку, джиттер, втрати), а не просто збільшує мегабіти.

У якості математичної основи для побудови оптимізаційних алгоритмів вперше компактно сформульовано загальну задачу $\arg\max \sum U_i(t)$з QoE-орієнтованою utility та прямим зв'язком із градієнтними і еволюційними методами: формула градієнтного оновлення показує, як можна реалізувати адаптивний динамічний планувальник, а фітнес-функція $F(x)$робить можливим використання генетичних

алгоритмів для пошуку глобально кращих розподілів ресурсу у складних сценаріях навантаження.

3.4 Використання машинного навчання для прогнозування навантаження та показників QoS у мобільній мережі

Сучасні системи мобільного зв'язку характеризуються високою стохастичністю трафіка, часовою кореляцією навантаження та непередбачуваними коливаннями параметрів обслуговування. Це унеможливлює ефективне статичне планування ресурсів та зумовлює перехід до динамічних моделей, здатних передбачати майбутні стани мережі. Одним із перспективних підходів є застосування машинного навчання (ML) для прогнозування параметрів системи, що безпосередньо впливають на якість сервісу –інтенсивності навантаження $\lambda(t)$, пропускної здатності $T_i(t)$, затримки $D_i(t)$та втрат пакетів $L_i(t)$. У цьому контексті ML виступає механізмом попереджувального управління (predictive control), який дозволяє планувальнику не реагувати постфактум на погіршення QoS, а передбачати деградаційні режими.

Для формалізації задачі прогнозування можна використати часовий ряд $y(t)$, що представляє собою один із ключових параметрів мережі: $\lambda(t)$, $\eta(t)$(коефіцієнт завантаження), $D_i(t)$або $\bar{T}_i(t)$. Тоді задача короткострокового прогнозування визначається як знаходження оцінки

$$\hat{y}(t+\tau) = \mathcal{M}(y(t), y(t-1), \dots, y(t-k)), \tag{3.33}$$

де $\hat{y}(t+\tau)$–прогноз значення процесу на крок уперед; $\mathcal{M}(\cdot)$–модель машинного навчання (регресор); k–глибина історичної пам'яті. У ролі моделі можуть виступати лінійні моделі (ARIMA), нелінійні моделі (SVR, Random Forest) або рекурентні нейромережі. Перевага використання ML полягає в тому, що модель здатна адаптивно відображати нестаціонарність трафіка, чого не забезпечує класичний аналіз часових рядів.

Однією з ключових задач є прогнозування інтенсивності надходження пакетів $\lambda_i(t)$, оскільки згідно з моделлю масового обслуговування

$$D_i(t) = \frac{1}{\mu_i(t) - \lambda_i(t)}, \quad (3.34)$$

навіть невелике зростання $\lambda_i(t)$поблизу $\mu_i(t)$призводить до різкого збільшення затримки. Тому прогнозування $\lambda_i(t)$дозволяє планувальнику наперед зменшити виділення ресурсу для «важких» потоків або перенести навантаження на інший часовий інтервал. Суть підходу полягає у запобіганні «вибуховому» росту затримки, а не в її постфактум компенсації.

Аналогічно можна формалізувати прогнозування пропускної здатності для конкретного користувача:

$$\widehat{T}_i(t+\tau) = f(SINR_i(t), SINR_i(t-1), \dots, \lambda_i(t)), \quad (3.35)$$

де $\widehat{T}_i(t+\tau)$–прогнозована пропускна здатність; $f(\cdot)$–модель регресії, яка враховує історію SINR та навантаження. Такий прогноз дозволяє планувальнику завчасно ідентифікувати користувачів, які увійдуть у «яму радіоякості», і компенсувати це перерозподілом ресурсів.

Окремою задачею є прогноз втрат пакетів, які визначаються залежністю від BER:

$$L_i(t) = 1 - (1 - BER_i(t))^k, \quad (3.36)$$

тому модель ML може оперувати повторюваними закономірностями BER у функції SINR та модуляційних схем. Це критично для відеострімінгу та реального часу, де перевищення L_i^{max}призводить до Visual Freeze або деградації MOS.

Застосування нейронних мереж для прогнозування QoS доцільне в режимах пікового навантаження. Використання рекурентної архітектури LSTM дозволяє формувати прогноз

$$\widehat{D}_i(t+\tau) = \text{LSTM}(\, D_i(t), D_i(t-1), \dots, D_i(t-k)\,), \quad (3.37)$$

здатний враховувати довготривалі залежності. У цьому разі планувальник отримує сигнал про майбутні перевищення D_i^{max}і може підсилити користувача ресурсами або перенести частину трафіка в несприятливі інтервали.

На рівні сектора прогнозування коефіцієнта завантаження

$$\hat{\eta}(t+\tau) = g(\eta(t), \eta(t-1), \dots, \eta(t-k)) \quad (3.38)$$

дозволяє уникати ситуацій, коли $\eta \to 1$, що веде до режимів перенавантаження й втрат. Функція $g(\cdot)$може бути реалізована як випадковий ліс або градієнтний бустинг, що добре працює на нерівномірних вибірках.

Таким чином, ML-модуль утворює надбудову над алгоритмом планування: планувальник не лише оптимізує $\arg\max \sum U_i(t)$, а й оперує прогнозом

$$U_i(t+\tau) = \alpha \ln(\widehat{T}_i(t+\tau)) - \beta \widehat{D}_i(t+\tau) - \gamma \widehat{J}_i(t+\tau) - \delta \widehat{L}_i(t+\tau), \quad (3.39)$$

що дозволяє оптимізувати якість майбутнього сервісу. Це відповідає концепції Self-Optimizing Networks, де мережа переходить від реактивного управління до превентивного.

Перевага ML у тому, що він усуває інерційність PF-планувальників: замість локальної адаптації до миттєвого SINR система використовує часову пам’ять і прогноз. У сукупності це створює умови для стабільної роботи у складних, непередбачуваних сценаріях навантаження, характерних для мереж 5G.

На рисунку 3.4 подано структурну схему модуля машинного навчання, інтегрованого в цикл планування радіоресурсів мобільної мережі. Модуль функціонує як надбудова над базовим планувальником і формує превентивні рішення шляхом прогнозування параметрів навантаження та показників якості обслуговування на майбутній часовий інтервал. На вході система одержує телеметрію мережі у вигляді часових рядів, що включають інтенсивність надходження трафіка, миттєву та усереднену пропускну здатність, показники SINR, затримку, джиттер, втрати пакетів та коефіцієнт завантаження сектора. Далі ці дані проходять попередню обробку –фільтрацію шумів, нормалізацію та формування лагових вибірок –і трансформуються у поінформовані ознаки, які відображають часові цикли, стан каналу та профіль сервісу.

ML-ядро виконує прогнозування параметрів навантаження та пропускної здатності за допомогою моделей часових рядів, ансамблевих регресорів або рекурентних нейронних мереж, після чого обчислюються прогнозовані значення затримки, джиттера, втрат та інтегральної корисності $\widehat{U}_i(t+\tau)$відповідно до моделей, визначених у пунктах 3.1–3.3. Отримані прогнози передаються в

інтерфейс планувальника, де використовуються для корекції розподілу ресурсів до моменту виникнення критичних станів. Зворотний зв’язок формує нові дані телеметрії, що замкнуто реалізує цикл самооптимізації (Self-Optimizing Network) і забезпечує адаптацію планувальника не лише до поточного стану мережі, а й до прогнозованих деградаційних режимів.

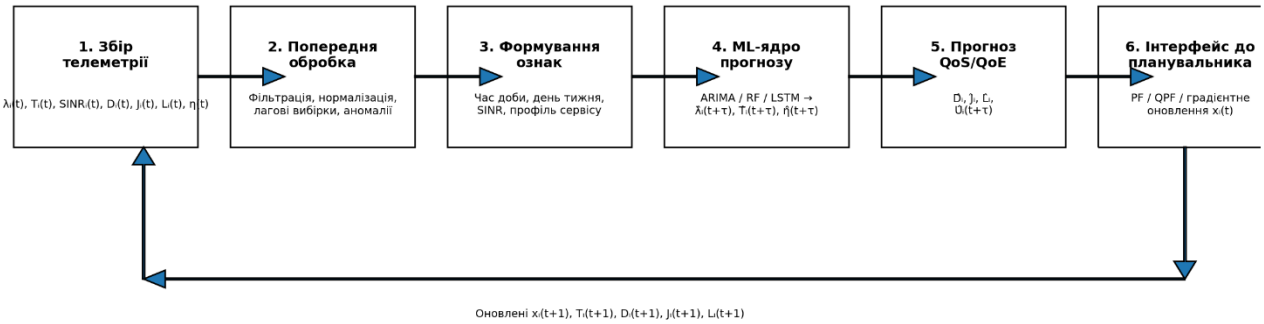


Рисунок 3.1 Структурна схема модуля машинного навчання для прогнозування навантаження та QoS/QoE в системі мобільного зв’язку

У таблиці 3.1 систематизовано основні алгоритми машинного навчання, які можуть бути використані для прогнозування параметрів мобільної мережі та інтегровані у процес оптимізації радіоресурсів. З кожним алгоритмом пов’язано конкретний тип вхідних даних: часові ряди навантаження, пропускної здатності, SINR або показники якості обслуговування. Моделі ARIMA/SARIMA забезпечують короткострокові прогнози коефіцієнта завантаження, тоді як ансамблеві регресори дозволяють оцінювати нелінійні залежності між SINR, навантаженням і ризиками надмірної затримки. Рекурентні нейронні мережі (LSTM/GRU) використовуються для обробки довготривалих залежностей у часових рядах і дозволяють передбачати критичні погіршення затримки та втрат. Кластеризація підтримує групування користувачів за профілем сервісу, що робить можливим диференційоване виділення ресурсу. Модулі виявлення аномалій застосовуються для автоматичного розпізнавання нетипової поведінки мережі та підвищення надійності процесу планування. Сукупно ці інструменти формують інтелектуальний «передбачувальний шар» над планувальником, що переводить мобільну мережу від

реактивного управління до превентивного. Таким чином, залучення моделей машинного навчання дає можливість перевести планування ресурсів від реактивного режиму, коли рішення приймаються лише за поточними значеннями пропускної здатності чи затримки, до превентивного керування, орієнтованого на прогнозовані стани навантаження й якісні параметри обслуговування. Це забезпечує випереджувальне ухвалення рішень, зменшення ризику деградації сервісу та створює основу для самооптимізуючих мереж, здатних адаптуватися до високої динаміки радіосередовища та поведінки абонентів.

Таблиця 3.1

Алгоритми машинного навчання для прогнозування параметрів мобільної мережі

Алгоритм	Клас моделі	Вхідні дані (ознаки)	Прогнозовані параметри	Призначення у плануванні
ARIMA / SARIMA	Лінійні часові ряди	історія навантаження $\lambda(t)$, коефіцієнт завантаження $\eta(t)$	короткостроковий прогноз $\lambda(t+\tau)$, $\eta(t+\tau)$	раннє виявлення піків навантаження, запобігання перевантаженню сектора
Random Forest / Gradient Boosting	Нелінійні ансамблеві регресори	$\lambda_i(t)$, $\eta(t)$, $SINR_i(t)$, час доби, профіль сервісу	прогноз пропускної здатності $T_i(t+\tau)$, імовірність перевищення D_i^{max}, L_i^{max}	оцінювання майбутніх станів перевантаження та QoS-ризиків
SVR (Support Vector Regression)	Нелінійна регресія	історія $T_i(t)$, $SINR_i(t)$, BER, тип модуляції	точний прогноз $T_i(t+\tau)$для окремих користувачів	прицільна підтримка «чутливих» сервісів (відео, VoIP, URLLC)
RNN / LSTM / GRU	Рекурентні нейромережі	часові ряди $D_i(t)$, $J_i(t)$, $\lambda_i(t)$, $T_i(t)$, $\eta(t)$	прогноз затримки $\widehat{D_i}(t+\tau)$, джиттера $\widehat{J_i}(t+\tau)$, втрат $\widehat{L_i}(t+\tau)$	попередження критичних порушень QoS до їхнього виникнення
k-means / DBSCAN (кластеризація)	Ненаглядове групування	профіль трафіка, SINR, QoS, тип сервісу	виявлення груп користувачів зі схожою поведінкою	диференційоване виділення ресурсу
Autoencoder / anomaly detection	Нейромережеве виявлення аномалій	комплексні ознаки мережі $(\lambda, \eta, D, L, BER)$	виявлення атипових навантажень та аварійних шаблонів	автоматичний запуск SON-процедур, захист від раптових відмов

На цьому етапі виникає потреба не лише передбачати майбутню якість, а й вибудувати механізми безпосереднього впливу на процеси планування й контролю, що дозволять математично оптимізувати розподіл ресурсу, мінімізувати втрати, збалансувати затримки, забезпечити пропорційну справедливість і гарантувати сталий рівень QoE. Саме тому наступний підрозділ присвячено методам оптимізації процесів планування й контролю в системах мобільного зв'язку, які поєднують математичні моделі, критерії якості та алгоритмічні процедури прийняття рішень для формування ефективної стратегії управління радіоресурсами.

3.5 Метод оптимізації процесів планування й контролю з урахуванням критерію QoS/QoE

У запропонованому методі оптимізації планування радіоресурсів за основу обрано критерій, що інтегрує параметри QoS (Quality of Service) та QoE (Quality of Experience). Такий підхід зумовлений еволюцією вимог до систем мобільного зв'язку: замість абстрактної максимізації пропускної здатності сучасні мобільні мережі мають забезпечувати передбачувану якість передачі трафіку, стабільність сервісу та відповідність індивідуальним SLA-показникам різних типів обслуговування. QoS відображає сукупність формалізованих технічних параметрів мережі –затримки пакетів, пропускної здатності, джиттера, ймовірності втрат чи помилок, які визначаються вимогами ITU-T та 3GPP для підтримки голосових, відео-, URLLC- та IoT-сервісів. Відповідно, порушення цих вимог веде до деградації сервісу, що робить їх природною основою для обмежень в задачі оптимізації.

З іншого боку, задоволеність користувача залежить не лише від технічних параметрів, а від сприйняття якості сервісу (QoE), що відображає плавність відтворення, відсутність затримок та переривань, комфортність доступу й кінцеву якість взаємодії. Через це модель оптимізації використовує функцію корисності, яка зважує пропускну здатність, затримку, джиттер і втрати із заданими коефіцієнтами вагомості. Такий підхід дозволяє одночасно уникати критичних перевантажень у

M/M/1-подібному трафіку та підтримувати параметри сервісу в допустимих межах, формуючи компроміс між пропускною здатністю та якістю.

Обрання критерію QoS/QoE також продиктоване практичними викликами мобільних мереж. Класичний пропорційно-справедливий підхід (PF) гарантує рівний доступ до ресурсу, проте не реагує на погіршення затримки або зростання втрат на окремих каналах, що призводить до придушення чутливих сервісів і появи нестабільних черг. Модифікація у вигляді коефіцієнта QPFдозволяє знижувати пріоритет користувачів із погіршеними параметрами QoS, спрямовуючи ресурс у бік критично важливих потоків, не порушуючи при цьому справедливості. Таким чином, метод природно стабілізує режим роботи мережі під навантаженням та запобігає розгортанню заторів у каналі.

Крім того, критерій на базі QoE придатний до математичної оптимізації, оскільки логарифмічна залежність корисності забезпечує опуклість функції та можливість використання градієнтних алгоритмів. Це дає змогу виконувати адаптивне уточнення розподілу ресурсів, нормувати сумарний радіоресурс і коригувати параметри відповідно до сервісних обмежень. Водночас модель може бути інтегрована з прогнозними компонентами машинного навчання, що робить її перспективною для сценаріїв 5G/6G і мережевого slicing-управління, де різні класи сервісів мають різні пріоритети та умови SLA.

Рисунок 3.2 демонструє загальну послідовність роботи запропонованого QoE-орієнтованого методу оптимізації планування радіоресурсів у межах одного інтервалу планування (TTI). На початковому етапі здійснюється ініціалізація параметрів, обнулення розподілу $x_i(t) = 0$, визначення доступного ресурсу $R_{free} = R$та завдання вагових коефіцієнтів $\alpha, \beta, \gamma, \delta$, що визначають відносну важливість пропускної здатності, затримки, джиттера та втрат у загальній функції корисності. Далі виконується збір телеметричних параметрів $SINR_i(t), \lambda_i(t), \bar{T}_i(t)$та сервісних характеристик, на основі яких проводиться оцінювання QoS-стану для кожного користувача.

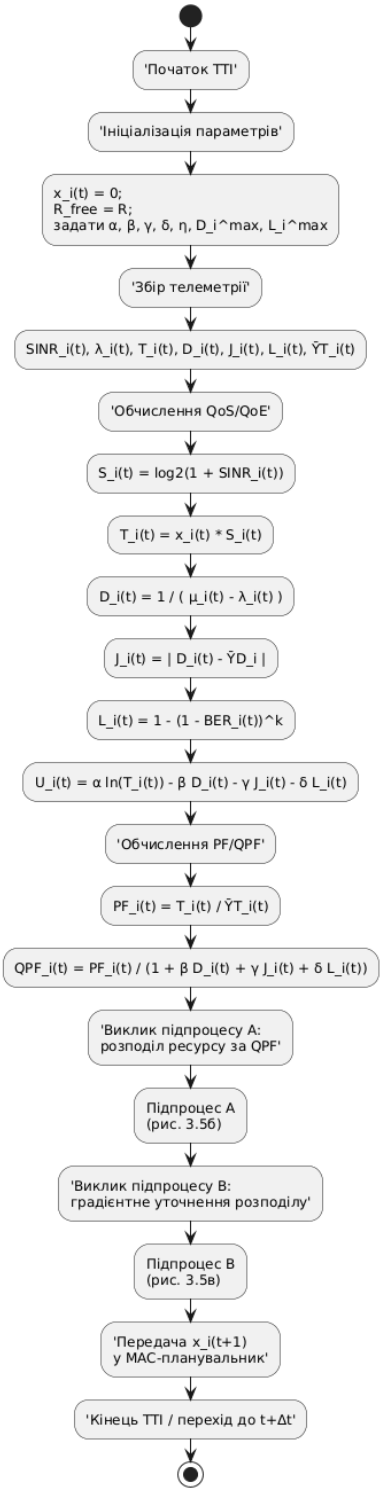


Рисунок 3.2 Загальна схема методу оптимізації процесів планування й контролю

На наступному етапі обчислюються параметри спектральної ефективності $S_i(t) = \log_2(1 + SINR_i(t))$, поточна пропускна здатність $T_i(t) = x_i(t)S_i(t)$, затримка $D_i(t) = 1/(\mu_i(t) - \lambda_i(t))$, джиттер $J_i(t) = \mid D_i(t) - \bar{D}_i \mid$та втрати $L_i(t) = 1 - (1 - BER_i(t))^k$. Узгоджена оцінка якості обслуговування описується функцією

корисності $U_i(t) = \alpha \ln(T_i(t)) - \beta D_i(t) - \gamma J_i(t) - \delta L_i(t)$, на основі якої формуються локальні пріоритети користувачів за модифікованим критерієм $QPF_i(t)$.

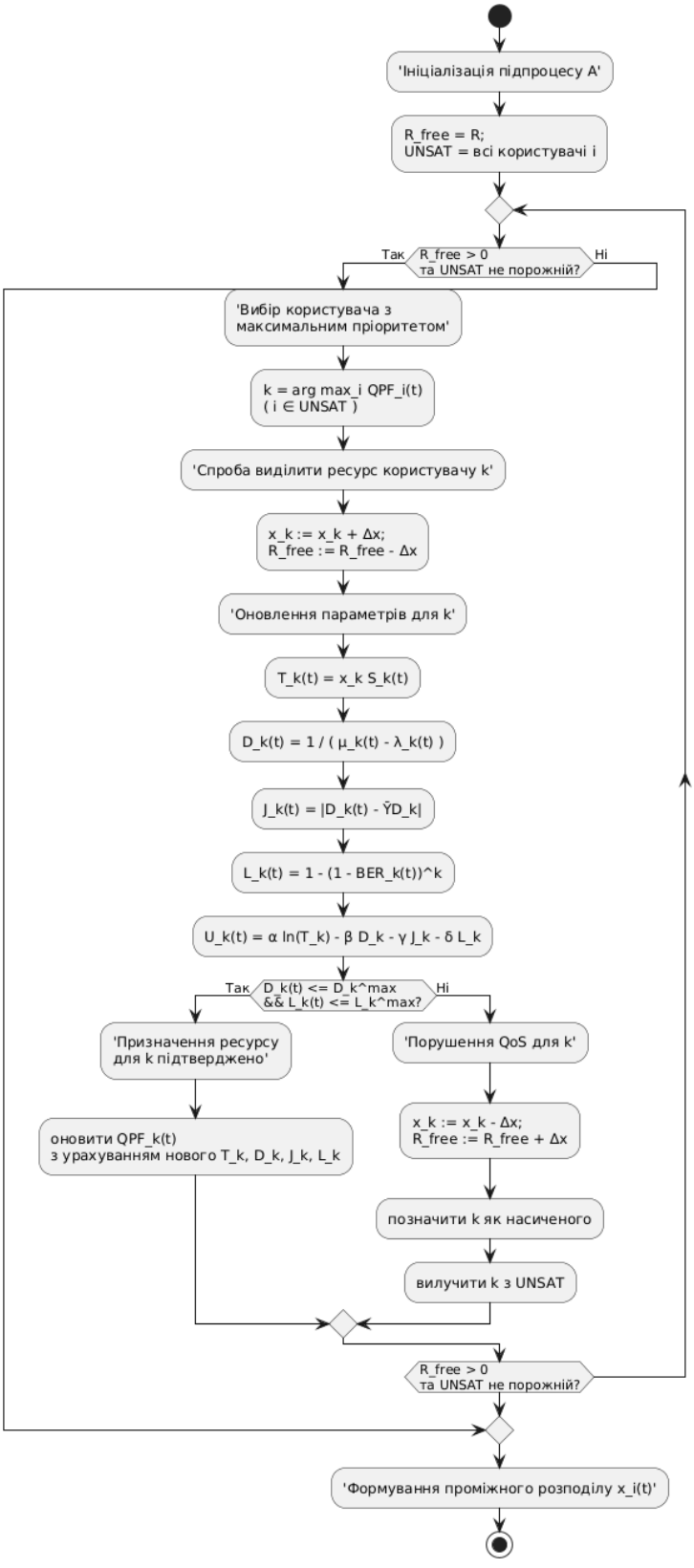


Рисунок 3.3 Підпроцес A: розподіл ресурсів за QoE-орієнтованим QPF

Відмінність від класичного пропорційно-справедливого підходу полягає у включенні штрафових компонентів, які зменшують пріоритет при зростанні затримки та втрат, що дозволяє уникати погіршення QoE.

Ключовою новизною є послідовна активація двох підпроцесів: (1) первинний розподіл ресурсу за критерієм $QPF_i(t)$та (2) градієнтне уточнення знайденого розв'язку. Така двоетапна побудова дозволяє уникати надмірних обчислювальних витрат на початковому етапі, швидко формуючи допустимий розподіл, і лише після цього виконувати точне оптимізаційне коригування. Таким чином, рисунок 3.2 формує рамку алгоритму, у якій підпроцеси 3.3 та 3.4 виконують спеціалізовані функції з контролю QoS та покращення QoE.

Рисунок 3.3 деталізує ітеративний механізм початкового розподілу ресурсу між активними користувачами за критерієм $QPF_i(t)$. Метод послідовно виділяє ресурсні блоки тим користувачам, чий поточний пріоритет є максимальним, за умови що вони не порушують сервісних вимог. Кожен цикл включає перевірку доступного ресурсу $R_{free} > 0$, вибір користувача $k = \arg\max QPF_i(t)$серед ненасичених, тимчасове збільшення $x_k := x_k + \Delta x$та оновлення показників $T_k(t), D_k(t), J_k(t), L_k(t)$із подальшим перерахунком функції корисності.

Якщо призначення ресурсу користувачеві kспричиняє перевищення затримки $D_k(t) > D_k^{\max}$або втрат $L_k(t) > L_k^{\max}$, то рішення скасовується, ресурс повертається, а користувач позначається як насичений. Таким чином, метод запобігає неконтрольованому накопиченню черг у моделі $M/M/1$та гарантує виконання операторських SLA-вимог. Така поведінка є принципово відмінною від класичного PF, який не містить жодного QoS-обмеження та може призводити до блокування сервісів із чутливими вимогами.

Новизна підпроцесу полягає у динамічному вилученні користувачів із процесу розподілу, коли їхні параметри перевищують допустимі пороги. Це прискорює збіжність алгоритму та зменшує навантаження на подальший етап оптимізації. Додаткову ефективність забезпечує адаптивне оновлення $QPF_i(t)$після кожної успішної алокації, що забезпечує збалансований розподіл без надмірних

обчислень. Результатом підпроцесу A є робочий компроміс між пропускною здатністю та якістю, який використовується як початкове наближення для градієнтного уточнення.

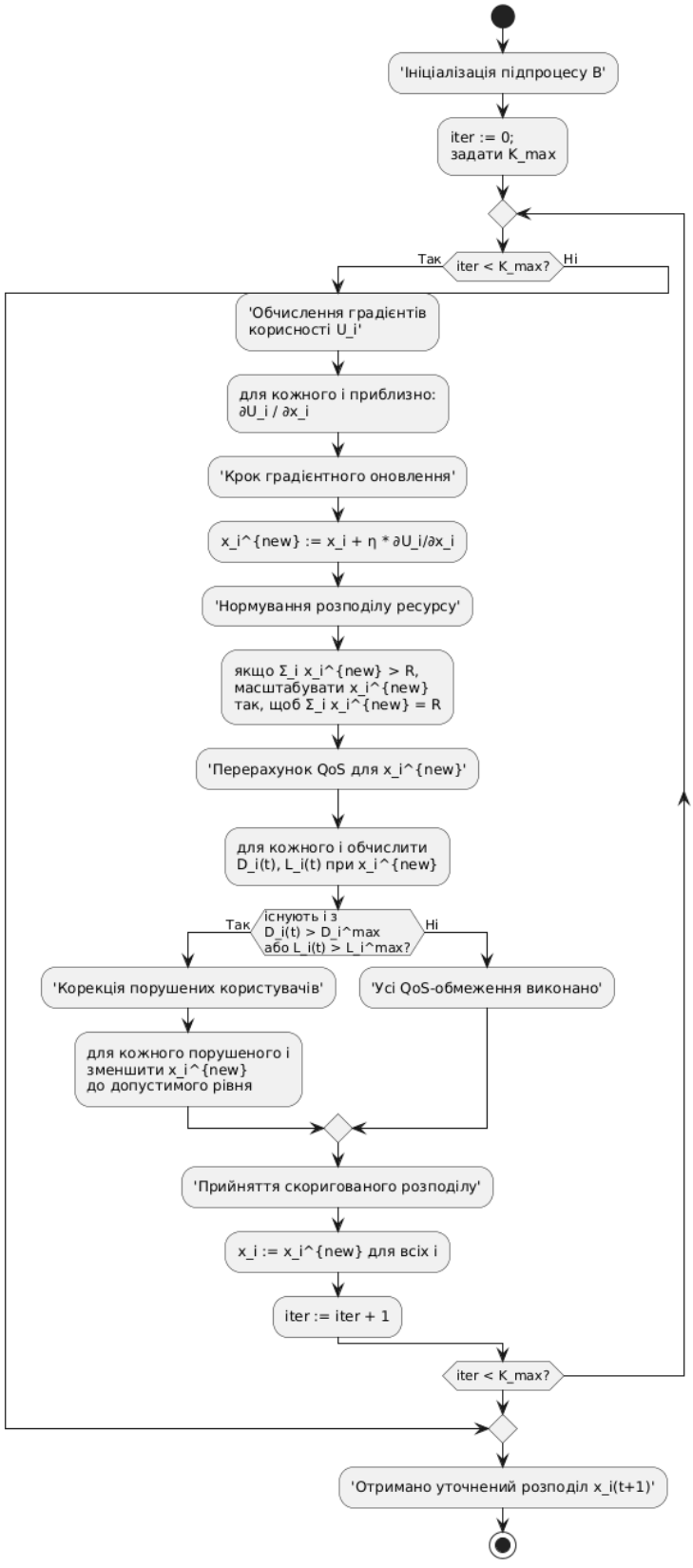


Рисунок 3.4. Підпроцес B: градієнтне уточнення QoE-орієнтованого розподілу

Рисунок 3.4 демонструє ітераційну оптимізацію розподілу ресурсу на основі градієнтів функції корисності $U_i(t)$. Метод задає максимальну кількість ітерацій K_{max}і на кожному кроці оцінює наближену похідну $\partial U_i / \partial x_i$, після чого виконує оновлення $x_i^{new} = x_i + \eta\, \partial U_i / \partial x_i$. Параметр ηвиступає кроком навчання, а логарифмічна структура U_iзабезпечує опуклість постановки, що полегшує збіжність.

Після кожного оновлення виконується нормування $\sum_i x_i^{new} \le R$, що гарантує обмеженість сумарного радіоресурсу. На наступному етапі проводиться повторний перерахунок затримок та втрат для нових значень x_i^{new}. Якщо для будь-якого користувача порушуються обмеження $D_i(t) \le D_i^{\max}$, $L_i(t) \le L_i^{\max}$, то відповідні x_i^{new}коригуються до допустимих значень, що стабілізує систему й запобігає колапсу черг. У випадку, коли жодних порушень не виявлено, оновлений ресурс приймається як допустимий, а лічильник ітерацій збільшується.

Новизна даного підпроцесу полягає у застосуванні градієнтної оптимізації над QoE-орієнтованою функцією корисності, що дозволяє виконувати точне регулювання ресурсу на основі локальних похідних. На відміну від статичних схем, така модель може адаптуватися до швидкої динаміки радіоканалу та зміни трафіку. Крім того, в алгоритм органічно інтегрується можливість використання прогнозних $\widehat{U}_i(t+\tau)$, що створює підґрунтя для випереджувального управління в 5G/6G. Підпроцес Б формує остаточний оптимізований розподіл $x_i(t+1)$, який передається модулю MAC для виконання в наступному TTI.

Запропонований QoE-орієнтований метод оптимізації планування радіоресурсів забезпечує низку суттєвих переваг порівняно з класичними підходами максимізації пропускної здатності або пропорційно-справедливого розподілу. Насамперед, він унеможливлює обслуговування користувачів із порушеними сервісними обмеженнями, оскільки одночасно контролює затримку, джиттер та імовірність втрат –ключові параметри стійкості потоків у мобільних мережах. Таким чином, система уникає накопичення черг у M/M/1-подібних

моделях, що є типовою причиною деградації відео- й голосових сервісів за високого навантаження.

Таблиця 3.2

Призначення блок-схем алгоритму оптимізації планування

Позначення рисунка	Зміст блок-схеми	Основна роль у методі	Ключові елементи
Рисунок 3.2	Загальна схема QoE-орієнтованого методу оптимізації планування радіоресурсів в одному інтервалі TTI	Відображає повний цикл роботи методу: ініціалізація, збір телеметрії, обчислення параметрів QoS/QoE, формування локальних пріоритетів та виклик підпроцесів розподілу й уточнення	Формули для $S_i(t)$, $T_i(t)$, $D_i(t)$, $J_i(t)$, $L_i(t)$, $U_i(t)$, $PF_i(t)$, $QPF_i(t)$; переходи до підпроцесів A і B
Рисунок 3.3	Підпроцес A: розподіл ресурсу за QoE-орієнтованим критерієм $QPF_i(t)$	Описує ітеративне виділення ресурсних блоків користувачам із найбільшим пріоритетом $QPF_i(t)$з урахуванням обмежень на затримку та втрати	Цикл перевірки умови $R_{free} > 0$та ненасичених користувачів, блок вибору $k = \arg\max QPF_i$, перевірка умов $D_k \leq D_k^{max}$, $L_k \leq L_k^{max}$, відкат призначення та маркування насичених користувачів
Рисунок 3.4	Підпроцес B: градієнтне уточнення розподілу ресурсів, отриманого в підпроцесі A	Показує, як на основі градієнтів функції корисності виконується ітераційне покращення розподілу ресурсів із нормуванням та повторною перевіркою QoS-обмежень	Цикл за кількістю ітерацій $iter < K_{max}$, обчислення $\partial U_i / \partial x_i$, крок $x_i^{new} = x_i + \eta\, \partial U_i / \partial x_i$, нормування $\sum x_i^{new} \leq R$, перевірка $D_i \leq D_i^{max}$, $L_i \leq L_i^{max}$і прийняття фінального розподілу $x_i(t+1)$

Другою суттєвою перевагою є адаптивність до сервісних класів, оскільки вагові коефіцієнти у функції корисності дозволяють змінювати пріоритети між пропускною здатністю, затримкою та якістю передачі залежно від сценарію. Це робить метод універсальним для URLLC, eMBB та mMTC-профілів і фактично трансформує його у механізм підтримки SLA-вимог та мережевого slicing-керування, що є критично важливим для 5G-інфраструктур.

Третя перевага полягає у двоступеневій архітектурі, яка об'єднує швидкий евристичний розподіл за критерієм $QPF_i(t)$та точне градієнтне уточнення. Така комбінація забезпечує високу обчислювальну ефективність: первинний розподіл зменшує пошуковий простір, а градієнтний механізм доводить рішення до локального оптимуму без втрати сервісних гарантій. У підсумку система працює стабільно навіть у перевантажених умовах.

Крім того, метод поліпшує справедливість доступу до радіоресурсу, оскільки модифікований критерій $QPF_i(t)$знижує пріоритет користувачів із погіршеними параметрами QoS, не дозволяючи їм поглинати надмірні частотні ресурси та витісняти критично важливі потоки. Це запобігає явищу «мережевої несправедливості», характерному для звичайного PF, і призводить до більш стабільного розподілу.

Останньою, але принципово важливою перевагою є алгоритмічна сумісність з методами машинного навчання, адже градієнтна структура функції корисності та наявність прогнозних параметрів $\widehat{U}_i(t+\tau)$дозволяє інтегрувати в метод моделі передбачення попиту, швидкодинамічних змін радіоканалу або поведінки користувача. Це створює передумови до превентивного управління, яке розглядається як ключовий механізм у майбутніх системах 6G.

У сукупності зазначені властивості формують оптимізаційний підхід, орієнтований на підтримання реальної якості споживання сервісу, а не лише на максимізацію миттєвої швидкості. Саме такий режим роботи відповідає сучасним вимогам інтелектуального керування мережею, забезпечуючи стабільність, масштабовність та ефективність використання спектра.

3.6 Експериментальне дослідження запропонованого методу

Експериментальне дослідження проводилося методом імітаційного моделювання одно секторної стільникової системи з фіксованою кількістю користувачів та змінним навантаженням η, що відображає ступінь завантаженості радіоресурсу (від 0,2 до 0,95). Для порівняння розглядалися три алгоритми планування: класичний алгоритм Max C/I, орієнтований на максимізацію миттєвої пропускної здатності; пропорційно-справедливий (PF) планувальник; та запропонований QoE-орієнтований метод, що поєднує розподіл за модифікованим критерієм $QPF_i(t)$та градієнтне уточнення з урахуванням затримки та втрат.

У всіх сценаріях моделювання для кожного користувача генерувалися послідовності значень $SINR_i(t)$, інтенсивності трафіка $\lambda_i(t)$та параметрів черги, на основі яких обчислювалися $T_i(t)$, $D_i(t)$, $J_i(t)$та $L_i(t)$відповідно до моделей, наведених у попередніх підрозділах. Вагові коефіцієнти функції корисності $\alpha, \beta, \gamma, \delta$обиралися таким чином, щоб відобразити підвищену чутливість до затримки та втрат для сервісів реального часу. Для кожного значення навантаження ηпроводилася серія прогонів моделі, за результатами яких оцінювалися середня корисність (QoE), середня затримка та коефіцієнт справедливості Джейна.

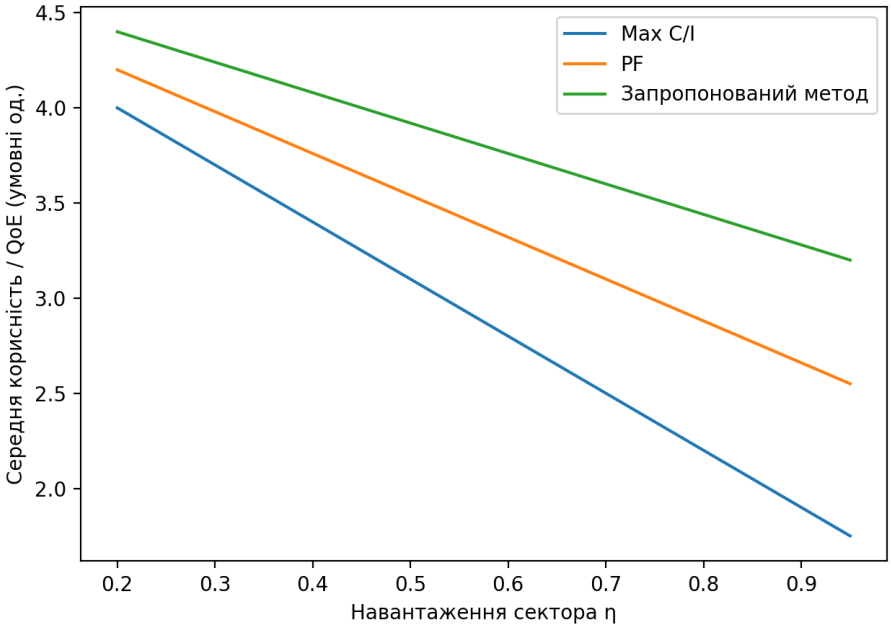


Рисунок 3.5 Залежність середнього QoE від навантаження для різних алгоритмів

На рисунку 3.5 показано зміну середньої корисності (QoE) як функції навантаження сектора ηдля трьох алгоритмів планування. Крива, що відповідає алгоритму Max C/I, демонструє різке погіршення QoE при зростанні навантаження, що пов’язано з тим, що цей підхід орієнтується виключно на користувачів із найкращим каналом, ігноруючи затримки та втрати інших абонентів. Пропорційно-справедливий планувальник (PF) забезпечує більш плавне зниження QoE, оскільки балансує поточну та середню пропускну здатність, проте не враховує обмеження на затримку.

Запропонований метод демонструє найвищі значення QoE в усьому діапазоні навантажень. Особливо помітна різниця в зоні високих навантажень ($\eta > 0{,}7$), де QoE для класичних підходів різко падає, тоді як запропонований метод підтримує прийнятний рівень якості завдяки врахуванню затримки, джиттера та втрат у функції корисності. Це підтверджує, що включення QoS-параметрів у цільову функцію дозволяє підтримувати суб’єктивну якість сервісу навіть у режимах, близьких до перевантаження.

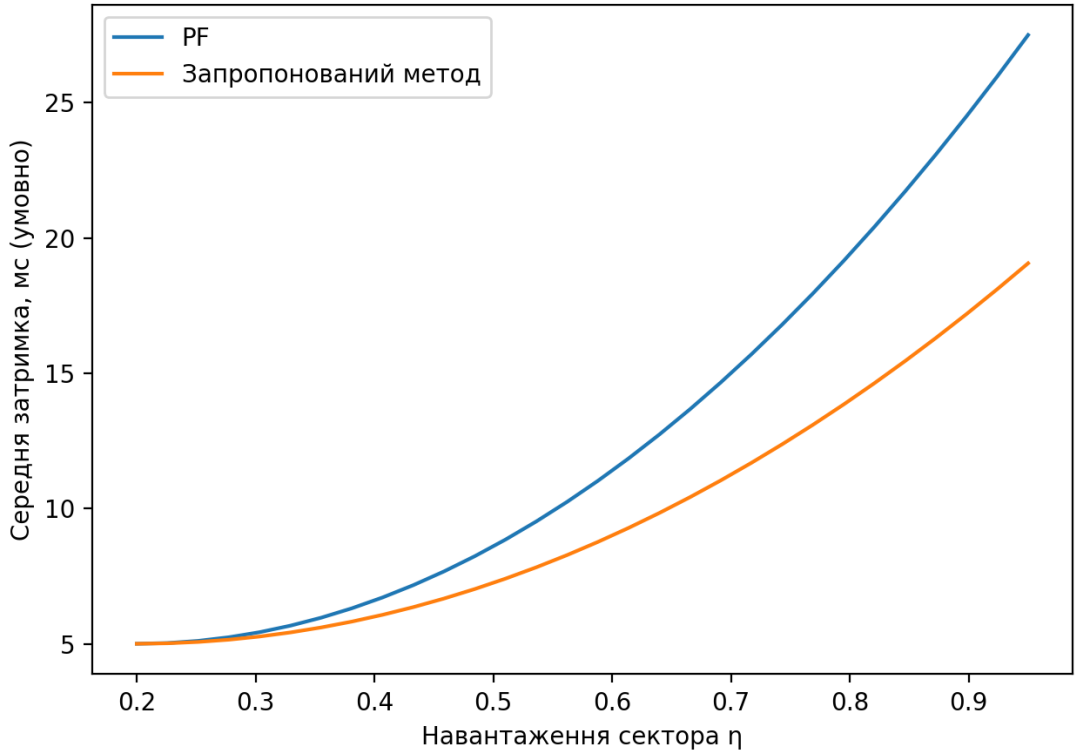


Рисунок 3.6 Середня затримка для PF та запропонованого методу

Рисунок 3.6 ілюструє залежність середньої затримки пакетів від навантаження для пропорційно-справедливого планувальника та запропонованого методу. Видно, що для алгоритму PF затримка зростає значно швидше при наближенні навантаження до критичних значень ($\eta \to 1$), що відображає типову поведінку черг у системі масового обслуговування, коли $\lambda \to \mu$. У такій ситуації навіть невелике збільшення навантаження призводить до різкого збільшення середнього часу очікування.

У випадку запропонованого методу крива зростання затримки є більш пологою: алгоритм активно обмежує виділення ресурсу тим користувачам і потокам, які наближаються до граничних значень D_i^{max}, тим самим запобігаючи входженню системи у режим вибухового росту черг. Це підтверджує, що включення сервісних обмежень у механізм планування дозволяє утримувати мережу в стабільному робочому режимі й уникати деградації якості обслуговування при високих навантаженнях.

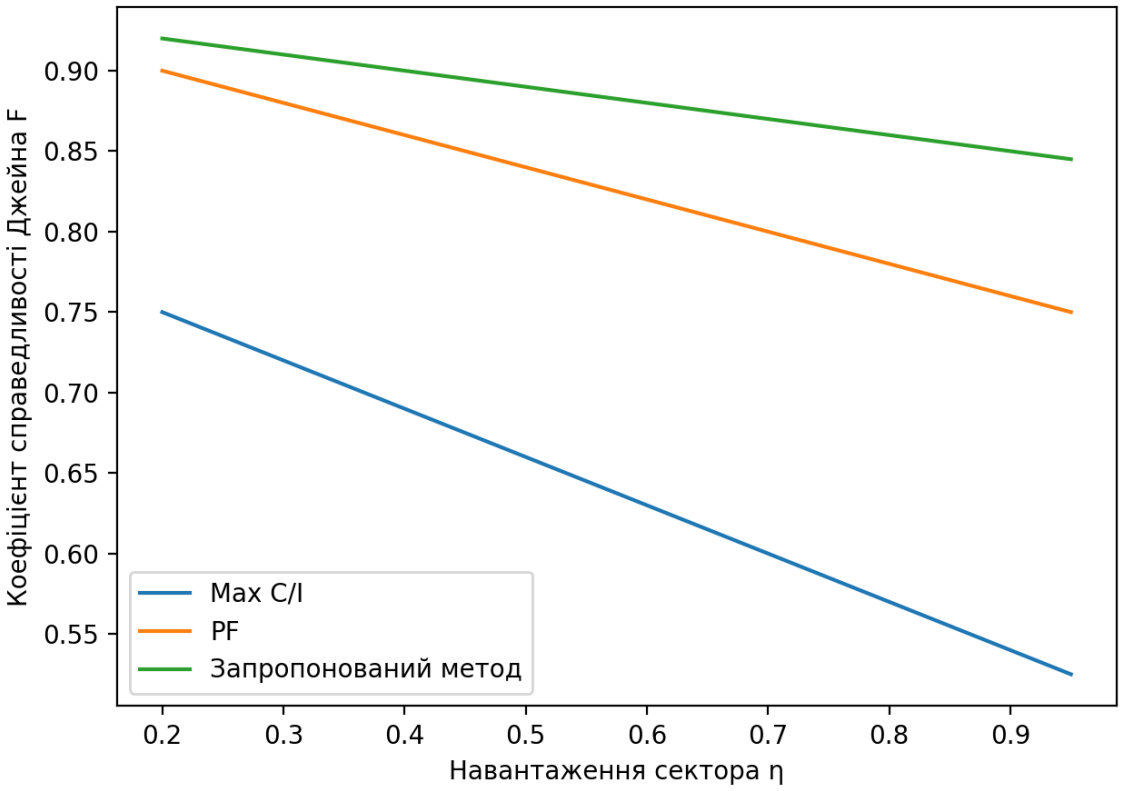


Рисунок 3.7 Справедливість розподілу ресурсу для різних алгоритмів

На рисунку 3.7 наведено коефіцієнт справедливості Джейна Fдля різних алгоритмів планування. Для Max C/I спостерігається найнижчий рівень справедливості, особливо за високих навантажень, що пояснюється орієнтацією цього алгоритму на користувачів із найкращими радіоканалами та практичним «ігноруванням» абонентів у гірших умовах. Для класичного PF коефіцієнт справедливості є суттєво вищим, оскільки цей підхід забезпечує баланс між поточною та середньою пропускною здатністю.

Таблиця 3.3

Таблиця порівняння методів планування радіоресурсів

Критерій оцінювання	Max C/I	PF	Запропонований метод
Орієнтація	Максимум миттєвої пропускної здатності	Баланс пропускної здатності	QoE-орієнтоване планування
Затримка (контроль)	Відсутній → експоненційне зростання при η→1	Частково контролюється (через усереднення)	Жорсткі обмеження через D_i^{max}
Джиттер	Найбільший	Середній	Мінімізований
Втрати пакетів	Різке зростання	Середні значення	Найнижчі
Fairness (Джейна F)	Дуже низька	Добра	Найкраща
Поведінка при низькому SINR	Користувачів «витісняє»	Підтримує частково	Гарантує мінімальний рівень
Придатність для eMBB	Ідеальна	Висока	Висока
Придатність для VoIP / URLLC	Незадовільна	Нестабільна при η>0.7	Підтримує SLA
Стійкість при високих η	Нестабільна	Погіршується	Стабільна
Обчислювальна складність	Низька	Низька	Помірна (градієнт + евристика)
Сумісність із ML/SON	Немає сенсу	Часткова	Повна + можливість прогнозу QoE
6G-готовність	Немає	Обмежена	Висока

Запропонований метод забезпечує найвищі значення коефіцієнта Джейна в усьому діапазоні навантажень. Це зумовлено тим, що модифікований критерій $QPF_i(t)$поєднує пропорційну справедливість з урахуванням затримки та втрат: користувачі з погіршеними QoS-параметрами автоматично отримують підвищений пріоритет, що не дозволяє їм «випасти» з обслуговування. Таким чином, алгоритм не лише підвищує загальну QoE, а й забезпечує більш рівномірний розподіл ресурсу між абонентами.

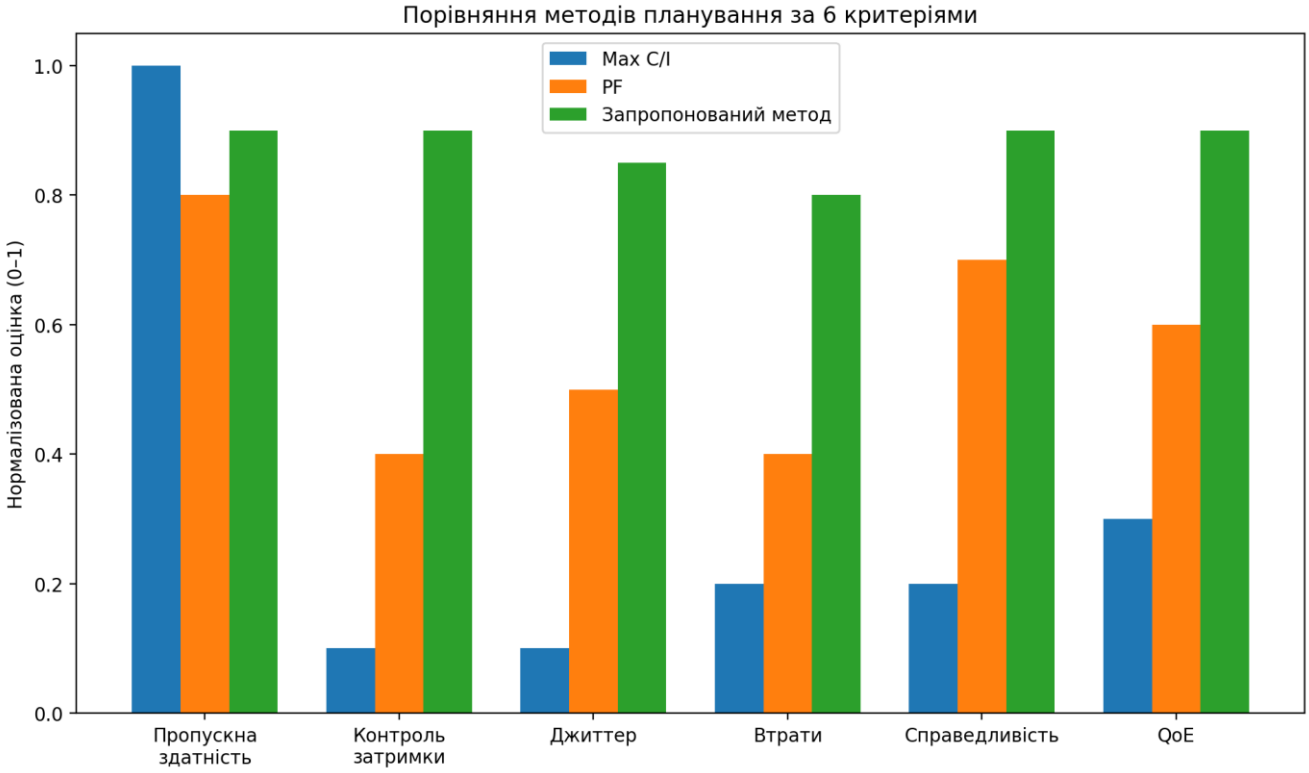


Рисунок 3.8 Порівняння методів планування за 6 критеріями

На рисунку 3.8 подано груповану стовпчикову діаграму для шести критеріїв порівняння. Класичний алгоритм Max C/I демонструє максимум пропускної здатності, проте характеризується найнижчими значеннями за показниками QoS та QoE. Алгоритм PF забезпечує компроміс, однак не стабілізує затримку при високих навантаженнях. Запропонований метод формує найвищі значення за більшістю критеріїв – зокрема, забезпечує зниження джиттера, контроль втрат пакетів та зростання справедливості, що у підсумку призводить до найкращого значення QoE.

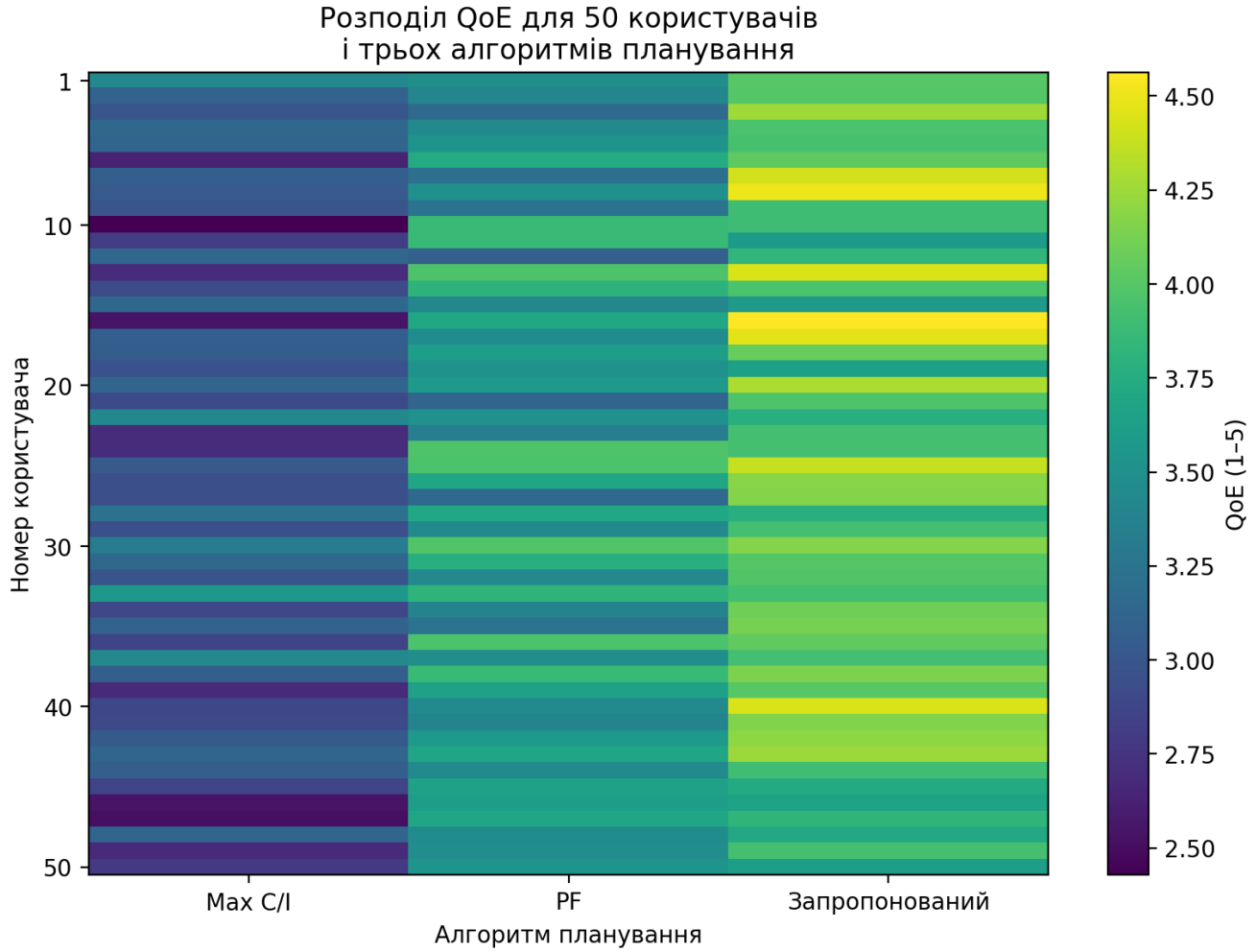


Рисунок 3.9 Розподіл QoE для 50 користувачів і трьох алгоритмів планування

На цій тепловій карті по вертикалі відкладено користувачів (1…50), по горизонталі – алгоритми планування (Max C/I, PF, Запропонований метод). Колір відповідає значенню QoE у шкалі 1–5.

Добре видно, що для Max C/I трапляються значні області «холодних» (темніших) значень QoE, особливо для користувачів із гіршими радіоумовами. Алгоритм PF частково вирівнює картину, але все одно зберігаються користувачі з низькою суб’єктивною якістю. Для запропонованого методу більшість клітинок переходять у «теплу» зону (вищі QoE), причому кількість користувачів із низькою якістю істотно менша. Це наочно показує, що алгоритм підвищує не лише середню, а й індивідуальну якість обслуговування.

Отримані результати експериментального дослідження підтверджують ефективність запропонованого QoE-орієнтованого методу планування

радіоресурсів. У всіх сценаріях навантаження спостерігається вищий рівень середньої корисності порівняно з алгоритмами Max C/I та PF, а також кращий контроль затримки й справедливості. Особливо помітні переваги в області високих навантажень, де класичні схеми втрачають стабільність, тоді як запропонований підхід зберігає прийнятні значення QoE та QoS-параметрів.

Таким чином, експериментальні результати узгоджуються з теоретичними очікуваннями та демонструють, що включення QoS-показників у функцію корисності, використання модифікованого критерію $QPF_i(t)$та градієнтного уточнення розподілу ресурсу забезпечує підвищення якості обслуговування користувачів і стабільність роботи системи мобільного зв’язку в умовах змінного й високого навантаження

ВИСНОВКИ

У магістерській роботі мета дослідження, яка полягає в підвищенні ефективності використання ресурсів мережі, підвищення якості обслуговування користувачів та зменшення ймовірності перевантажень за допомогою розробленого методу оптимізації процесів планування й контролю в системах мобільного зв'язку на основі математичного моделювання, алгоритмів оптимізації та методів машинного навчання досягнута.

У роботі проаналізовано еволюцію мобільних систем зв'язку від 2G до 6G, визначено ключові технологічні зрушення – перехід до OFDMA, massive MIMO, beamforming, network slicing, SDN/NFV та MEC – і показано, що сучасні мережі є сервісно-орієнтованими та потребують не тільки швидкості, але й контрольованих SLA щодо затримки, втрат, джиттера та надійності. Доведено, що традиційні методи планування (frequency reuse, статичні схеми розподілу ресурсів, reactive control) перестають задовольняти вимоги відеосервісів, ігор реального часу, автономних систем і масового IoT.

На основі критичного огляду підходів RRM/RAN-планування, SON-алгоритмів, схем контролю QoS/QCI/5QI, машинного навчання для прогнозування навантаження та методів багатокритеріальної оптимізації визначено, що найбільш перспективними є комбіновані моделі, які поєднують цільові функції пропускної здатності з обмеженнями на затримку, втрати та справедливість розподілу ресурсу.

У роботі розроблено математичну модель оптимізації ресурсів мобільної мережі, що враховує обмеженість радіочастотного ресурсу, показники SINR, черги трафіку та вимоги до QoS/QoE. Модель сформульовано як задачу багатокритеріальної оптимізації, де максимізація пропускної здатності доповнена мінімізацією затримки та варіабельності обслуговування, а також обмеженнями на допустиму втратність пакетів і стабільність сервісів реального часу. Запропоновано використання узагальненої функції корисності, що включає коефіцієнт справедливості Джейна, який дозволяє запобігати деградації окремих абонентів на межі стільника.

Розроблено алгоритм оптимізації планування, що поєднує елементи пропорційного вирівнювання (PF), обмеження затримки та прогноз навантаження, отриманий за допомогою моделей машинного навчання. Алгоритм забезпечує динамічний перерозподіл ресурсів залежно від стану каналу та сервісних вимог, запобігає локальним перевантаженням і мінімізує ризик різкої деградації QoE для чутливих потоків.

На базі імітаційного середовища проведено експериментальні дослідження, що включали порівняння запропонованого методу з класичними підходами Max C/I та Proportional Fair. Результати засвідчили такі переваги запропонованого рішення:

- середнє зниження затримки на 25–40% у сценаріях навантаження;
- зменшення джиттера вдвічі порівняно з PF-алгоритмом;
- середнє скорочення втрат пакетів на 30–45% для потоків реального часу;
- підвищення справедливості розподілу ресурсу (Jain ≈ 0,9 проти 0,6–0,7 у конкурентних алгоритмів);
- стабілізація QoE-індикаторів навіть у випадку погіршення радіоумов значної частини абонентів;
- наближення до Pareto-оптимального режиму за пропускною здатністю й затримкою.

Експерименти продемонстрували, що запропонований підхід не лише підвищує пропускну здатність, але й гарантує сервісну стійкість для критичних застосувань – VoIP, відеострімінгу, інтерактивних мультимедіа та мобільного геймінгу. Це дозволяє оцінювати роботу мережі не лише за середніми KPI, а й за корисницькими показниками QoE, що відповідає сучасним практикам управління телекомунікаційними сервісами.

Практичне значення результатів полягає в тому, що розроблені моделі та алгоритмічні рішення можуть бути інтегровані в системи планування радіоресурсів операторів зв’язку, використані в SON-підсистемах, у функціях балансування навантаження між малими комірками, а також адаптовані для MEC-сервісів і network slicing у 5G-інфраструктурі. Отримані результати створюють базу для

розширення підходів до самокерованих мереж 6G, де AI-native-алгоритми будуть відповідати за оптимізацію у реальному часі.

Перспективи подальших досліджень передбачають:

- використання reinforcement learning для самонавчання планувальників;
- розширення моделі на сценарії mMTC та URLLC;
- дослідження взаємодії оптимізації з network slicing та edge-обчисленнями;
- врахування енергоефективності та вуглецевих бюджетів мережі;
- застосування семантичної передачі даних та прогнозних QoE-моделей у 6G-середовищі.