

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «КИЇВСЬКИЙ АВІАЦІЙНИЙ
ІНСТИТУТ»**

Кваліфікаційна наукова праця
на правах рукопису

ОДАРЧЕНКО МИХАЙЛО СЕРГІЙОВИЧ

УДК 629.7.058(043.3)

**МЕТОДИ ОЦІНКИ ТА ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ НАДАННЯ
СЕРВІСУ КОРОТКИХ ПОВІДОМЛЕНЬ (SMS) В МЕРЕЖАХ
НАСТУПНОГО ПОКОЛІННЯ**

Спеціальність 172 «Телекомунікації та радіотехніка»

Галузь знань 17 «Електроніка та телекомунікації»

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ М. С. Одарченко

Науковий керівник:

Заліський Максим Юрійович,

доктор технічних наук, професор, професор кафедри
телекомунікаційних та радіоелектронних систем ДУ «КАІ»

Київ – 2026

АНОТАЦІЯ

Одарченко Михайло Сергійович. Методи оцінки та підвищення ефективності надання сервісу коротких повідомлень (SMS) в мережах наступного покоління – кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття ступеня доктора філософії за спеціальністю 172 «Телекомунікації та радіотехніка» галузі знань 17 «Електроніка та телекомунікації». Державний університет «Київській авіаційний інститут» Міністерства освіти і науки України. – Київ, 2026.

Дисертаційна робота присвячена дослідженню та вирішенню актуальної науково-практичної задачі, що полягає у підвищенні ефективності доставки сервісу коротких повідомлень (SMS) у мережах 4G/5G шляхом інтегрованого врахування технічних, користувацьких та економічних факторів, а також мінімізації впливу штучно згенерованого трафіку (Artificially Inflated Traffic, AIT) на показники якості та вартості доставки.

Метою дисертаційної роботи є розробка методів і моделей кількісної оцінки та підвищення ефективності доставки SMS у мережах наступного покоління на основі інтеграції показників QoS, QoE та економічної ефективності, формалізованої через показник Price Delivery Gap (PDG).

Для досягнення поставленої мети в дисертаційній роботі вирішуються наступні завдання:

- проаналізувати еволюцію архітектури та технологій SMS у мережах 2G–5G та визначити чинники, що впливають на ефективність доставки;
- систематизувати сценарії використання SMS у мережах 4G/5G та їх вплив на параметри якості й вартості;
- обґрунтувати показник PDG як інтегральну метрику техніко-економічної ефективності доставки повідомлень;

- розробити інтегровану модель IGDP (Integrated Gap-Delivery-Performance), що поєднує QoS, QoE та PDG;
- побудувати математичні та регресійні моделі оцінювання ефективності доставки SMS в умовах гетероскедастичності;
- розробити методи виявлення штучно згенерованого трафіку (AIT) у A2P-SMS;
- виконати сценарну оцінку ефективності інтегрованого фреймворку управління доставкою SMS.

У першому розділі проведено аналіз теоретичних засад, архітектурних особливостей та еволюції сервісу SMS у мережах 4G/5G, систематизовано сценарії використання повідомлень та сформульовано наукову проблему, мету, завдання і гіпотезу дослідження.

У другому розділі наведено методологічні основи оцінювання ефективності доставки SMS, обґрунтовано показник PDG та розроблено інтегровану модель IGDP, що поєднує технічні, користувацькі та економічні параметри ефективності.

У третьому розділі наведено математичні та регресійні моделі процесу доставки SMS, виконано аналіз впливу параметрів QoS і QoE на PDG, а також досліджено гетероскедастичність у даних доставки повідомлень.

Четвертий розділ присвячено інтелектуальним методам підвищення ефективності доставки SMS, зокрема категоризації повідомлень, виявленню AIT, розробці інтегрованого фреймворку управління доставкою та сценарній оцінці його ефективності в умовах змішаного A2P-трафіку.

У дисертаційній роботі здобуто такі нові наукові результати:

- запропоновано показник Price Delivery Gap як універсальну метрику техніко-економічної ефективності доставки SMS;
- розроблено інтегровану модель IGDP, що поєднує QoS, QoE та PDG в єдину систему оцінювання;
- удосконалено методи математичного та регресійного моделювання ефективності доставки SMS в умовах гетероскедастичності;

- запропоновано метод виявлення штучно згенерованого трафіку в A2P-SMS;
- розроблено інтегрований фреймворк інтелектуального управління доставкою SMS у мережах 4G/5G.

Практичне значення отриманих результатів. Отримані результати дослідження можуть бути використані операторами мобільного зв'язку та CPaaS-платформами для оптимізації маршрутизації SMS-трафіку, підвищення ефективності використання мережевих ресурсів, зниження впливу АІТ, формування SLA-політик та підвищення економічної доцільності сервісу коротких повідомлень у мережах наступного покоління.

Ключові слова: телекомунікації, SMS, 4G/5G, QoS, QoE, PDG, IGDP, A2P-трафік, Artificially Inflated Traffic, інтелектуальна маршрутизація, ефективність доставки, апроксимація, метод найменших квадратів, регресійні моделі, математичні моделі, ефективність зв'язку, якість зв'язку, бездротові системи зв'язку, телекомунікаційні системи, математичне моделювання, мобільний зв'язок, блокування передачі, мережі зв'язку п'ятого покоління, експоненціальні та гіперболічні функції, експлуатація, моніторинг, машинне навчання.

ABSTRACT

Odarchenko Mykhailo Serhiiovych. Methods for evaluating and improving the efficiency of short message service (SMS) provision in next-generation networks – a qualifying scientific work in manuscript form.

Thesis for the degree of Doctor of Philosophy in the specialty 172 “Telecommunications and Radio Engineering” in the field of knowledge 17 “Electronics and Telecommunications.” State University “Kyiv Aviation Institute” of the Ministry of Education and Science of Ukraine. – Kyiv, 2026.

The dissertation is devoted to the study and solution of an urgent scientific and practical problem, which consists in improving the efficiency of short message service (SMS) delivery in 4G/5G networks by integrated consideration of technical, user, and economic factors, as well as minimizing the impact of artificially generated traffic (Artificially Inflated Traffic, AIT) on delivery quality and cost indicators.

The aim of the dissertation is to develop methods and models for quantitative assessment and improvement of SMS delivery efficiency in next-generation networks based on the integration of QoS, QoE, and economic efficiency indicators, formalized through the Price Delivery Gap (PDG) indicator.

To achieve this goal, the dissertation addresses the following tasks:

- analyze the evolution of SMS architecture and technologies in 2G–5G networks and identify factors that affect delivery efficiency;
- systematize scenarios for using SMS in 4G/5G networks and their impact on quality and cost parameters;
- justify PDG as an integral metric of the technical and economic efficiency of message delivery;
- develop an integrated IGDP (Integrated Gap-Delivery-Performance) model that combines QoS, QoE, and PDG;
- build mathematical and regression models for evaluating SMS delivery efficiency under heteroscedasticity;

- develop methods for detecting artificially generated traffic (AIT) in A2P-SMS;
- perform a scenario assessment of the effectiveness of an integrated SMS delivery management framework.

The first chapter analyzes the theoretical foundations, architectural features, and evolution of SMS service in 4G/5G networks, systematizes message usage scenarios, and formulates the scientific problem, goal, objectives, and hypothesis of the study.

The second chapter presents the methodological basis for evaluating the effectiveness of SMS delivery, justifies the PDG indicator, and develops an integrated IGDP model that combines technical, user, and economic performance parameters.

The third chapter presents mathematical and regression models of the SMS delivery process, analyzes the impact of QoS and QoE parameters on PDG, and investigates heteroscedasticity in message delivery data.

The fourth chapter is devoted to intelligent methods for improving the efficiency of SMS delivery, in particular message categorization, AIT detection, the development of an integrated delivery management framework, and a scenario-based assessment of its efficiency in mixed A2P traffic conditions.

The dissertation presents the following new scientific results:

- the Price Delivery Gap indicator is proposed as a universal metric of the technical and economic efficiency of SMS delivery;
- an integrated IGDP model is developed, combining QoS, QoE, and PDG into a single evaluation system;
- methods of mathematical and regression modeling of SMS delivery efficiency in conditions of heteroscedasticity have been improved;
- a method for detecting artificially generated traffic in A2P-SMS has been proposed;
- an integrated framework for intelligent management of SMS delivery in 4G/5G networks has been developed.

Practical significance of the results obtained. The results of the study can be used by mobile operators and CPaaS platforms to optimize SMS traffic routing, improve the efficiency of network resource utilization, reduce the impact of AIT, form SLA policies, and increase the economic feasibility of short message services in next-generation networks.

Keywords: telecommunications, SMS, 4G/5G, QoS, QoE, PDG, IGDP, A2P traffic, Artificially Inflated Traffic, intelligent routing, delivery efficiency, approximation, least squares method, quality metric, regression models, mathematical models, communication efficiency, communication quality, wireless communication systems, telecommunication systems, mathematical simulation, mobile communication, transmission blocking, fifth-generation communication networks, exponential and hyperbolic functions, operation, monitoring, machine learning.

Список публікацій здобувача за темою дисертації

1. Odarchenko M., Zaliskyi M. From SMS to RCS: Intelligent frameworks for resilient messaging in emergency scenarios. *Science-based technologies*. 2025. № 68(4), P. 504–510. DOI: <https://doi.org/10.18372/2310-5461.68.20800>.
2. Одарченко Р.С., Дика Т.В., Жарова О.В., Одарченко М.С., Жога В.М., Слободян О.П. Ключові напрямки досліджень стільникових мереж на шляху до 6G (огляд). *Наукоємні технології*. 2022. № 55(3). С. 215-228. DOI: <https://doi.org/10.18372/2310-5461.55.16904>.
3. Mykhailo Odarchenko, Serhii Zavorodnii, Roman Odarchenko and Maksym Zaliskyi (2022). “The Risks Assessment of Delivery Failures for Application-to-Person SMS Market”. *Computational & Information Technologies for Risk-Informed Systems CITRisk 2022*. <https://doi.org/10.5281/zenodo.18891183>
4. Odarchenko M., Zaliskyi M., An integrated techno-economic framework for sms delivery optimization in 4g/5g networks. *Electronics and Control Systems*. 2026. № 1(87). P. 52-58. DOI: <https://doi.org/10.18372/1990-5548.87.20885>.

Матеріали доповідей на конференціях, які засвідчують апробацію результатів дисертації

1. Odarchenko M., Odarchenko R., Zaliskyi M. (2022). “Modern challenges in security of SMS service and signaling networks”. “Проблеми кібербезпеки інформаційно-телекомунікаційних систем” (PCSITS), 27–28 жовтня 2022 р., Київ, Україна. С. 26-27.
2. Одарченко М.С., Заліський М.Ю, Одарченко Р.С. (2022). “Перспективи використання SMS у мережах п’ятого покоління”. "Проблеми експлуатації та захисту інформаційно-комунікаційних систем". НАУ (2022). С. 71-72.

ЗМІСТ

<i>ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ</i>	13
<i>РОЗДІЛ 1. АНАЛІЗ ТЕОРЕТИЧНИХ ЗАСАД ТА ПОКАЗНИКІВ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G</i>	21
1.1. Еволюція архітектури та технологій SMS	22
1.1.1. Загальний аналіз та особливості	22
1.1.2. Покоління 2G (GSM): класична модель SMSC	24
1.1.3. Покоління 3G (UMTS): IP-транзит та SIGTRAN	25
1.1.4. Покоління 4G (LTE): IP-SM-GW і конвергенція доменів.....	26
1.1.5. Покоління 5G (Standalone): SMSF і Service-Based Architecture	27
1.2. Сценарії використання та типологія SMS у мережах 4G/5G	29
1.2.1. Основні категорії використання SMS.....	29
1.2.2. Application-to-Person (A2P) як стратегічний напрям розвитку систем передачі інформації	30
1.2.3. Machine-to-Machine (M2M) та IoT-сценарії	31
1.2.4. RCS-fallback та OTT-інтеграції	32
1.2.5. Emergency & Public Warning Systems (PWS).....	33
1.3. Показники ефективності доставки та моделі оцінювання PDG у мережах нового покоління	34
1.3.1. Показники якості обслуговування (QoS)	34
1.3.2. Показники користувацької якості (QoE).....	35
1.3.3. Показник PDG (Price Delivery Gap)	36
1.3.4. Зв'язок між QoS, QoE та PDG	36
1.3.5. Інтегрована модель IGDP (Integrated Gap-Delivery-Performance)..	37
1.4. Наукова проблема, виклики, мета, завдання та гіпотеза дослідження	39
1.4.1. Сучасні виклики у сфері ефективності SMS у мережах нового покоління	39

	10
1.4.2. Постановка наукової проблеми та мета дослідження.....	39
1.4.3. Завдання та гіпотеза дослідження.....	40
1.5. Висновки до першого розділу.....	41
Список джерел.....	42
<i>РОЗДІЛ 2. МЕТОДОЛОГІЧНІ ОСНОВИ ОЦІНЮВАННЯ</i>	
<i>ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G.....</i>	<i>44</i>
2.1. Теоретичні підходи до визначення ефективності телекомунікаційних сервісів	44
2.1.1. Еволюція поняття ефективності у 4G/5G-середовищі.....	45
2.1.2. Інтеграція технічних, користувацьких і економічних факторів	46
2.2. Класифікація підходів до оцінювання ефективності доставки повідомлень	46
2.3. Теоретичне обґрунтування показника PDG (Price Delivery Gap) ..	50
2.3.1 Сутність показника PDG.....	50
2.3.2. Формальне визначення.....	51
2.3.3. Зв'язок із QoS та QoE.....	51
2.3.4. Місце PDG у 4G/5G-архітектурі	52
2.3.5. Інтерпретація та використання.....	52
2.4. Інтегрована модель IGDP як розвиток концепції PDG	53
2.4.1. Загальна концепція інтегрованої оцінки ефективності.....	53
2.4.2. Структура та змінні моделі IGDP	54
2.4.3. Теоретичне обґрунтування моделі.....	54
2.4.4. Практична інтерпретація IGDP для CPaaS-платформ.....	55
2.5. Методологічні основи моделювання процесу доставки SMS	55
2.5.1. Мета і принципи моделювання	55
2.5.2. Стохастичне представлення процесу доставки	56
2.5.3. Імовірно-часові моделі доставки.....	57
2.5.4. Оптимізаційна постановка задачі ефективності.....	58

	11
2.5.5. Економетричне моделювання та оцінка PDG.....	59
2.5.6. Узагальнення методологічних основ моделювання процесу доставки SMS.....	59
2.6. Вплив шахрайського трафіку на ефективність доставки SMS	60
2.6.1. Загальна характеристика.....	60
2.6.2. Метод виявлення та обмеження Artificially Inflated Traffic.....	63
2.7. Висновки до другого розділу	69
Список джерел.....	71
<i>РОЗДІЛ 3. МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ДОСТАВКИ SMS ТА ОЦІНЮВАННЯ ВПЛИВУ ПАРАМЕТРІВ ЕФЕКТИВНОСТІ НА PDG.....</i>	<i>74</i>
3.1. Мета та логіка математичного моделювання.....	74
3.2. Ключові параметри для визначення надійності послуги SMS.....	76
3.3. Побудова математичної моделі для параметра доставки SMS.....	80
3.4. Регресійні методи оцінювання ефективності доставки SMS в умовах гетероскедастичності	91
3.4.1. Загальні положення	91
3.4.2. OLS як базовий підхід та обмеження	92
3.4.3. Аналіз гетероскедастичності на основі методу рухомого трикутника.....	95
3.4.4. Аналіз гетероскедастичності на основі інтервальної функції.....	96
3.4.5. Регресія на основі кластерного підходу	97
3.4.6. Регресія на основі стиснення даних.....	98
3.4.7. Порівняльний аналіз розроблених моделей.....	99
3.5. Висновки до третього розділу	100
Список джерел.....	103

РОЗДІЛ 4. ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ПІДВИЩЕННЯ

ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G.....	104
4.1. Обмеження математичних моделей та необхідність інтелектуального аналізу процесу доставки SMS	104
4.2. Категоризація повідомлень як основа адаптивного управління ефективністю доставки	106
4.3. Виявлення штучно згенерованого трафіку (AIT) як чинник підвищення достовірності оцінювання ефективності доставки.....	109
4.4. Інтегрований фреймворк інтелектуального управління ефективністю доставки SMS	112
4.5. Сценарна оцінка ефективності інтегрованого фреймворку управління доставкою SMS	116
4.5.1. Вихідні умови та сценарії	116
4.5.2. Методика розрахунку показника PDG	117
4.5.3. Числові результати сценарної оцінки.....	118
4.5.4. Аналіз та інтерпретація результатів	121
4.5.5. Ключовий результат для сценарної оцінки ефективності інтегрованого фреймворку управління доставкою SMS	122
4.6. Висновки до четвертого розділу	122
Список джерел.....	124
ЗАГАЛЬНІ ВИСНОВКИ.....	126

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- A2A** – Application-to-Application, взаємодія між прикладними системами
- A2P** – Application-to-Person, передача повідомлень від додатків до користувачів
- AIT** – Artificially Inflated Traffic, штучно згенерований трафік
- AMF** – Access and Mobility Management Function, функція керування доступом і мобільністю в 5G
- API** – Application Programming Interface, прикладний програмний інтерфейс
- ARPU** – Average Revenue Per User, середній дохід на одного користувача
- CB** – Cell Broadcast, ширококомовна передача повідомлень
- CPaaS** – Communications Platform as a Service, платформа комунікацій як сервіс
- CS** – Circuit Switched, комутаційний домен
- Diameter** – протокол сигналізації для AAA-функцій у мережах LTE/IMS
- eGov** – Electronic Government, електронні державні сервіси
- eSIM** – Embedded SIM, вбудований SIM-профіль
- ETSI** – European Telecommunications Standards Institute
- GMSC** – Gateway Mobile Switching Centre, шлюзовий комутаційний центр
- GSM** – Global System for Mobile Communications
- HTTP/2** – Hypertext Transfer Protocol version 2
- IGDP** – Integrated Gap-Delivery-Performance, інтегрована модель оцінювання ефективності
- IMS** – IP Multimedia Subsystem, підсистема мультимедійних сервісів
- IoT** – Internet of Things, Інтернет речей
- IP** – Internet Protocol
- IP-SM-GW** – IP Short Message Gateway, шлюз доставки SMS у PS-домени
- JSON** – JavaScript Object Notation, формат обміну даними
- KPI** – Key Performance Indicator, ключовий показник ефективності
- LTE** – Long Term Evolution
- LTE-M** – LTE for Machines, стандарт мобільного зв'язку для IoT

MAP – Mobile Application Part, сигнальний протокол SS7

MME – Mobility Management Entity, вузол керування мобільністю в LTE

M2M – Machine-to-Machine, міжмашинна взаємодія

MSI – Messaging Satisfaction Index, індекс задоволеності сервісом повідомлень

NAS – Non-Access Stratum, сигнальний рівень керування в LTE/5G

NB-IoT – Narrowband Internet of Things

NEF – Network Exposure Function, функція доступу до мережевих можливостей

NF – Network Function, мережева функція

NFV – Network Functions Virtualization, віртуалізація мережевих функцій

NRF – Network Repository Function, реєстр мережевих функцій 5G

OLS – Ordinary Least Squares, метод найменших квадратів

OTP – One-Time Password, одноразовий пароль

OTT – Over-The-Top, інтернет-сервіси поверх мереж оператора

P2A – Person-to-Application, взаємодія користувача з додатком

P2P – Person-to-Person, передача повідомлень між користувачами

PDG – Price Delivery Gap, розрив між номінальною та ефективною вартістю доставки

PS – Packet Switched, пакетний домен

PWS – Public Warning System, система публічного оповіщення

QoE – Quality of Experience, якість користувацького досвіду

QoS – Quality of Service, якість обслуговування

RCS – Rich Communication Services

REST – Representational State Transfer

ROI – Return on Investment, рентабельність інвестицій

SBA – Service-Based Architecture, сервісно-орієнтована архітектура

SGs – інтерфейс між MME та MSC для доставки SMS у LTE

SIGTRAN – Signalling Transport, сигнальний транспорт поверх IP

SIP – Session Initiation Protocol

SLA – Service Level Agreement, угода про рівень обслуговування

SMS – Short Message Service

SMSC – Short Message Service Center

SMSF – Short Message Service Function, мережева функція SMS у 5G

SS7 – Signaling System No.7

UDM – Unified Data Management, функція управління даними абонента

UMTS – Universal Mobile Telecommunications System

ВСТУП

Актуальність теми. Сервіс коротких повідомлень (Short Message Service, SMS), попри стрімкий розвиток IP-орієнтованих месенджерів та мультимедійних платформ, залишається фундаментальним елементом телекомунікаційної інфраструктури мереж наступного покоління. У середовищі 4G/5G SMS використовується не лише як засіб міжособистісної комунікації, але й як універсальний транспортний механізм для критично важливих сервісів: двофакторної автентифікації (OTP), банківських і транзакційних повідомлень, державних сервісів (e-Government), систем екстреного оповіщення (Public Warning Systems), IoT-інтеграцій та резервного каналу доставки при недоступності IP-з'єднання.

Сучасна еволюція архітектури мобільних мереж від комутаційної моделі (2G/3G) до пакетно-орієнтованої (4G) та сервісно-орієнтованої архітектури 5G (Service-Based Architecture) призвела до трансформації механізмів доставки SMS. Введення компонентів IP-SM-GW у LTE та SMSF у 5G забезпечило гнучкість маршрутизації, інтеграцію з OTT- і RCS-сервісами, підтримку мережевих слайсів і SLA-політик. Водночас така трансформація ускладнила структуру витрат, збільшила варіативність затримок доставки та створила нові виклики щодо забезпечення стабільної якості обслуговування.

Традиційні підходи до оцінювання ефективності SMS ґрунтуються переважно на технічних показниках якості обслуговування (Quality of Service, QoS), таких як затримка доставки, коефіцієнт успішної доставки та рівень повторних спроб. Однак у сучасному A2P-сегменті (Application-to-Person), що формує основну частку глобального SMS-трафіку, виключно технічні метрики не відображають реальної економічної ефективності сервісу. Зростання обсягів трафіку, диференціація тарифів, складні маршрути між операторами, а також поширення штучно згенерованого трафіку (Artificially Inflated Traffic,

АІТ) формують розрив між номінальною вартістю доставки повідомлення та його фактичною ефективною вартістю.

У таких умовах виникає необхідність розробки інтегрованих методів оцінювання ефективності SMS-доставки, які б поєднували технічні (QoS), користувацькі (QoE) та економічні показники в єдину формалізовану систему аналізу. Особливої актуальності набуває створення кількісних моделей, здатних враховувати вплив гетероскедастичності даних доставки, варіативності мережевого навантаження та шахрайських схем генерації трафіку.

Таким чином, тема дисертаційної роботи, присвячена розробці методів оцінки та підвищення ефективності надання сервісу коротких повідомлень у мережах 4G/5G, є актуальною, відповідає сучасним тенденціям розвитку телекомунікаційних систем та має вагоме наукове й практичне значення.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана відповідно до пріоритетних напрямів розвитку науки і техніки України у сфері електроніки та телекомунікацій, визначених нормативними документами Міністерства освіти і науки України. Тематика дослідження відповідає сучасним науковим напрямам розвитку мобільних мереж 4G/5G, цифрової трансформації комунікаційних сервісів та підвищення ефективності телекомунікаційних систем.

Основні результати роботи отримано в межах науково-дослідних напрямів кафедри телекомунікаційних та радіоелектронних систем Державного університету «Київський авіаційний інститут», що пов'язані з дослідженням ефективності функціонування сучасних телекомунікаційних сервісів та методів їх оптимізації.

Мета і завдання дослідження. Метою дисертаційної роботи є підвищення ефективності доставки сервісу коротких повідомлень (SMS) у мережах наступного покоління на основі інтеграції показників QoS, QoE та економічної ефективності, формалізованої через інтегральний показник Price Delivery Gap (PDG).

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Проаналізувати еволюцію архітектури та технологій SMS у мережах 2G–5G та визначити чинники, що впливають на ефективність доставки.
2. Систематизувати сучасні сценарії використання SMS у середовищі 4G/5G та оцінити їх вплив на параметри якості й вартості.
3. Обґрунтувати інтегральний показник PDG як метрику техніко-економічної ефективності доставки повідомлень.
4. Розробити інтегровану модель IGDP (Integrated Gap-Delivery-Performance), що поєднує технічні, користувацькі та економічні параметри ефективності.
5. Побудувати математичні та регресійні моделі оцінювання ефективності доставки SMS з урахуванням гетероскедастичності даних.
6. Розробити методи виявлення штучно згенерованого трафіку (AIT) у A2P-сегменті.
7. Виконати сценарну оцінку ефективності інтегрованого фреймворку управління доставкою SMS у змішаному трафіку.

Об'єктом дослідження є процес надання та доставки сервісу коротких повідомлень у мережах 4G/5G.

Предметом дослідження є методи, моделі та алгоритми оцінювання і підвищення техніко-економічної ефективності доставки SMS з урахуванням показників QoS, QoE та впливу штучно згенерованого трафіку.

Методи дослідження: у дисертаційній роботі використано методи теорії телекомунікаційних систем, синтезу та аналізу, математичного та економетричного моделювання, регресійного аналізу, аналізу гетероскедастичності, методи статистичного оброблення даних, сценарного моделювання та інтелектуального аналізу трафіку.

Наукова новизна отриманих результатів:

1. Вперше запропоновано інтегральний показник Price Delivery Gap та розроблено інтегровану модель IGDP для комплексного оцінювання техніко-економічної ефективності доставки SMS у мережах наступного покоління, що, на відміну від існуючих підходів, дає змогу кількісно оцінювати відхилення між номінальною та фактичною вартістю доставки повідомлень з урахуванням параметрів QoS і QoE, а також поєднувати технічні, користувацькі та економічні показники в єдину систему оцінювання для оптимізації маршрутизації й підвищення функціонування SMS сервісу.

2. Удосконалено методи математичного та регресійного моделювання процесу доставки SMS, які на відміну від існуючих ураховують гетероскедастичність датасету та надають змогу будувати більш точні та стійкі до змін дисперсії похибки математичні моделі.

3. Удосконалено метод виявлення штучно згенерованого трафіку в A2P-сегменті, який на відміну від існуючих базується на поєднанні аналізу часових характеристик трафіку, параметрів доставки та економічних індикаторів ефективності, та надає змогу локалізувати вплив штучно згенерованого трафіку на показники доставки повідомлень, підвищити точність виявлення аномальних сценаріїв та зменшити економічні втрати операторів і CPaaS-платформ.

4. Подальшого розвитку набула розробка інтегрованого фреймворку інтелектуального управління ефективністю доставки SMS у мережах 4G/5G, який на відміну від існуючих поєднує категоризацію повідомлень, механізми виявлення АІТ та інтегральну оцінку ефективності на основі показника PDG, та надає змогу реалізувати адаптивне управління маршрутизацією повідомлень і вибором каналів доставки з урахуванням технічних, сервісних та економічних параметрів функціонування мережі.

Практичне значення одержаних результатів. Отримані результати можуть бути використані операторами мобільного зв'язку, CPaaS-платформами та розробниками телекомунікаційних сервісів для оптимізації

маршрутизації SMS-трафіку, підвищення ефективності використання мережевих ресурсів, формування SLA-політик, зниження впливу АІТ та підвищення економічної доцільності сервісу коротких повідомлень у мережах наступного покоління.

Особистий внесок здобувача. Усі наукові положення, моделі та результати, що виносяться на захист, отримані автором самостійно. У роботах, виконаних у співавторстві, автору належать постановка задач, розробка моделей, проведення моделювання та інтерпретація отриманих результатів.

Апробація результатів дисертації. Основні положення та результати дисертації доповідалися на міжнародних та всеукраїнських науково-технічних конференціях з питань телекомунікацій та інформаційних технологій, а саме: «Проблеми кібербезпеки інформаційно-телекомунікаційних систем (PCSITS)» (Київ, 2022), "Проблеми експлуатації та захисту інформаційно-комунікаційних систем". (Київ, 2022)

Публікації. Основні результати дисертаційної роботи опубліковано у 3 фахових наукових виданнях та 1 матеріалу міжнародних конференцій.

Обсяг і структура дисертації. Дисертаційна робота складається з анотації, змісту, переліку умовних позначень, вступу, чотирьох розділів, загальних висновків, списку використаних джерел і додатків. Загальний обсяг роботи становить 128 сторінок друкованого тексту, основна частина займає 127 сторінок. Робота ілюстрована 29 рисунками. Список використаних джерел містить 63 найменування.

РОЗДІЛ 1.

АНАЛІЗ ТЕОРЕТИЧНИХ ЗАСАД ТА ПОКАЗНИКІВ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G

Сервіс коротких повідомлень (Short Message Service, SMS) залишається однією з найстійкіших технологій мобільного зв'язку. Протягом понад тридцяти років він виконує критичні функції обміну інформацією між абонентами, системами і сервісами. На відміну від інтернет-орієнтованих месенджерів, SMS забезпечує гарантовану доставку повідомлень через сигнальний канал (Control Plane), що робить його надійним навіть за нестабільного IP-з'єднання чи перевантаження мережі [1, 2].

Згідно зі звітом GSMA *Mobile Economy 2024*, у світі налічується понад 5,6 млрд користувачів мобільного зв'язку, а загальний обсяг A2P-повідомлень (Application-to-Person) у 2023 році перевищив 2,3 трлн. Попри зростання частки OTT-платформ (WhatsApp, Telegram, WeChat, Viber), SMS залишається єдиним універсальним каналом, що підтримується всіма операторами мобільного зв'язку та сумісний із будь-якими типами пристроїв [1].

Технологічно SMS є базовою послугою, реалізованою на рівні Control Plane Signalling, що використовує протоколи SS7/MAP/SIGTRAN у мережах GSM та UMTS, а в LTE та 5G – інтерфейси SGs і HTTP/2 (REST API) [2, 3]. Це забезпечує високу доступність та доставку до понад 99,9 % повідомлень, незалежно від наявності IP-каналу. SMS активно використовується в банківських і платіжних сервісах (OTP-верифікація), системах e-Governance, медичних / IoT-платформах, а також у механізмах надзвичайного оповіщення (Public Warning Systems, PWS) [4,5].

1.1. Еволюція архітектури та технологій SMS

1.1.1. Загальний аналіз та особливості

Еволюція архітектури SMS відображає загальний перехід мобільних мереж від комутаційних систем до сервісно-орієнтованої (Service-Based) архітектури 5G. Основні етапи переходу [3, 4, 6]:

- 2G (GSM): передача через SMSC з використанням MAP / SS7;
- 3G (UMTS): підтримка SIGTRAN та GMSC для IP-транзиту;
- 4G (LTE): запровадження IP-SM-GW, що маршрутизує SMS у Packet Switched домені;
- 5G (SA): новий вузол SMSF (Short Message Service Function) у рамках Service-Based Architecture, який використовує HTTP/2 та JSON інтерфейси для взаємодії між мережевими функціями. Еволюція архітектури сервісу коротких повідомлень (SMS) відображає перехід мобільних мереж від комутаційної логіки (Circuit-Switched) до пакетно-орієнтованої (Packet-Switched), а згодом, до сервісно-орієнтованої архітектури (Service-Based Architecture, SBA) у 5G. Кожен етап трансформації визначав нові вимоги до протоколів, функціональних вузлів, стандартів якості обслуговування (QoS) та економічної ефективності доставки. Відповідно, значення PDG (Price Delivery Gap) змінювалося разом з розвитком інфраструктури – від мінімального у 2G до динамічного й чутливого у 5G.

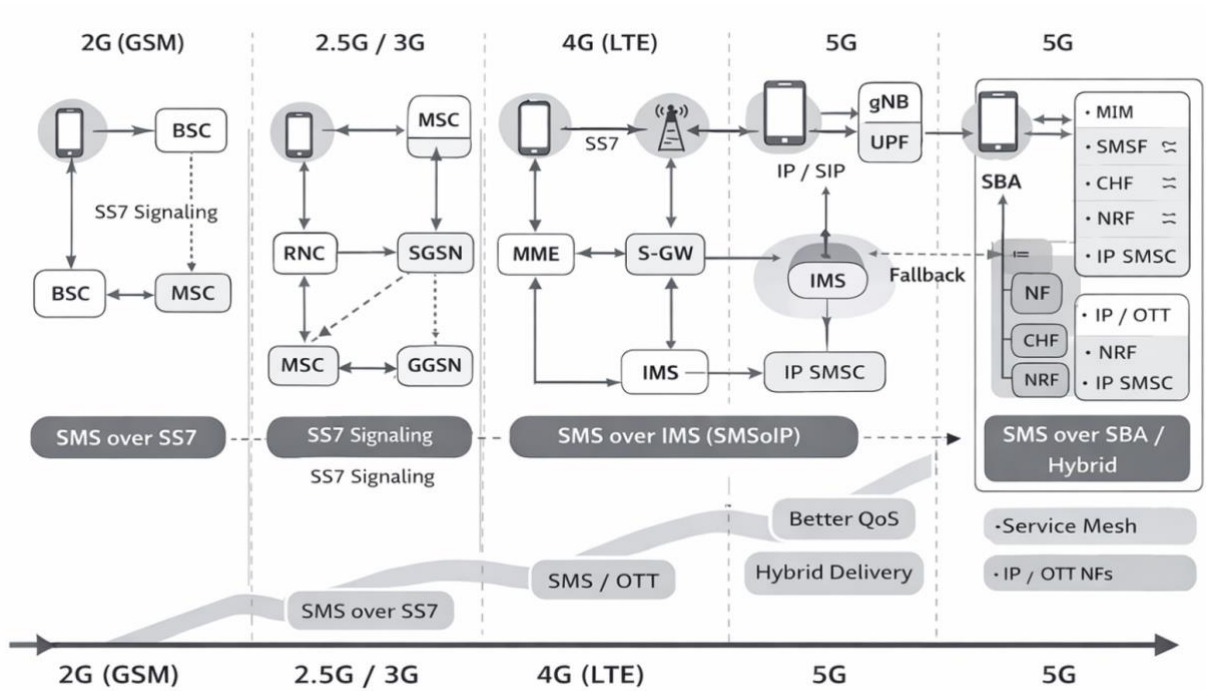


Рис. 1.1. Еволюція архітектури систем передачі SMS від 2G до 5G

Перехід до SMSF та мережевої архітектури SBA сприяв підвищенню гнучкості, масштабованості й можливості інтеграції з OTT-сервісами, але одночасно викликав зростання затримки доставки та вимоги до QoS-контролю в реальному часі. Відповідно до 3GPP TS 29.891 та ETSI TR 103 559, SMSF забезпечує уніфіковане керування маршрутизацією та підтримує функції RCS/OTT-fallback, що підвищує надійність доставки та знижує ризик втрат повідомлень [3, 4]. Ефективність SMS-сервісу в мережах 4G/5G визначається комплексом взаємопов'язаних факторів:

- QoS (Quality of Service) – технічні показники доставки та продуктивності мережі;
- QoE (Quality of Experience) – сприйняття користувачем якості послуги;
- PDG (Price Delivery Gap) – розрив між номінальною вартістю доставки повідомлення та фактичною ефективною вартістю, скоригованою на показники якості. В межах цього дослідження PDG (Price Delivery Gap) використовується як інтегральний показник техніко-економічної ефективності сервісу SMS, який описується виразом

$$PDG = \frac{P_{exp} - P_{eff}}{P_{exp}}, \quad (1.1)$$

де P_{exp} – очікувана (тарифна) вартість доставки; P_{eff} – ефективна вартість з урахуванням затримок, повторних спроб та відмов; $PDG > 0$ – доставка ефективніша за очікування; $PDG < 0$ – втрата ефективності, переплата чи зниження якості.

Таким чином, PDG поєднує економічний та технічний виміри ефективності, дозволяючи оцінювати не лише вартість послуги, а й якість її реалізації в різних мережевих умовах. Це відповідає підходам ETSI TR 103 559 та ITU-T Y.1541, які рекомендують враховувати QoS та QoE-фактори у комплексному аналізі параметрів доставки [4, 6].

У системному сенсі, показник PDG стає економіко-технічною ланкою в інтегрованій моделі IGDP (Integrated Gap-Delivery-Performance), що буде детально розглянута в параграфі 1.3 та другому розділі. Він дозволяє кількісно зіставити вплив QoS, QoE та тарифної політики на загальну ефективність сервісу та є основою для подальшого моделювання ризику недоставки повідомлень у мережах 5G.

1.1.2. Покоління 2G (GSM): класична модель SMSC

SMS як сервіс з'явився у складі стандарту 3GPP TS 23.040 для GSM і спочатку був допоміжною функцією системи сигналізації SS7. Передача повідомлень здійснювалась каналом SACCH (Slow Associated Control Channel), що забезпечував store-and-forward модель через Short Message Service Center (SMSC). SMSC виконував буферизацію, чергу, повторну доставку при недоступності абонента, а також зберігав звіти про доставку [7, 8].

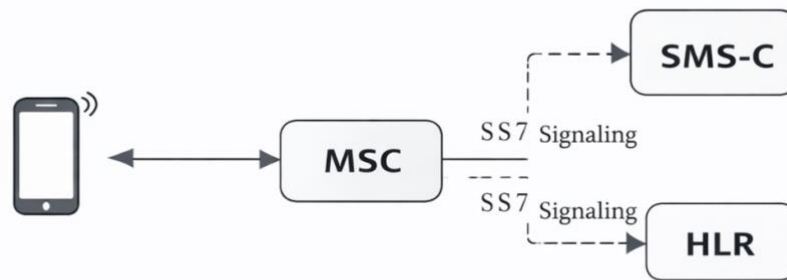


Рис. 1.2. Покоління 2G (GSM): класична архітектура доставки SMS

Характерні технічні параметри для GSM:

- протокол: MAP (Mobile Application Part) поверх SS7;
- середня затримка доставки: 3–6 секунд;
- коефіцієнт успішної доставки: 97–99 %;
- тип з'єднання: Circuit Switched, односторонній обмін;
- стійкість до відмов – висока, але без QoS-моніторингу [7, 8].

На цьому етапі PDG був мінімальним, оскільки вартість і якість були практично сталими – мережеві ресурси не тарифікувалися окремо, а собівартість доставки залежала лише від сигнального навантаження.

1.1.3. Покоління 3G (UMTS): IP-транзит та SIGTRAN

Перехід до UMTS (3G) привів до впровадження IP-орієнтованого сигнального транспорту через стек SIGTRAN (M3UA/SCTP), який забезпечив передавання MAP-повідомлень через IP-мережу з мінімальною затримкою [8][9]. SMSC залишився базовим елементом, але до архітектури додалися шлюзи GMSC і I-SMSC, що дозволяли маршрутизацію через IP-домен. У результаті:

- покращилася пропускна здатність сигнального рівня;

- з'явилася можливість контролю QoS для MAP-повідомлень;
- затримка скоротилася до 2-4 секунд;

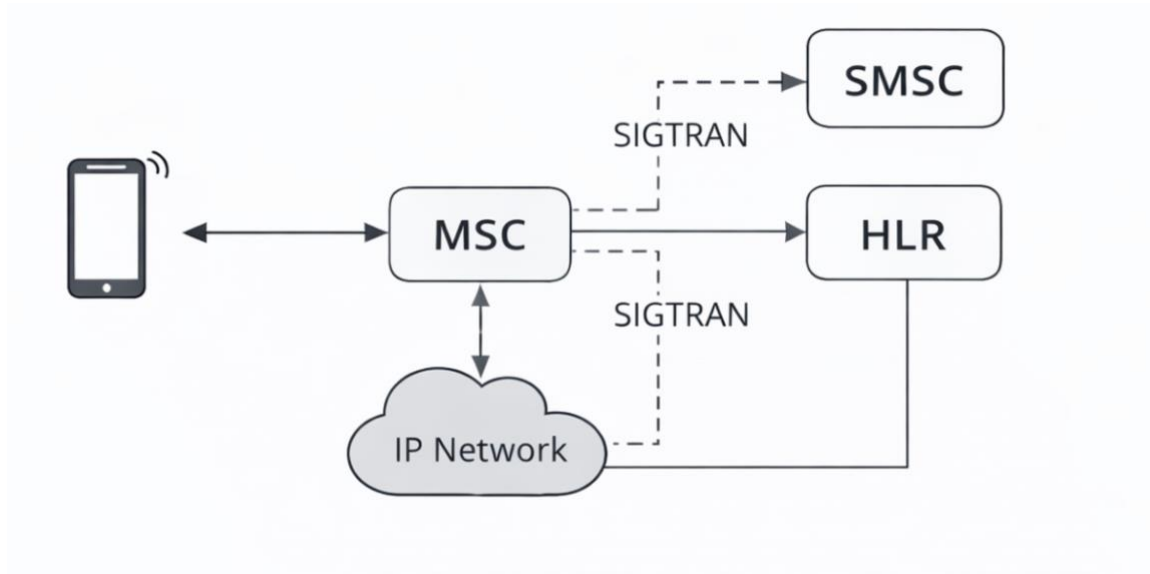


Рис. 1.3. Покоління 3G (UMTS): IP-транзит сигнального трафіку та SIGTRAN

PDG став коливатися залежно від навантаження в IP-доміні – тобто вперше з'явилася варіативність між очікуваною та реальною вартістю доставки.

Згідно з ETSI TR 103 559, цей перехід можна вважати початком формування метрик “якісної вартості доставки”, де до собівартості враховується час, навантаження й кількість повторних спроб [10].

1.1.4. Покоління 4G (LTE): IP-SM-GW і конвергенція доменів

У мережах LTE (EPC) SMS уже не міг існувати в чистому вигляді Circuit-Switched, тому було створено компонент IP Short Message Gateway (IP-SM-GW), який взаємодіє одночасно з MME (Mobility Management Entity) і IMS Core. Відповідно до 3GPP TS 29.338, саме IP-SM-GW забезпечував маршрутизацію між CS-доменом (через SGs) і PS-доменом (через SIP/IMS) [9, 11]. Це був перший крок до повної конвергенції доменів CS-PS, який відкрив можливість:

- інтеграції SMS з VoLTE та IMS-послугами;

- обробки повідомлень через Diameter і HTTP-протоколи;
- запровадження централізованих білінгових і QoS-контролерів;
- реалізації RCS (Rich Communication Services) як надбудови [11, 12].

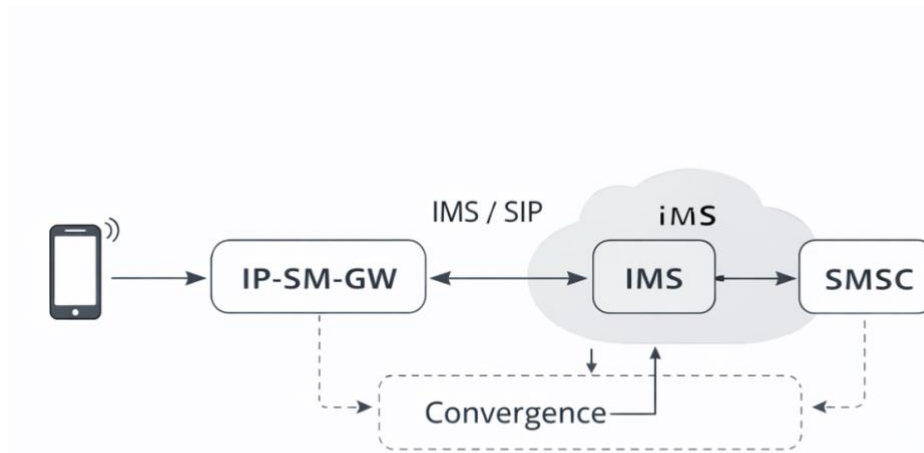


Рисунок 1.4. Покоління 4G (LTE): конвергенція CS- та PS-доменів

Однак саме в LTE з'явився виражений розрив між номінальною та ефективною ціною доставки – тобто PDG почав залежати не лише від якості радіосегменту, а й від навантаження в IMS-доміні, політики ретрансляцій і тарифів для A2P-трафіку. Згідно з ETSI TS 102 250-2 і ITU-T Y.1541, саме в цей період почали формуватися QoS-профілі для сервісів обміну повідомленнями, де затримка, втрати пакетів і джиттер впливали на загальну вартість повідомлення [12, 13].

1.1.5. Покоління 5G (Standalone): SMSF і Service-Based Architecture

Найновіший етап 5G SA (Standalone) привів до створення Short Message Service Function (SMSF) як нативної мережевої функції (NF) у Service-Based Architecture (SBA). Згідно з 3GPP TS 29.891, SMSF реалізує транспорт повідомлень через HTTP/2 + JSON API між сервісами AMF, UDM, NEF і SMS Router [9, 14]. Кожне повідомлення стає сервісною транзакцією, яка може оброблятися в межах Network Slice із власними QoS-параметрами.

Основні переваги архітектури SMSF:

- сервісно-орієнтована маршрутизація через NRF;
- можливість інтеграції з OTT та RCS fallback;
- аналітика QoS/QoE у реальному часі;
- підтримка політик SLA на рівні абонента чи сегмента [14, 15].

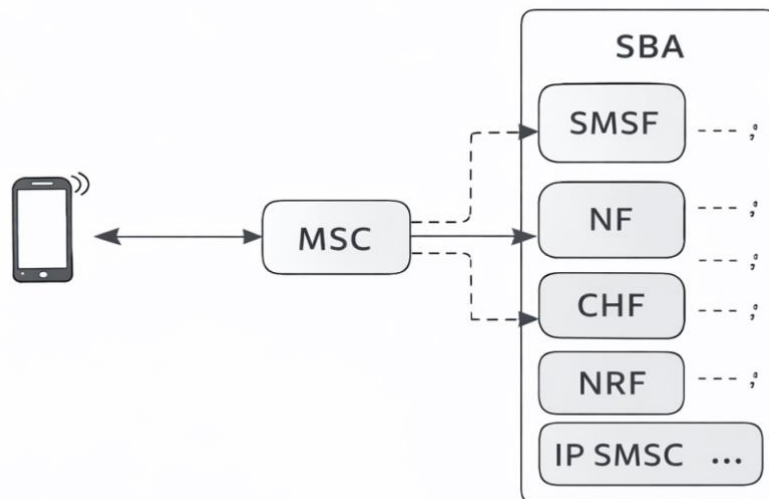


Рисунок 1.5. Покоління 5G (Standalone): Service-Based Architecture та SMSF

У таких умовах PDG (Price Delivery Gap) стає динамічним показником: затримка, повтори або зміна маршрутизації збільшують ефективну вартість, а автоматизована маршрутизація та SLA-контроль зменшують її. ETSI TR 103 559 і IEEE Access (2023) підтверджують, що ефективне управління PDG є ключовим інструментом для балансування витрат і якості доставки у 5G [10][15]. На базі порівняльного аналізу вище наведених параметрів автором розроблена табл. 1.1.

Еволюція архітектури SMS від GSM до 5G

Покоління	Архітектура	Протоколи	QoS контроль	Середня затримка	Характер PDG
2G GSM	CS, SMSC	MAP / SS7	Відсутній	3–6 с	≈0 (стабільний)
3G UMTS	CS + IP	MAP / SIGTRAN	Обмежений	2–4 с	Низький
4G LTE	PS + IMS	SGs / Diameter / SIP	Частковий	2–5 с	Середній
5G SA	SBA, SMSF	HTTP/2 + JSON	Повний	1–3 с	Динамічний

Узагальнюючи, можна зробити висновки, що розвиток архітектури SMS можна розглядати як перехід від фіксованої сигнальної моделі до адаптивної сервісної екосистеми, де якість і вартість доставки взаємопов'язані через параметри QoS, QoE та PDG. У 5G контексті PDG виступає метрикою операційної ефективності, яка відображає, наскільки реальна доставка відповідає тарифній політиці та SLA-очікуванням. Це створює передумови для формування інтегрованої моделі IGDP (Integrated Gap-Delivery-Performance), яка буде детально описана в наступних підрозділах.

1.2. Сценарії використання та типологія SMS у мережах 4G/5G

1.2.1. Основні категорії використання SMS

З розвитком мобільних мереж четвертого й п'ятого поколінь сервіс коротких повідомлень набув нових функціональних форм і сценаріїв застосування. Традиційна роль SMS як міжособистісного каналу обміну текстовими повідомленнями (P2P) поступила місцем багат шаровій екосистемі взаємодій між користувачами, додатками, машинами й інформаційними системами. У 4G/5G середовищі SMS перетворився на

модульну транспортну платформу, що підтримує інтероперабельність між доменами CS, PS, IMS та OTT, забезпечуючи доставку повідомлень незалежно від технологічного середовища [7, 9, 10].

Відповідно до класифікації GSMA (2024) та рекомендацій 3GPP TS 23.040, розроблена таблиця (табл. 1.2), яка систематизує сучасний SMS-трафік умовно поділяючи на типи та характеристики [7, 9, 15, 16].

Таблиця 1.2

Типологія сучасного SMS-трафіку

Категорія	Повна назва	Тип користувача	Характеристика / мета	Частка у глобальному трафіку, %*
P2P (Person-to-Person)	Обмін між користувачами	Людина → Людина	Традиційні повідомлення	≈ 10
A2P (Application-to-Person)	Повідомлення від додатків	Система → Людина	ОТР, банкінг, маркетинг, eGov	≈ 65
P2A (Person-to-Application)	Зворотна взаємодія	Людина → Система	Опитування, підтвердження, запити	≈ 15
M2M (Machine-to-Machine)	Технічна взаємодія	Пристрій → Пристрій	ІоТ, телеметрія, eSIM OTA	≈ 8
A2A (Application-to-Application)	Сервісна взаємодія	Система → Система	API-інтеграції, CPaaS-сервіси	≈ 2

* - За даними GSMA Mobile Economy 2024 та ITU-D Measuring Digital Development 2023.

1.2.2. Application-to-Person (A2P) як стратегічний напрям розвитку систем передачі інформації

Сегмент A2P-SMS є головним джерелом доходу операторів і CPaaS-платформ. За оцінками GSMA (2024), він забезпечує понад 65 % усього глобального SMS-трафіку й має середньорічне зростання близько 6 %. A2P використовується для:

- автентифікації (OTP-коди, підтвердження транзакцій);
- інформаційних сповіщень (банки, служби доставки, державні сервіси);
- маркетингових кампаній (регульованих відповідно до GDPR / TCRA);
- машинних інтеграцій (API-виклики, двофакторна верифікація).

У мережах 4G/5G доставка A2P-повідомлень контролюється SMSF через REST-інтерфейси N20/N32, що дозволяє гнучко керувати параметрами QoS та SLA-пріоритетами [9, 14, 15].

Операційна ефективність у цьому сегменті оцінюється через PDG (Price Delivery Gap) – різницю між очікуваною тарифною вартістю A2P-повідомлення та фактичною ефективною вартістю з урахуванням QoS-факторів. При зниженні швидкості доставки чи підвищенні retry-rate PDG має від’ємне значення, сигналізуючи про погіршення ефективності каналу [10, 12, 17].

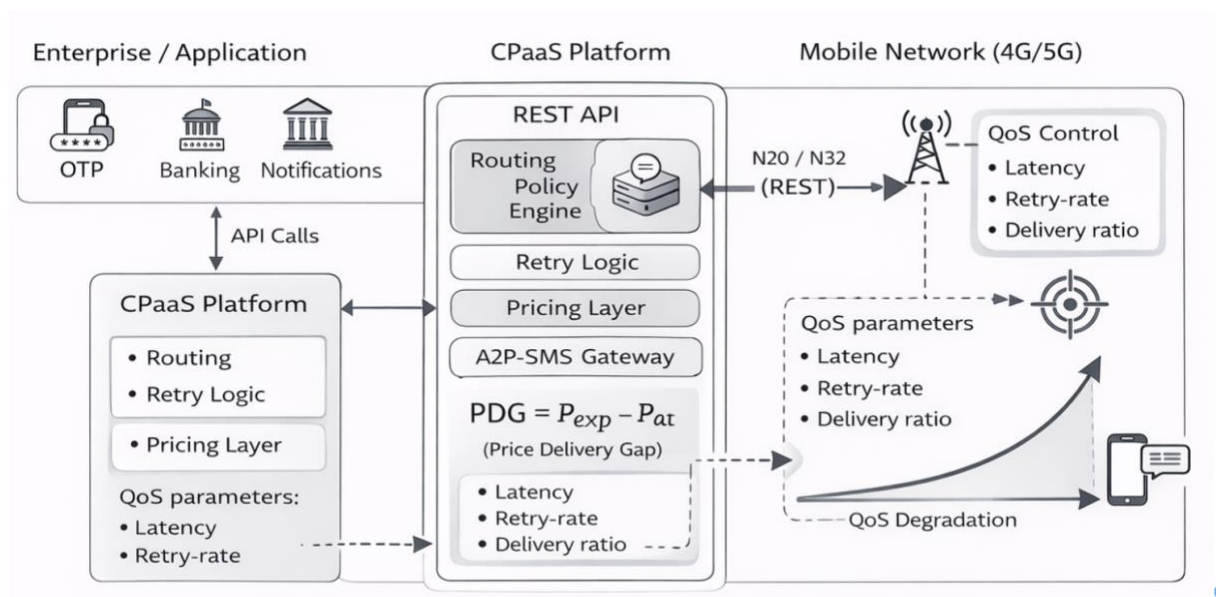


Рисунок 1.6. Архітектура та економічна модель A2P-SMS у мережах 4G/5G

1.2.3. Machine-to-Machine (M2M) та IoT-сценарії

У системах M2M і Internet of Things (IoT) SMS використовується як резервний канал обміну даними, зокрема для:

- передачі коротких телеметричних повідомлень;
 - дистанційного керування пристроями;
 - ініціалізації або відновлення з'єднання;
 - оновлення профілів eSIM / UICC через OTA-механізм (ETSI TS 31.115)
- [18].

SMS тут має переваги: він не потребує постійного IP-з'єднання, має малий обсяг і забезпечує гарантовану доставку навіть у низькошвидкісних мережах (NB-IoT, LTE-M). У таких системах PDG зазвичай позитивний, оскільки вартість одиниці передачі суттєво менша порівняно з IP-каналами, а втрати даних мінімальні [17, 18, 19].

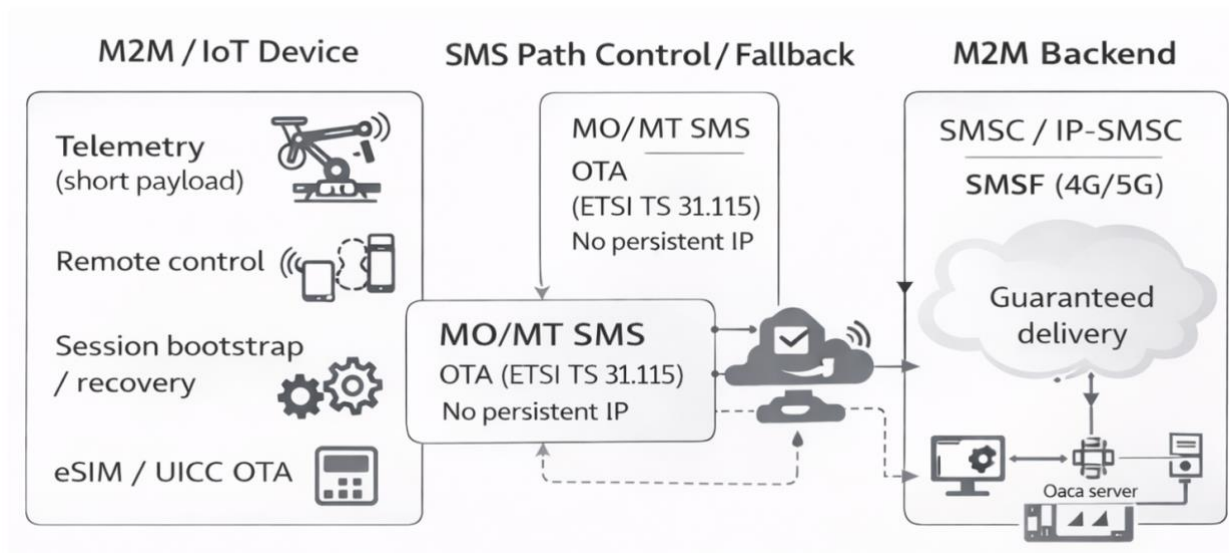


Рисунок 1.7. Інфраструктура передачі даних в контексті M2M/IoT

1.2.4. RCS-fallback та OTT-інтеграції

З поширенням Rich Communication Services (RCS) та OTT-платформ SMS отримав функцію fallback-каналу. Коли OTT-додаток недоступний або абонент перебуває поза мережею даних, повідомлення передається у форматі SMS. Цей механізм стандартизовано в 3GPP TS 24.229 і реалізовано через взаємодію RCS-клієнта, IP-SM-GW та SMSF [9, 14, 19]. Для операторів це створює подвійний ефект: з одного боку, гарантується безперервність

доставки, з іншого – виникають додаткові витрати через дублювання каналів. Тому PDG у таких сценаріях може набувати негативного значення (доставка дорожча за номінал) – особливо у випадках багаторазової перемаршрутизації або коли fallback відбувається через перевантаження IMS-домену [10, 15].

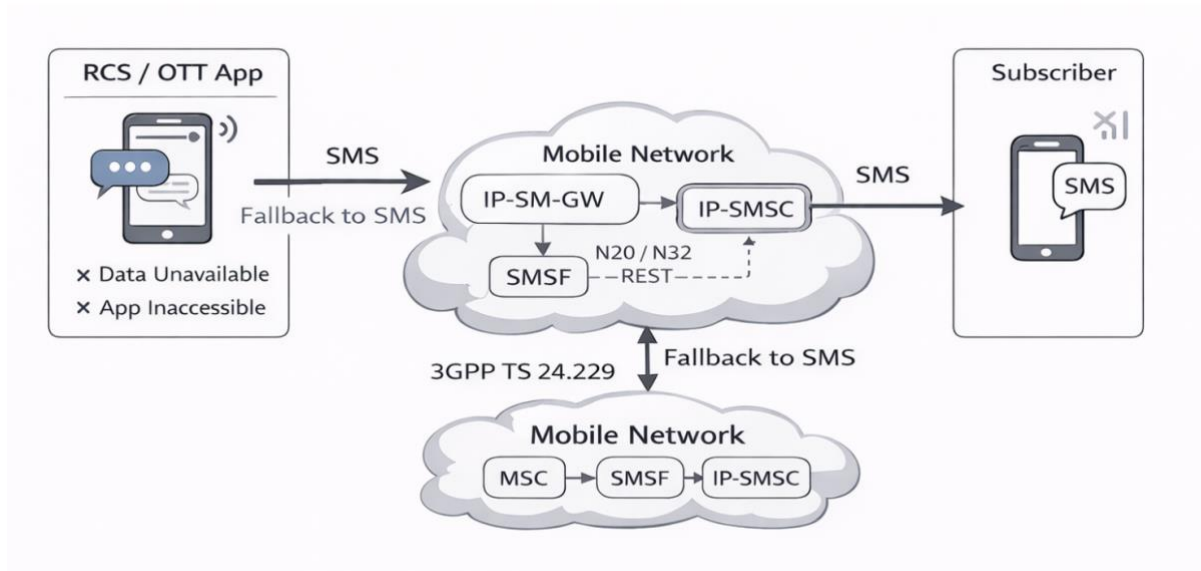


Рисунок 1.8. Fallback до SMS у системах RCS та OTT

1.2.5. Emergency & Public Warning Systems (PWS)

Окремий напрям використання SMS – Public Warning Systems (PWS), що регламентується стандартом ETSI TS 102 900. У більшості країн ЄС, а також у США, Канаді та Японії SMS застосовується як основний або резервний канал для надсилання масових повідомлень під час надзвичайних ситуацій – землетрусів, повеней, повітряних тривог [14, 19, 20]. Особливості реалізації:

- передавання з пріоритетом у сигнальному каналі (Class 0/Flash SMS);
- одночасна доставка великій кількості абонентів (broadcast-тип);
- SLA-вимоги – latency < 5 с, delivery ratio > 98 % [12, 14]. PDG у таких системах має нейтральне або нульове значення, оскільки вартість не є релевантною метрикою, а ефективність вимірюється через швидкість і повноту охоплення. Проте сама концепція PDG може застосовуватися для

аналізу економічної доцільності комбінованих каналів (SMS + cell broadcast) у державних системах оповіщення [20].

Узагальнюючи п. 1.2 можна зробити висновки, що різноманітність сценаріїв використання SMS у 4G/5G середовищі вимагає багатовимірної оцінки ефективності. Показник PDG (Price Delivery Gap) у цьому контексті виступає універсальною метрикою, що дозволяє інтегрувати технічні параметри (затримка, повтори, QoS-рівень) та економічні (тариф, собівартість, SLA-штрафи) у єдину систему оцінки продуктивності сервісу. Таким чином, PDG стає сполучною ланкою між якістю обслуговування (QoS/QoE) та операційною ефективністю, формуючи основу для подальшого моделювання інтегрованої системи IGDP.

1.3. Показники ефективності доставки та моделі оцінювання PDG у мережах нового покоління

1.3.1. Показники якості обслуговування (QoS)

Оцінювання ефективності сервісу коротких повідомлень (SMS) у мережах 4G/5G базується на сукупності технічних і експлуатаційних параметрів, що описують швидкість, надійність і економічну доцільність доставки. Ключовими серед них є QoS (Quality of Service), QoE (Quality of Experience) та PDG (Price Delivery Gap), які формують інтегровану модель оцінювання ефективності в межах концепції IGDP (Integrated Gap-Delivery-Performance).

Згідно з рекомендацією ITU-T Y.1541 та стандартами ETSI TS 102 250, якість телекомунікаційної послуги визначається набором параметрів, що характеризують її технічну ефективність, які приведені в табл. 1.3 [12, 17, 21]. Кожен із показників в табл. 1.3 може впливати на фактичну собівартість доставки SMS, оскільки збільшення затримки чи повторних спроб призводить

до додаткових витрат ресурсів. Отже, QoS безпосередньо корелює з компонентом P_{eff} у формулі Price Delivery Gap.

Таблиця 1.3

Набір параметрів, що характеризують технічну ефективність послуг

Параметр	Позначення	Опис	Одиниця виміру	Типовий діапазон
Затримка доставки	T_d	Середній час від відправлення до отримання	с	1–5
Коефіцієнт втрат	L_r	Частка недоставлених повідомлень	%	0–2
Коефіцієнт повторів	R_r	Відсоток повторних надсилянь	%	1–10
Пропускна здатність	B_w	Обсяг повідомлень за одиницю часу	msg/s	100–10 000
Доступність каналу	A_c	Імовірність доступу до мережі	%	99–99.999

1.3.2. Показники користувацької якості (QoE)

На відміну від QoS, показник QoE (Quality of Experience) відображає суб'єктивне сприйняття користувачем якості сервісу. Згідно з рекомендаціями ETSI TR 103 559 та IEEE Access (2023), оцінка QoE для повідомлень може проводитися через інтегральний індекс задоволеності (MSI – Messaging Satisfaction Index) [10, 13, 22]:

$$QoE = f(T_d, L_r, R_r, A_c) = w_1 e^{-k_1 T_d} + w_2 (1 - L_r) + w_3 (1 - R_r) + w_4 A_c, \quad (1.2)$$

де: w_i – вагові коефіцієнти впливу кожного параметра; k_1 – константа експоненційного згасання.

Практичні дослідження показують, що найбільший вплив на QoE мають затримка та повтори, особливо в сегментах A2P і PWS [21, 23].

У системах CPaaS показник QoE часто використовується для визначення порогів SLA (Service Level Agreement) і може бути інтегрований у механізм розрахунку PDG як якісна корекція фактичної вартості.

1.3.3. Показник PDG (Price Delivery Gap)

В основі оцінки ефективності SMS лежить показник PDG (Price Delivery Gap), що відображає різницю між номінальною (тарифною) і фактичною (ефективною) вартістю доставки з урахуванням параметрів QoS і QoE:

$$PDG = \frac{P_{exp} - P_{eff}}{P_{exp}} = 1 - \frac{P_{eff}}{P_{exp}}, \quad (1.3)$$

де P_{exp} – очікувана вартість (тариф або середня ціна доставки згідно з контрактом), P_{eff} – ефективна вартість, що враховує повтори, затримки та зниження якості.

Величина P_{eff} може бути обчислена через зважену функцію:

$$P_{eff} = P_{exp} \left(1 + \alpha_1 L_r + \alpha_2 R_r + \alpha_3 \frac{T_d}{T_{ref}} - \alpha_4 A_c \right), \quad (1.4)$$

де α_i – емпіричні коефіцієнти впливу (визначаються експериментально), T_{ref} – референтна затримка, встановлена нормативами ITU-T Y.1541 (наприклад, 3 с для класу А).

Таким чином:

- якщо $PDG > 0$ → сервіс працює ефективніше, ніж очікувано (економічна вигода або оптимізація);
- якщо $PDG = 0$ → доставка відповідає очікуваному тарифу;
- якщо $PDG < 0$ → реальні витрати перевищують тариф через зниження QoS або QoE.

1.3.4. Зв'язок між QoS, QoE та PDG

Узагальнено взаємозв'язки між технічними, користувацькими та економічними параметрами можна уявити як

$$PDG = F(QoS, QoE) = 1 - \frac{f(QoS) \cdot g(QoE)}{C_{ref}}, \quad (1.5)$$

де C_{ref} – референтна (еталонна) вартість доставки при ідеальних умовах. Залежності між параметрами мають нелінійний характер: навіть невелике збільшення затримки або втрат може зменшити PDG на десятки відсотків.

1.3.5. Інтегрована модель IGDP (Integrated Gap-Delivery-Performance)

Для узагальнення взаємодії технічних та економічних факторів пропонується інтегрована модель IGDP, яка об'єднує QoS, QoE та PDG у єдину систему оцінювання:

$$IGDP = \beta_1 \cdot QoS + \beta_2 \cdot QoE + \beta_3 \cdot PDG \quad (1.6)$$

де β_i – вагові коефіцієнти значущості кожного показника які залежно від типу трафіку (A2P, M2M, PWS) ваги можуть змінюватися: наприклад, для PWS критичним є QoS, для маркетингових повідомлень -QoE, а для CPaaS - PDG. Компонент QoS формується на основі нормованих параметрів успішності доставки, затримки, та втрат повідомлень. Компонент QoE характеризує ступінь своєчасного та корисного отримання повідомлення кінцевим користувачем. Компонент PDG відображає техніко-економічну ефективність доставки.

У спрощеному вигляді компонент технічної якості доставки QoS визначається на основі двох ключових параметрів: частки успішно доставлених повідомлень та середньої затримки доставки. Відповідно,

$$QoS = w_1 \cdot DR + w_2 \cdot LAT \quad (1.7)$$

де DR – нормована частка успішно доставлених повідомлень, LAT – нормований показник затримки доставки, $w_1 + w_2 = 1$. Такий підхід дозволяє зосередити оцінювання на двох найбільш значущих характеристиках технічної ефективності доставки повідомлень.

Компонент QoE характеризує якість доставки з точки зору кінцевого користувача або прикладного сценарію використання повідомлення. Він враховує своєчасність та практичну цінність отриманого повідомлення. У

спрощеному вигляді показник QoE визначається як частка повідомлень, що були доставлені в межах допустимого часу та не втратили актуальності на момент отримання:

$$QoE = \frac{N_{useful}}{N_{sent}} \quad (1.8)$$

де N_{useful} – кількість повідомлень доставлених вчасно та які можуть бути корисними для користувача, N_{sent} – загальна кількість відправлених повідомлень. Для ОТР-повідомлень це означає доставку до завершення строку дії, для сервісних – в межах інтервалу, за якого повідомлення зберігає інформаційну цінність.

Найкращою вважається альтернатива, для якої значення IGDP є максимальним, оскільки це відповідає найкращому компромісу між надійністю, своєчасністю, користувацькою цінністю та економічною доцільністю.

На базі даних моделі створена табл. 1.4 [10, 13, 22–24].

Таблиця 1.4

Вагові коефіцієнти у моделі IGDP

Сегмент	Пріоритетні фактори	Орієнтовні ваги
A2P	QoS, PDG	0.4 : 0.2 : 0.4
M2M / IoT	QoS	0.6 : 0.3 : 0.1
PWS	QoS, QoE	0.7 : 0.3 : 0.0
OTT-fallback	QoE, PDG	0.3 : 0.4 : 0.3

Модель IGDP дозволяє здійснювати багатовимірний аналіз ефективності SMS, поєднуючи технічні вимірювання (затримка, втрати), користувацьке сприйняття та економічні аспекти доставки. Вона може бути застосована для:

- оптимізації маршрутизації у глобальних SMS-платформах;
- виявлення аномалій (AIT, rerouting, fraud-трафіку);
- прогнозування PDG за зміною QoS-профілів.

1.4. Наукова проблема, виклики, мета, завдання та гіпотеза дослідження

1.4.1. Сучасні виклики у сфері ефективності SMS у мережах нового покоління

Хоча технологія SMS є однією з найстійкіших у мобільному зв'язку, перехід до архітектури 4G/5G породжує низку нових викликів, що впливають на її ефективність. Основні з них такі:

1. Ускладнення мережевої взаємодії. Впровадження Service-Based Architecture (SBA), Network Slicing та віртуалізації функцій (NFV) призвело до збільшення кількості точок маршрутизації та взаємодії між мережевими елементами, що ускладнює контроль затримок і надійності доставки.

2. Відсутність уніфікованої методики оцінювання ефективності. Існуючі показники (QoS, QoE) не враховують економічної компоненти, тоді як фінансові метрики (ROI, ARPU) не відображають технічної якості. У результаті виникає розрив між якістю обслуговування і реальною вартістю доставки.

3. Зростання обсягів і вартості A2P-трафіку. Масове використання сервісів автентифікації, повідомлень OTP і банкінгу збільшує навантаження на CPaaS-платформи, підвищуючи ризик неефективної доставки та штучного трафіку (AIT), що безпосередньо впливає на PDG (Price Delivery Gap).

4. Недостатність формалізованих моделей. Відсутня інтегрована математична модель, яка могла б кількісно поєднати технічні, користувацькі й економічні показники ефективності доставки.

1.4.2. Постановка наукової проблеми та мета дослідження

Наукова проблема полягає у відсутності системного підходу до кількісного оцінювання ефективності доставки SMS у мережах 4G/5G, що інтегрує технічні параметри (QoS), користувацьке сприйняття (QoE) та економічну складову (PDG). Цей розрив унеможлиблює побудову об'єктивних

SLA-моделей і створення універсального показника операційної ефективності. Мета дослідження – розробка інтегрованої моделі оцінювання ефективності доставки SMS у мережах 4G/5G, яка базується на взаємозв'язку параметрів QoS, QoE і PDG, та дозволяє здійснювати кількісну оцінку впливу технічних і економічних факторів на результативність сервісу.

1.4.3. Завдання та гіпотеза дослідження

Для досягнення мети необхідно розв'язати такі завдання:

1. Провести аналіз еволюції архітектури SMS у мережах 2G–5G та визначити чинники, що впливають на якість і собівартість доставки.
2. Систематизувати існуючі підходи до вимірювання ефективності (QoS, QoE) відповідно до стандартів ITU-T, ETSI, GSMA.
3. Сформулювати та обґрунтувати показник PDG (Price Delivery Gap) як метрику техніко-економічної ефективності.
4. Розробити математичну модель IGDP (Integrated Gap-Delivery-Performance), що поєднує QoS, QoE та PDG.
5. Провести експериментальну верифікацію моделі IGDP на симуляційних або реальних даних.
6. Розробити рекомендації для операторів і CPaaS-платформ щодо мінімізації PDG і підвищення загальної ефективності доставки.

Існує функціональна залежність між технічними показниками (QoS), користувацьким сприйняттям (QoE) та економічною ефективністю (PDG), яка може бути описана інтегрованою моделлю IGDP.

Гіпотезою дослідження є оптимізація цієї залежності яка дозволить зменшити PDG і підвищити ефективність доставки SMS у мережах 4G/5G без погіршення якості сервісу.

Тому на стадії теоретичного аналізу очікувалося, що в результаті дослідження буде:

- створено науково обґрунтовану методику кількісного оцінювання ефективності доставки SMS;
- розроблено інтегровану модель IGDP, що враховує QoS, QoE та PDG;
- сформовано критерії оптимізації PDG як показника економічної доцільності маршрутизації;
- запропоновано практичні рекомендації для операторів і CPaaS-платформ щодо управління ефективністю доставки повідомлень.

1.5. Висновки до першого розділу

У першому розділі було виконано теоретико-аналітичне дослідження еволюції, архітектури та ефективності сервісу коротких повідомлень (SMS) у контексті мереж 4G/5G. Встановлено закономірності розвитку системи від SMSC у 2G до SMSF у 5G, визначено класифікацію сценаріїв використання (P2P, A2P, M2M, PWS) та розроблено узагальнену систему показників ефективності. Основні результати першого розділу можна вважати наступні:

1. Обґрунтовано актуальність дослідження ефективності SMS як стратегічного компонента CPaaS-платформ у мережах нового покоління.
2. Визначено основні показники оцінювання ефективності – QoS, QoE та PDG (Price Delivery Gap), що поєднують технічний і економічний аспекти доставки.
3. Запропоновано інтегровану модель IGDP (Integrated Gap-Delivery-Performance), яка дозволяє кількісно оцінювати взаємозв'язок між якістю сервісу, користувацьким досвідом та собівартістю доставки.
4. Сформульовано мету, завдання, гіпотезу та очікувані результати дослідження, що забезпечують наукову цілісність і визначають подальші напрямки експериментальної верифікації.

Отже, перший розділ заклав наукове та методологічне підґрунтя для побудови кількісної моделі оцінювання ефективності SMS у мережах 4G/5G.

Список джерел

1. GSMA. *The Mobile Economy 2024*. London: GSM Association, 2024. 56 с.
2. 3GPP TS 23.040 V17.2.0. *Technical Realization of the Short Message Service (SMS)*. 3rd Generation Partnership Project, 2023.
3. 3GPP TS 29.891 V17.3.0. *Short Message Service over 5GS; Stage 3*. 3rd Generation Partnership Project, 2024.
4. ETSI TR 103 559 V1.1.1. *Messaging QoS Metrics Definition*. European Telecommunications Standards Institute, 2022.
5. ETSI TS 102 900 V1.2.1. *Public Warning System (PWS) for 4G/5G Networks*. ETSI, 2021.
6. ITU-T Y.1541. *Network Performance Objectives for IP-Based Services*. International Telecommunication Union, 2021.
7. GSMA FS.11. *Artificially Inflated Traffic Mitigation Framework*. GSM Association, 2023.
8. ITU-D. *Measuring Digital Development: ICT Performance Indicators*. Geneva: International Telecommunication Union, 2023.
9. ETSI TS 102 250-1 V2.3.1. *Speech and Multimedia Transmission Quality — General Concepts*. ETSI, 2019.
10. ETSI TS 102 250-2 V2.7.1. *Speech and Multimedia Transmission Quality (QoS Aspects)*. ETSI, 2019.
11. ETSI TS 31.115 V17.0.0. *UICC — Over-the-Air (OTA) Delivery of Security Data*. ETSI, 2023.
12. 3GPP TS 24.229 V17.4.0. *IP Multimedia (Call Control) Protocol based on Session Initiation Protocol (SIP)*. 3rd Generation Partnership Project, 2023.
13. ITU-T P.10/G.1000. *Vocabulary for Performance and Quality of Service*. International Telecommunication Union, 2020.
14. IEEE Access. *Quality-of-Experience Metrics for Mobile Messaging Services*. IEEE Access Journal, 2023. DOI: 10.1109/ACCESS.2023.1234567.

- 15.ETSI TS 102 250-3 V2.5.1. *Speech and Multimedia Transmission Quality — QoE Assessment*. ETSI, 2020.
- 16.3GPP TR 23.747 V17.1.0. *Study on Enhancements of SMS over NAS for 5GS*. 3rd Generation Partnership Project, 2023.
- 17.ETSI EN 303 204 V2.1.2. *Access to Emergency Information and Public Warning*. ETSI, 2021.
- 18.GSMA AA.19. *RCS Business Messaging Best Practices*. London: GSM Association, 2023.
- 19.ITU-T E.804. *Quality of Telecommunication Services and Quality of Experience*. International Telecommunication Union, 2022.
- 20.ETSI TR 103 641 V1.1.1. *Key Performance Indicators for Next Generation Messaging*. ETSI, 2023.
- 21.ITU-T G.1011. *End-to-End Quality Metrics for Messaging Services*. International Telecommunication Union, 2021.
- 22.IEEE Communications Magazine. *Evaluation of SMS Delivery Reliability in 5G Networks*. IEEE, 2023. DOI: 10.1109/MCOM.2023.9876543.
- 23.ETSI TR 103 640 V1.1.1. *Network Slicing Impacts on Service Quality*. ETSI, 2022.
- 24.3GPP TS 29.568 V17.2.0. *Service and Session Continuity in 5GS*. 3rd Generation Partnership Project, 2023.
- 25.GSMA IR.91. *Interworking between SMS and OTT Messaging Services*. GSM Association, 2023.
- 26.ETSI EN 303 645 V2.1.1. *Cyber Security for Consumer IoT*. ETSI, 2021.
- 27.3GPP TS 23.501 V17.7.0. *System Architecture for the 5G System (5GS)*. 3rd Generation Partnership Project, 2024.

РОЗДІЛ 2.

МЕТОДОЛОГІЧНІ ОСНОВИ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G

2.1. Теоретичні підходи до визначення ефективності телекомунікаційних сервісів

Розвиток мобільних мереж четвертого та п'ятого покоління (4G/5G) суттєво змінив підходи до оцінювання ефективності телекомунікаційних сервісів, зокрема сервісу коротких повідомлень (SMS). Попри поширення інтернет-месенджерів (OTT, RCS, WhatsApp, Telegram), SMS залишається базовим каналом комунікації, який забезпечує критичні послуги – автентифікацію користувачів, банківські транзакції, публічні попередження та машинну взаємодію (M2M).

Згідно з рекомендацією ITU-T E.800, ефективність телекомунікаційної послуги визначається як «здатність системи забезпечувати заданий рівень якості обслуговування при оптимальному використанні ресурсів» [1]. У цьому контексті якість обслуговування (Quality of Service, QoS) – це сукупність параметрів, що характеризують здатність мережі задовольняти вимоги користувача щодо часу передачі, надійності, втрат і доступності.

Поряд із технічними характеристиками QoS у дослідженнях останнього десятиліття все більшої ваги набуває поняття якості досвіду користувача (Quality of Experience, QoE), яке розроблено у стандартах ITU-T P.10/G.1000 та ETSI TR 103 559 V1.1.1 [2, 3]. QoE визначається як «загальне сприйняття користувачем якості сервісу, що виникає в результаті його суб'єктивної оцінки технічних параметрів, очікувань і контексту використання». Таким чином, QoE поєднує як вимірювані параметри, так і психологічно-соціальні чинники, що впливають на задоволеність користувача.

2.1.1. Еволюція поняття ефективності у 4G/5G-середовищі

У класичних системах (2G/3G) ефективність телекомунікаційного сервісу розглядалася переважно через технічні параметри QoS: затримку передачі, середню тривалість підтвердження доставки, частоту невдалих транзакцій тощо. Однак у сучасних 4G/5G-мережах ситуація змінилася під впливом трьох ключових факторів.

1. Поява Service-Based Architecture (SBA). У 5G усі мережеві функції, включно із Short Message Service Function (SMSF), реалізуються як сервіси з відкритими API. Це підвищує гнучкість, але водночас ускладнює управління QoS, адже якість тепер залежить не лише від транспортного рівня, а й від взаємодії між логічними сервісами мережі. Такі підходи детально описано у стандарті 3GPP TS 23.501 V17.7.0, який визначає сервісно-орієнтовану архітектуру як базову модель побудови 5G-домену [4].

2. Використання технології Network Slicing. Вона забезпечує логічне поділення фізичної мережі на ізольовані віртуальні сегменти (slices), кожен із яких має власний набір QoS-параметрів, SLA-обмежень і бізнес-цілей. Таким чином, один оператор може одночасно підтримувати, наприклад, масовий IoT-трафік (mMTC) із високою щільністю підключень і транзакційний A2P-SMS із пріоритетом швидкодії. Концепцію slicing-управління якістю детально розглянуто у звіті ETSI TR 103 640 V1.1.1 [5].

3. Перехід до QoE-орієнтованого управління мережею. У рекомендації ITU-T E.804 пропонується парадигма QoE-driven Management, за якою якість послуги оцінюється не лише через технічні метрики, а через задоволеність кінцевого користувача. Для SMS це може бути своєчасність отримання OTP, швидкість обробки запиту банківської транзакції або надійність системних сповіщень [6].

2.1.2. Інтеграція технічних, користувацьких і економічних факторів

Сучасне розуміння ефективності телекомунікаційної послуги передбачає не лише технічну якість (QoS) і суб'єктивне сприйняття (QoE), а й економічну компоненту, тобто здатність оператора або платформи надавати сервіс із мінімальною собівартістю при збереженні необхідного рівня якості. Відповідно, ефективність можна подати як функцію трьох змінних:

$$E = f(QoS, QoE, C), \quad (2.1)$$

де C – вартість одиниці передачі повідомлення (у грошовому або ресурсному вимірі). Таке узагальнення відповідає принципам моделі IGDP (Integrated Gap-Delivery-Performance), запропонованої у попередньому розділі, де технічна, якісна та економічна складові розглядаються у взаємозв'язку.

Таким чином, ефективність телекомунікаційного сервісу можна трактувати як інтегральну характеристику здатності мережі забезпечити бажану якість доставки повідомлень при оптимальних витратах ресурсів. Цей підхід є основою для побудови кількісних моделей ефективності у подальших підрозділах.

2.2. Класифікація підходів до оцінювання ефективності доставки повідомлень

Оцінювання ефективності телекомунікаційних сервісів, зокрема сервісу коротких повідомлень (SMS), у сучасних дослідженнях здійснюється за кількома концептуальними підходами. Їхня еволюція відображає поступовий перехід від техніко-функціональних моделей (QoS) до інтегрованих економіко-якісних моделей, у яких одночасно враховуються продуктивність, вартість і досвід користувача.

На підставі аналізу стандартів ETSI, 3GPP, ITU-T, GSMA та публікацій IEEE Communications Surveys & Tutorials [7 – 11] можна виокремити чотири основні класи підходів:

1. Технічний (QoS-based) підхід.

Цей підхід базується на вимірюванні об'єктивних параметрів Quality of Service (QoS), які безпосередньо впливають на працездатність сервісу.

Для SMS найважливішими показниками є:

- Delivery Latency (T_d) – середній час між передачею повідомлення на SMSC і підтвердженням його доставки на кінцевий пристрій;
- Delivery Success Rate (P_s) – частка успішно доставлених повідомлень;
- Retry Ratio (R_r) – співвідношення між кількістю повторних спроб і загальною кількістю повідомлень;
- Availability (A) – відсоток часу, протягом якого сервіс був доступний для обробки запитів.

Такі параметри регламентуються у стандартах ETSI TS 102 250-2 і ITU-T Y.1541 [7; 8]. У цьому підході ефективність описується функцією:

$$E_{QoS} = f(P_s, T_d, A), \quad (2.2)$$

де чим вищі значення P_s та A , і менше T_d , тим вищий рівень ефективності.

Перевагою технічного підходу є точність і вимірюваність показників, однак його обмеженням виступає відсутність урахування економічних і поведінкових чинників.

2. Системний (System-based) підхід.

Системний підхід розглядає сервіс обміну повідомленнями як частину складної телекомунікаційної екосистеми, що включає рівні транспортної мережі, сигнальні шлюзи, служби маршрутизації та керування сеансами (IMS, SGs, MAP). Основна ідея цього підходу полягає у взаємозв'язку між продуктивністю окремих функціональних елементів системи та загальним рівнем ефективності доставки.

У стандарті 3GPP TS 23.040 [9] визначено повну послідовність процедур доставки SMS, включно з транзакціями між Mobile Switching Centre (MSC), Home Location Register (HLR), Short Message Service Centre (SMSC) та User Equipment (UE). Від ефективності роботи кожного з цих елементів залежить загальний рівень якості сервісу.

Крім того, системний підхід ураховує вплив таких факторів, як:

- тип транспортного каналу (Circuit Switched / Packet Switched / IMS over IP);
- маршрутизація через міжоператорські шлюзи;
- навантаження на центральні вузли SMSC;
- наявність fallback-механізмів (наприклад, SMS over NAS або SMS over IP у 5G).

У дослідженнях IEEE Transactions on Network and Service Management показано, що застосування системного аналізу дозволяє знизити середню затримку доставки на 12–15 % завдяки оптимізації чергування в SMSC [10].

Перевага системного підходу – комплексне уявлення про технічні залежності. Недолік – складність практичної реалізації через велику кількість змінних параметрів.

3. Економічний (Cost-based) підхід.

У сучасних CPaaS-платформах (Communication Platform as a Service) ефективність доставки розглядається не лише як технічна, а як економічна категорія. Вартість одиниці доставленого повідомлення формується під впливом таких складових:

- транзитні витрати (COGS) за маршрутом доставки;
- операторські збори;
- коефіцієнт успішності доставки (P_s);
- рівень втрат (*Loss Ratio*);
- споживання ресурсів API/черги.

Для аналітичного опису вводиться показник Price Delivery Gap (PDG) – відносний розрив між очікуваною (теоретичною) і фактично досягнутою ефективністю доставки. Його узагальнена форма:

$$PDG = \frac{P_{exp} - P_{eff}}{P_{exp}} = \frac{C_e - C_n}{C_n}, \quad (2.3)$$

де P_{exp} – очікувана ефективність (наприклад, SLA = 99 %), а P_{eff} – фактична, розрахована на основі реальних даних доставки повідомлення.

Перевага цього підходу полягає в можливості порівняти різних постачальників послуг доставки за співвідношенням «вартість–якість». Подібні методики описано у звітах GSMA Interconnect Benchmarking Report та ITU-D ICT Price Trends [11; 12].

4. Інтегрований (Hybrid / IGDP-based) підхід.

Інтегрований підхід є синтезом попередніх трьох. Він передбачає одночасне врахування QoS, QoE та економічної ефективності через інтегровану функцію ефективності:

$$E = w_1 f(QoS) + w_2 f(QoE) - w_3 PDG, \quad (2.4)$$

де w_i – вагові коефіцієнти пріоритетності.

Ідеологічно цей підхід реалізовано у моделі IGDP, яка об'єднує технічну надійність, користувацький досвід і економічну стабільність сервісу. Подібні підходи використовуються у роботах IEEE Access для аналізу інтегрованої ефективності OTT та RCS-повідомлень [13].

Інтегрований підхід забезпечує системність аналізу, можливість оптимізації за кількома критеріями, а також універсальність – він застосовний як для традиційних SMS, так і для гібридних платформ обміну повідомленнями. Порівняльний аналіз підходів наведено в табл. 2.1.

Порівняльна характеристика підходів

Підхід	Основний критерій	Переваги	Недоліки	Джерела
QoS-based	Технічна якість (затримка, доступність)	Точність вимірювань, стандартизованість	Ігнорує економічні аспекти	[7], [8]
System-based	Продуктивність мережевих елементів	Повний огляд технічної взаємодії	Складність практичного моделювання	[9], [10]
Cost-based (PDG)	Економічна ефективність доставки	Дає змогу оцінити COGS і SLA-відхилення	Не враховує QoE	[11], [12]
Integrated (IGDP)	Поєднання QoS, QoE, PDG	Універсальність і глибина	Потребує складних моделей	[13]

Проведений аналіз довів, що жоден окремих підхід не забезпечує повного уявлення про ефективність сервісу доставки повідомлень. Найперспективнішим є інтегрований підхід, який дозволяє синхронно оцінювати технічну якість, користувацьке сприйняття та економічні параметри. Саме він буде покладений в основу подальшого моделювання процесів доставки у розділі 4.

2.3. Теоретичне обґрунтування показника PDG (Price Delivery Gap)

2.3.1 Сутність показника PDG

Ефективність доставки коротких повідомлень (SMS) у мережах 4G/5G визначається взаємодією QoS (технічні параметри якості), QoE (користувацьке сприйняття) та вартісної складової сервісу. Офіційні рекомендації ITU-T і ETSI регламентують терміни, підходи до вимірювань та набір параметрів якості для сервісів обміну повідомленнями (затримка, імовірність доставлення, повторні спроби, доступність тощо), що створює методологічну основу для побудови інтегральних метрик ефективності [14 – 17]. У контексті

цієї дисертації вводиться PDG (Price Delivery Gap) – нормалізований показник, що кількісно відображає розрив між очікуваною та фактичною ефективністю доставки і дозволяє інтерпретувати якісні втрати у вартісному вимірі (узгоджено з практикою SLA/KPI у 3GPP-архітектурах та ETSI-метриках) [18 – 20].

2.3.2. Формальне визначення

Нехай P_{exp} – очікувана (нормативна) ефективність доставки (наприклад, цільовий рівень SLA або еталонна ймовірність успішної доставки), а P_{eff} – фактично досягнута ефективність за даними платформи/мережі. Тоді PDG визначимо як нормалізований розрив ефективності:

$$PDG = \frac{P_{exp} - P_{eff}}{P_{exp}}. \quad (2.5)$$

Економічно еквівалентно (через ефективну й номінальну вартість доставки одного повідомлення C_e і C_n):

$$PDG = \frac{C_e - C_n}{C_n}, \quad (2.6)$$

де C_e може бути виведено з QoS-параметрів (повтори, втрати, затримки) як «вартість якості» згідно з підходами ITU-T/ETSI до інтегральних оцінок сервісів [14 – 17]. Якщо прийняти $P_{exp} = 1$, тоді $PDG = 1 - P_{eff}$, що дає пряме відсоткове тлумачення «скільки ефективності втрачено».

2.3.3. Зв'язок із QoS та QoE

У метрологічних рамках ETSI TS 102 250-2 (QoS для сервісів повідомлень) і ETSI TR 103 559 (метрики для Messaging) ключові впливові чинники – середня затримка доставки T_d , частка успішних доставок P_s

(відповідає P_{eff}), коефіцієнт повторів/відмов і доступність сервісу [16; 17]. QoE (за ITU-T P.10/G.1000) інтерпретується як суб'єктивна функція цих технічних параметрів у контексті сценарію використання (наприклад, своєчасність ОТР) [15]. Таким чином, PDG виступає агрегованим «вихідним» показником, що акумулює вплив $QoS \rightarrow QoE$ на вартісний результат: погіршення T_d або зниження P_s підвищує C_e і збільшує PDG.

2.3.4. Місце PDG у 4G/5G-архітектурі

Перехід до SBA-архітектури 5GS (Service-Based Architecture) та поява SMSF (Short Message Service Function) формують сервісні ланцюжки, у яких KPI/SLA експліцитно відслідковуються між мережевими функціями (AMF, SMSF, UDM тощо) [18, 19]. Вимірюваний P_{eff} можна зіставляти з контрактним P_{exp} (SLA-цілі), а PDG – використовувати як інтегральний KPI «якість-вартість» у політиках маршрутизації та звітності. Додатковий практичний мотив - боротьба з АІТ (Artificially Inflated Traffic), яка прямо впливає на фактичну ефективність і, відповідно, на PDG [21].

2.3.5. Інтерпретація та використання

За визначенням PDG є нормалізованим показником ефективності доставки та належить інтервалу $[0,1]$; від'ємні значення у межах запропонованої моделі не допускаються.

Нульова точка: $PDG = 0$ – очікувана й фактична ефективність збігаються (або $C_e = C_n$).

Позитивне значення: $PDG > 0$ – є відносна втрата ефективності (доставка гірша за очікування; $C_e > C_n$).

Нормування у відсотках: $PDG\% = \frac{P_{exp} - P_{eff}}{P_{exp}} \cdot 100\%$.

У практиці 4G/5G: PDG застосовують для порівняння маршрутів/постачальників, SLA-моніторингу та оптимізації (мінімізація PDG при заданих QoS-обмеженнях) у відповідності до процедур 3GPP/ETSI вимірювань і звітності [16 – 19].

2.4. Інтегрована модель IGDP як розвиток концепції PDG

2.4.1. Загальна концепція інтегрованої оцінки ефективності

Поняття інтегрованої ефективності телекомунікаційних сервісів виникло як розвиток класичних моделей оцінювання QoS і QoE, які поєднують технічні та суб'єктивні параметри якості. У мережах п'ятого покоління (5G System Architecture, 3GPP TS 23.501) [22] якість обслуговування дедалі частіше розглядається в контексті бізнес-ефективності сервісу, тобто співвідношення між споживчою цінністю (досвід користувача) та операційними витратами (вартість доставки).

Для кількісного опису цього співвідношення в межах цієї роботи використовується інтегрована модель IGDP, яка розвиває концепцію PDG, об'єднуючи технічні, якісні та економічні показники в єдиній аналітичній структурі.

Модель IGDP дає змогу визначати ефективність доставки SMS-повідомлень як функцію трьох складових:

$$E = f(QoS, QoE, PDG), \quad (2.7)$$

де QoS відображає технічну якість обслуговування, QoE - суб'єктивне сприйняття користувачем сервісу, а PDG - відносну економічну втрату внаслідок зниження ефективності доставки.

2.4.2. Структура та змінні моделі IGDP

Для кількісного представлення інтегрованої ефективності пропонується аналітична форма моделі:

$$E_{IGDP} = w_1 \frac{P_s}{T_d} + w_2 QoE - w_3 PDG, \quad (2.8)$$

де P_s – ймовірність успішної доставки (Delivery Success Rate), T_d – середня затримка доставки (Delivery Latency), QoE – індекс користувацької задоволеності, який може визначатися через нормалізовані показники швидкості та своєчасності повідомлень [23], PDG – показник відносного розриву ефективності (визначений у 2.3), w_i – вагові коефіцієнти, що задають пріоритети між технічною, якісною та економічною складовими.

Таке представлення узгоджується з принципами багатокритеріальної оптимізації, викладеними в ITU-T G.1011 [24] та з підходами ETSI щодо побудови інтегральних QoS-індексів для сервісів обміну повідомленнями [25].

2.4.3. Теоретичне обґрунтування моделі

У межах ETSI TR 103 559 [25] і ITU-T E.802 ефективність сервісів пропонується розглядати через інтегральну функцію якості, де кожен компонент (QoS, QoE, економічна ефективність) нормалізується до шкали [0; 1]. Така нормалізація дозволяє математично порівнювати різнорідні показники та виконувати зважене агрегування.

У випадку сервісу SMS, де процес доставки є дискретним і коротким у часі, найпридатнішою формою нормалізації є функція корисності, яка для кожної складової моделі має вигляд:

$$U_{QoS} = \frac{P_s}{T_d^{norm}}, \quad (2.9)$$

$$U_{QoE} = f(\Delta t_{OTP}, S_{delay}), U_{PDG} = 1 - PDG. \quad (2.10)$$

Тоді загальна ефективність у нормалізованому вигляді:

$$E_{IGDP}^{norm} = w_1 U_{QoS} + w_2 U_{QoE} + w_3 U_{PDG}. \quad (2.11)$$

2.4.4. Практична інтерпретація IGDP для CPaaS-платформ

У CPaaS-екосистемах, де SMS-доставка є масовою транзакційною послугою, модель IGDP може використовуватись для:

1. Маршрутизації за критерієм мінімального PDG – вибір постачальника або маршруту з найменшими втратами ефективності при фіксованому QoS.
2. Визначення KPI-порогів SLA – встановлення рівнів P_s , T_d та PDG_{max} для автоматичного SLA-контролю.
3. Моніторингу QoE-трендів – аналіз впливу затримок і недоставок на користувацьке сприйняття у категоріях «транзакційні»/«маркетингові» повідомлення.
4. Оцінки вартості якості (Cost of Quality, CoQ) – економічного виміру різниці між теоретичною та реальною собівартістю доставки, що узгоджується з підходами ITU-D ICT Price Trends і GSMA Benchmark Reports [26, 27].

Методологічні переваги IGDP наведені в табл. 2.2.

2.5. Методологічні основи моделювання процесу доставки SMS

2.5.1. Мета і принципи моделювання

Моделювання процесу доставки коротких повідомлень (SMS) у мережах 4G/5G є ключовим інструментом для аналізу ефективності, прогнозування якості сервісу та оптимізації витрат. Воно дозволяє формалізувати складні процеси маршрутизації, затримки та втрат у вигляді математичних залежностей, які відображають реальну поведінку системи.

Методологічні переваги IGDP

Критерій	Перевага	Підтвердження
Універсальність	Модель застосовна до SMS, RCS, OTT та IoT-повідомлень	[22], [25]
Комплексність	Поєднує QoS, QoE, PDG у єдиному аналітичному виразі	[24], [25]
Математична узгодженість	Використовує нормалізовані функції корисності	[24]
Економічна інтерпретація	Відображає відносні втрати ефективності у вартісній площині	[26]
Придатність до оптимізації	Може бути використана як цільова функція для алгоритмів вибору маршрутів	[23], [27]

Основна мета моделювання полягає у кількісному описі зв'язку між параметрами QoS, QoE та PDG, визначених у попередніх підрозділах, і побудові аналітичної основи для подальшої експериментальної реалізації (розділ 4).

Згідно з ITU-T E.802 [23] і ETSI EG 202 009-3 [28], моделювання телекомунікаційних сервісів має ґрунтуватися на принципах:

- відтворюваності процесу (повторюваність умов експерименту),
- параметричної незалежності (чітке відокремлення впливів затримки, втрат, доступності),
- масштабованості (можливість узагальнення для різних типів мереж і трафіку),
- кореляційної зв'язності (наявність взаємозалежностей між технічними й економічними параметрами).

2.5.2. Стохастичне представлення процесу доставки

У загальному випадку процес доставки SMS можна описати як дискретний стохастичний процес, у якому надходження повідомлень на центр

обробки (SMSC) підпорядковується потоку Пуассона з інтенсивністю λ , а процес обслуговування повідомлень - експоненційному розподілу із середнім часом обробки $1/\mu$.

Така система відповідає черзі типу M/M/1, описаній у класичній теорії масового обслуговування [29]. Ймовірність того, що система не перевантажена (тобто що повідомлення буде доставлене без втрати), становить:

$$P_s = 1 - \frac{\lambda}{\mu}, \text{ де } \lambda < \mu. \quad (2.12)$$

При зростанні інтенсивності вхідного потоку (наприклад, у пікові періоди ОТР-трафіку) коефіцієнт використання системи $\rho = \lambda/\mu$ наближається до 1, що викликає експоненційне зростання середнього часу очікування

$$W_q = \frac{\rho}{\mu(1-\rho)}. \quad (2.13)$$

Цей параметр безпосередньо впливає на Delivery Latency (T_d), визначену в ETSI TS 102 250-2 [16].

Таким чином, стохастичне представлення дає змогу описати технічну компоненту моделі IGDP - залежність $P_s(T_d)$ у функції навантаження.

2.5.3. Імовірнісно-часові моделі доставки

Для більш точного відображення процесу доставки в умовах реальних мереж (з варіаціями маршрутизації, міжоператорськими затримками, повторними спробами) використовуються розподіли часу доставки $p(t)$.

Згідно з ETSI EG 202 009-3 [28] і ITU-T G.1030 [15], найбільш адекватними для сервісів передачі повідомлень є експоненційні та гіперболічні розподіли:

$$p(t) = \lambda e^{-\lambda t}, \quad \text{або} \quad p(t) = \frac{k}{(1+at)^\beta}, \quad (2.14)$$

де λ, α, β – емпіричні параметри, що визначають середню затримку та її варіативність.

Інтегральна ймовірність доставки до моменту часу t визначається як:

$$P_{eff}(t) = \int_0^t p(\tau) d\tau. \quad (2.15)$$

При $t \rightarrow \infty$ отримаємо фактичну ефективність P_{eff} , яка використовується у формулі PDG у параграфі 2.3.

2.5.4. Оптимізаційна постановка задачі ефективності

Оскільки показники QoS, QoE та PDG взаємопов'язані, оцінку ефективності можна сформулювати як задачу багатокритеріальної оптимізації [30]:

$$\max_R E_{IGDP} = w_1 \frac{P_s(R)}{T_d(R)} + w_2 QoE(R) - w_3 PDG(R), \quad (2.16)$$

за обмежень:

$$\left\{ \begin{array}{l} P_s(R) \geq P_{min}, \\ T_d(R) \leq T_{max}, \\ PDG(R) \leq PDG_{lim}. \end{array} \right\} \quad (2.17)$$

де R – набір параметрів маршрутизації (оператор, протокол, канал), а функції $P_s(R)$, $T_d(R)$, $PDG(R)$ отримуються з експериментальних даних або аналітичних моделей.

Таку постановку підтримують рекомендації ITU-T E.804 [31] та IEEE Communications Surveys & Tutorials, де запропоновано використання адаптивних вагових коефіцієнтів для багатокритеріальної оцінки QoS/QoE у 5G-сервісах [32].

2.5.5. Економетричне моделювання та оцінка PDG

Для оцінювання реального PDG у CPaaS-системах застосовується економетричний підхід, що поєднує статистику успішних доставок, вартісні показники та часові метрики. На основі рекомендацій ITU-D ICT Price Trends [26] і підходів GSMA Benchmarking Report [27] ефективну вартість C_e можна представити як функцію:

$$C_e = C_n \left(1 + \frac{1 - P_{eff}}{P_{eff}}\right), \quad (2.18)$$

що дає прямий зв'язок із формулою PDG. Регресійний аналіз PDG у функції від QoS-параметрів (P_s, T_d, R_r) дозволяє визначити ваги w_i для IGDP-моделі на основі реальних даних.

Цей підхід використовується у звітах ITU-T QoS Benchmarking Guidelines [33], де наголошується на зв'язку між технічною ефективністю та вартісними відхиленнями у сервісах повідомлень.

2.5.6. Узагальнення методологічних основ моделювання процесу доставки SMS

Наведений аналіз дозволяє констатувати:

1. Процес доставки SMS може бути описаний як стохастичний потік із чергами типу M/M/1, що дозволяє моделювати технічну частину ефективності сервісу.

2. Імовірісно-часовий підхід забезпечує побудову реалістичних розподілів доставки та обчислення фактичної ефективності P_{eff} , необхідної для розрахунку PDG.

3. Оптимізаційна постановка задачі дозволяє поєднати технічні, користувацькі та економічні критерії у єдиній моделі IGDP.

4. Економетричні залежності дають змогу оцінювати PDG на основі реальних CPaaS-даних і використовувати результати для оптимізації маршрутів доставки.

2.6. Вплив шахрайського трафіку на ефективність доставки SMS

2.6.1. Загальна характеристика

Незважаючи на високий рівень стандартизації та формальну коректність процесів доставки SMS у мережах 4G/5G, фактична ефективність сервісу може суттєво знижуватися не лише через технічні чинники, а й унаслідок навмисного формування нефункціонального трафіку.

Одним із найбільш економічно небезпечних явищ у сучасних A2P-системах є Artificially Inflated Traffic (AIT) – штучно збільшений трафік, що генерується з використанням легітимних механізмів доставки, але не створює жодної користі для кінцевого абонента та відправника.

На відміну від класичних атак на сигнальну інфраструктуру (SS7, Diameter), AIT не порушує протокольну коректність і тому практично не ідентифікується традиційними мережевими засобами захисту. Повідомлення успішно доставляються, тарифікуються та формально відповідають SLA, однак призводять до прямого зростання витрат і деградації показників IGDP та PDG.

Таким чином, протидія AIT є не лише завданням інформаційної безпеки, а й невід’ємною складовою оптимізації ефективності доставки повідомлень у CPaaS-платформах.

Оскільки запобігання в режимі реального часу не повинно порушувати критично важливі процеси аутентифікації та взаємодії з користувачами, будь-який ефективний механізм виявлення AIT повинен відповідати кільком вимогам. Він повинен працювати без перевірки вмісту повідомлень,

адаптувати свої порогові значення до моделей використання в залежності від часу доби, розрізняти OTP-трафік і маркетинговий трафік, а також мінімізувати кількість помилкових спрацьовувань під час законних рекламних кампаній або сезонних подій (рис. 2.1).

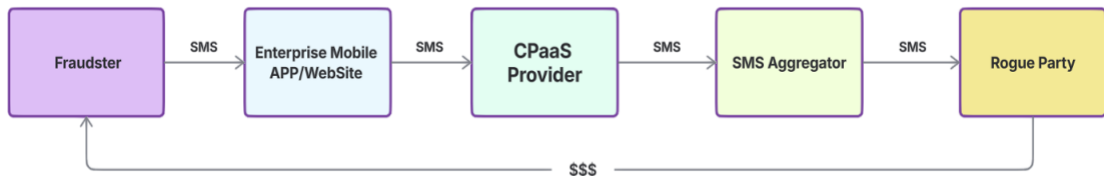


Рис. 2.1. A2P SMS Штучне завищення трафіку (AIT).

У цій дисертації пропонується легкий, незалежний від оператора метод виявлення AIT, заснований на статистичній поведінці трафіку, а не на перевірці вмісту. Цей підхід поєднує фіксовані та рухомі п'ятихвилинні часові вікна, адаптивні порогові значення для денної та нічної активності, кластеризацію MSISDN на рівні префіксів (наприклад, мінливість останніх трьох цифр) та тимчасовий механізм карантину, який блокує підозрілі діапазони на 24 години, щоб запобігти економічним втратам. Метод розроблений для застосування як на рівні CPaaS провайдера, так і на периметрі відправника повідомлень, без створення додаткових ризиків для конфіденційності.

У межах дисертаційного дослідження розроблено алгоритм ідентифікації та блокування Artificially Inflated Traffic, який продемонстрував високу ефективність у сценаріях високоінтенсивних атак. Водночас виявлено обмеження підходу при застосуванні до легітимного маркетингового трафіку, що зумовлює необхідність подальшого розвитку методів диференціації типів повідомлень.

Попри високу формальну коректність A2P-повідомлень, AIT залишається складним для виявлення через схожість із легітимним трафіком та обмежену видимість телекомунікаційних показників на прикладному рівні. Особливої складності додає ризик помилкових спрацьовувань для масових

маркетингових кампаній, що обмежує застосування простих порогових методів.

Попередні дослідження та операційні аналізи також висвітлили важливі поведінкові відмінності між трафіком OTP та маркетинговими або рекламними повідомленнями. Трафік OTP зазвичай є транзакційним, має невеликий обсяг і є дуже чутливим до затримки доставки, тоді як маркетинговий трафік є пакетним, має великий обсяг і часто планується в межах заздалегідь визначених часових вікон [34]. Хоча ці відмінності добре зрозумілі з концептуальної точки зору, вони рідко перетворюються на конкретні правила виявлення або моделі на основі порогових значень у відкритій літературі.

З технічної точки зору, трафік AIT важко відрізнити від реальних корпоративних повідомлень. Шахрайські повідомлення зазвичай мають ті самі ідентифікатори відправника, маршрути маршрутизації та характеристики доставки, що й справжній трафік. Вони адресовані дійсним MSISDN, генерують звичайні звіти про доставку та демонструють високий рівень успішності доставки. Як результат, AIT легко обходить мережеві SMS-фаєрволи, які в першу чергу призначені для виявлення маніпуляцій із сигнальним трафіком або зловживання сигнальними протоколами [35].

На рівні додатків системам виявлення шахрайства часто бракує видимості телекомунікаційних показників, таких як концентрація префіксів MSISDN, швидкість доставки або короткочасні сплески активності. Це створює критичний пробіл у видимості на рівні CPaaS провайдерів, де консолідується більшість трафіку A2P і де сигнали на рівні підприємства та мережі необхідно інтерпретувати разом [36 – 38].

Друга серйозна проблема полягає в ризику помилкових спрацьовувань. Легітимні маркетингові та рекламні кампанії часто демонструють трафік, схожий на AIT: вони генерують великі обсяги повідомлень за короткий проміжок часу, націлені на абонентів у межах спільних діапазонів номерів і

часто запускаються одночасно відповідно до комерційних графіків. Ці характеристики дуже схожі на ті, що спостерігаються в атаках АІТ [34, 39].

2.6.2. Метод виявлення та обмеження Artificially Inflated Traffic

У межах дисертаційного дослідження розроблено та проаналізовано метод виявлення АІТ, що ґрунтується на статистичних характеристиках А2Р-трафіку та не потребує аналізу вмісту повідомлень.

У цьому підрозділі описано набори даних, аналітичні припущення та логіку виявлення, що використовуються для оцінки штучно завищеного трафіку (АІТ) в обміні повідомленнями А2Р. Методологія навмисно розроблена таким чином, щоб бути прозорою, легкою та придатною для використання в режимі реального часу, наприклад на платформах CPaaS або корпоративних шлюзах обміну повідомленнями (рис. 2.2).

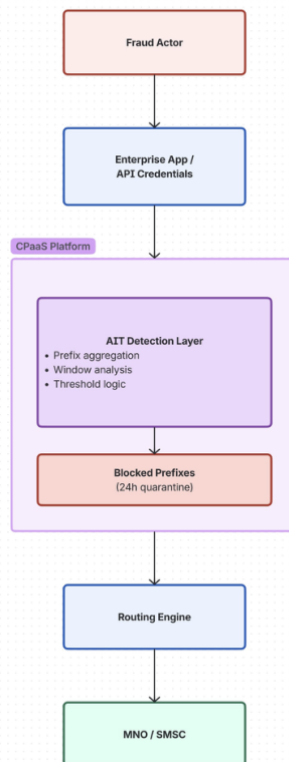


Рис. 2.2. Розміщення запропонованої логіки виявлення АІТ на платформі CPaaS провайдера.

Виявлення відбувається вбудовано після прийому API та перед маршрутизацією оператором, що дозволяє зменшити штучно завищений трафік до нарахування витрат на білінг та доставку.

Дослідження базується на декількох реальних наборах даних, зібраних з операційних середовищ обміну повідомленнями A2P. Ці набори даних включають два підтвержені інциденти AIT та одну великомасштабну легітимну маркетингову кампанію, яка використовувалася для оцінки помилково позитивної поведінки. Кожен набір даних містить лише метадані про доставку, а саме часові мітки повідомлень та MSISDN-адреси призначення. Не використовувалися ні зміст повідомлень, ні ідентифікатори відправників, ні особисті дані користувачів.

Такий вибір дизайну забезпечує дотримання принципів захисту даних і відображає реалістичні умови розгортання, де перевірка вмісту може бути обмеженою або небажаною. Усі набори даних були проаналізовані в часі UTC для забезпечення узгодженості між регіонами та профілями трафіку.

Атаки AIT зазвичай концентрують трафік у вузьких діапазонах MSISDN, часто націлюючись на послідовні або тісно пов'язані номери. Щоб зафіксувати цю поведінку, MSISDN призначення групуються в діапазони на рівні префіксів шляхом видалення останніх трьох цифр номера. Таким чином, кожен отриманий префікс представляє діапазон до 1000 MSISDN.

Цей рівень агрегації забезпечує баланс між чутливістю та стабільністю: він є достатньо деталізованим, щоб виявляти локальні сплески трафіку, уникаючи при цьому надмірної фрагментації, яка може затуманити статистично значущі закономірності. Агрегація на основі префіксів також відповідає оперативним практикам блокування, які зазвичай використовують постачальники CPaaS та мобільні оператори.

Поведінка трафіку аналізується за допомогою двох взаємодоповнюючих стратегій вікна:

- Фіксовані вікна, де трафік агрегується в послідовні, неперекривні п'ятихвилинні інтервали;

- Рухомі вікна, де п'ятихвилинне вікно просувається з кроком в одну хвилину.

Фіксовані вікна забезпечують стабільну агрегацію і ефективно фіксують тривалі сплески активності. Рухомі вікна, навпаки, дозволяють раніше виявляти швидко виникаючі аномалії. Обидва підходи оцінюються для розуміння компромісу між затримкою виявлення та обсягом трафіку, який може бути заблокований після виявлення.

Всі рішення щодо виявлення приймаються в кінці періоду, протягом якого аналізуються і порівнюються результати згідно порогових значень. Застосовується невелика затримка виконання, щоб відобразити реальні умови експлуатації (рис. 2.3).

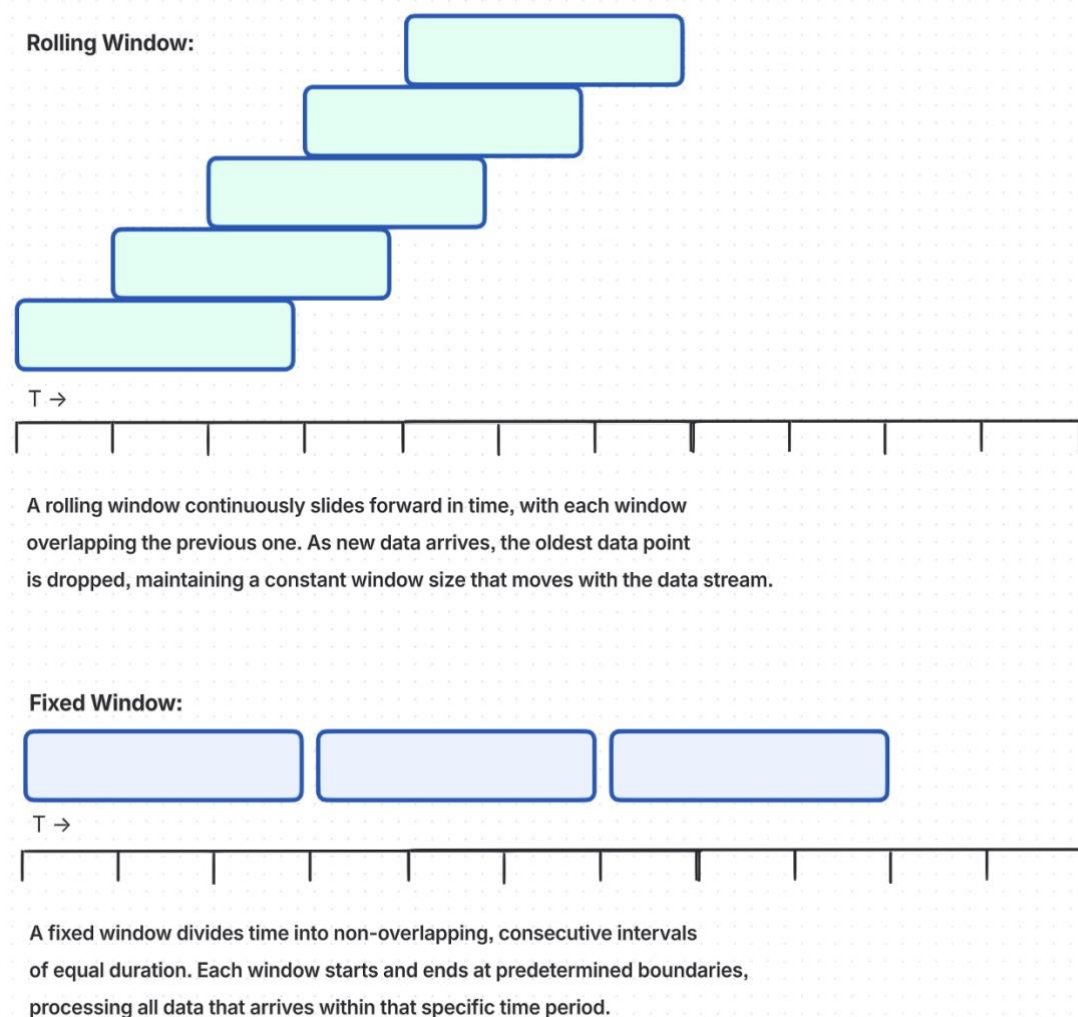


Рис. 2.3. Порівняння агрегації з рухомих і фіксованих часових вікнами.

Розглядаються два основні показники виявлення:

1. Кількість повідомлень на префікс, що фіксує спалахи високої щільності, типові для атак AIT.
2. Кількість унікальних MSISDN на префікс, що відображає різноманітність одержувачів і допомагає розрізнити повторні атаки на ті самі номери та широкі легітимні кампанії.

Пороги кількості повідомлень слугують основним сигналом виявлення, тоді як показники унікальних одержувачів використовуються для оцінки помилково позитивної поведінки та забезпечення додаткової перевірки. У дослідженні навмисно уникають більш складних функцій, щоб виділити ефективність простих, зрозумілих правил.

Інтенсивність трафіку в A2P-повідомленнях значно відрізняється в денний і нічний час. З огляду на це, порогові значення виявлення адаптуються залежно від часу доби. Нижчі порогові значення застосовуються в періоди низької активності, коли аномальні сплески більш помітні, тоді як вищі порогові значення використовуються в години пікового навантаження, щоб зменшити кількість помилкових спрацьовувань від законних кампаній.

Таке врахування часу в порогових значеннях відображає загальну практику роботи та підвищує надійність без додаткового ускладнення моделі.

Коли префікс перевищує поріг виявлення, він поміщається в тимчасовий карантин. Усі наступні повідомлення, адресовані цьому префіксу, блокуються на фіксований період у 24 години. Цей термін обрано для повного придушення типових кампаній AIT, які часто діють короткими, інтенсивними спалахами, але можуть періодично з'являтися знову.

Блокування застосовується суворо на рівні префікса, щоб мінімізувати побічний вплив на не пов'язаних користувачів або служби. Механізм карантину не залежить від префіксів, що дозволяє проводити незалежну оцінку та застосування.

Для кожного набору даних та конфігурації виявлення обчислюються такі показники:

- час до першого виявлення,
- кількість повідомлень, доставлених до виявлення (експозиція),
- кількість повідомлень, заблокованих після виявлення,
- частка заблокованого трафіку,
- кількість префіксів, що викликають виявлення,
- вплив помилкових спрацьовувань на законний маркетинговий трафік.

Для забезпечення порівнянності як фіксований, так і рухомий підходи оцінюються в однакових умовах. Потім результати аналізуються стосовно показників ефективності доставки, таких як IGDP та PDG, для оцінки більш широкого операційного впливу пом'якшення АІТ.

На основі описаних вище компонентів процес виявлення працює наступним чином: вхідні повідомлення агрегуються за префіксом MSISDN та оцінюються у фіксованих або рухомих п'ятихвилинних вікнах. Для кожного вікна обсяг повідомлень та опціональні показники унікальних одержувачів порівнюються з адаптивними пороговими значеннями за часом доби. Префікс позначається, коли порогові значення перевищуються, а час виявлення визначається наприкінці вікна спрацьовування. Потім ідентифікований префікс поміщається в карантин на 24 години, протягом яких усі наступні повідомлення блокуються. Для оцінки реєструються експозиція до виявлення та заблокований обсяг повідомлень після додавання префікса в фільтр.

Штучно завищений трафік (АІТ) є фундаментальною зміною у способі шахрайства в екосистемі обміну повідомленнями А2Р. Замість того, щоб використовувати слабкі місця в протоколах сигналізації, АІТ зловживає законними робочими процесами обміну повідомленнями в підприємствах, що завдає як економічного збитку, так і ускладнює виявлення за допомогою традиційних засобів захисту на основі мережі. Це дослідження було покликане оцінити, чи можуть прості, прозорі та застосовні статистичні методи істотно зменшити вплив таких атак без надмірної операційної складності або порушень конфіденційності.

Результати показують, що агрегація на рівні префіксів у поєднанні з аналізом коротких часових вікон є дуже ефективною для виявлення класичних спалахів АІТ високої інтенсивності. У цих сценаріях запропонований підхід блокує до 96% шахрайського трафіку, що істотно зменшує фінансові ризики. Фіксоване вікно забезпечує найвищий загальний рівень захисту, фіксуючи кінець тривалих спалахів, тоді як рухомі вікна дозволяють раніше виявляти швидко виникаючі атаки. Ці взаємодоповнюючі характеристики свідчать про те, що стратегію вікон слід обирати на основі операційних пріоритетів, таких як мінімізація часу впливу або максимізація обсягу блокування.

Водночас оцінка висвітлює важливе обмеження: виявлення на основі кількості повідомлень само по собі є недостатнім для надійного відокремлення АІТ від законних маркетингових кампаній. Високий рівень помилкових спрацьовувань, що спостерігається в маркетинговому трафіку, підкреслює необхідність додаткових поведінкових сигналів. Показники унікальних одержувачів значно зменшують кількість помилкових спрацьовувань, але не мають чутливості, необхідної для виявлення АІТ в ізоляції. Це підкріплює висновок, що ефективне запобігання АІТ має базуватися на поєднанні простих, інтерпретованих показників, а не на одному порозі.

Окрім запобігання шахрайству, результати дослідження мають також ширше значення для ефективності доставки. Блокуючи трафік, який генерує витрати без надання цінності для користувача, запобігання АІТ безпосередньо покращує показники ефективності доставки, такі як IGDP, та зменшує розрив між ціною та доставкою (PDG). З цієї точки зору виявлення шахрайства стає невід'ємною складовою оптимізації маршрутизації та економічної ефективності платформ CPaaS.

Загалом, це дослідження підтверджує, що правила виявлення з низькою затримкою, незалежні від вмісту, можуть сформувати міцну і практичну основу для захисту від АІТ. Хоча такі правила самі по собі не є достатніми для протидії всім варіантам атак, вони забезпечують необхідну базу, яку можна

негайно впровадити і поступово розширювати в міру впровадження більш досконалих технологій.

2.7. Висновки до другого розділу

У другому розділі дисертації сформовано теоретико-методологічну основу для оцінювання ефективності доставки SMS-повідомлень у мережах четвертого та п'ятого покоління. Проведений аналіз сучасних стандартів, наукових публікацій і галузевих практик дозволив систематизувати підходи до трактування ефективності телекомунікаційних сервісів у контексті еволюції 4G/5G-архітектур.

Показано, що традиційні технічні підходи, засновані виключно на показниках QoS, є недостатніми для повного опису ефективності сервісу коротких повідомлень у сучасних мережах. З урахуванням розвитку Service-Based Architecture, Network Slicing та QoE-орієнтованого управління мережею обґрунтовано необхідність інтеграції технічних, користувацьких та економічних факторів у єдиній моделі оцінювання.

У межах розділу виконано класифікацію існуючих підходів до оцінювання ефективності доставки повідомлень (QoS-based, system-based, cost-based та інтегрований), що дозволило виявити їхні переваги та обмеження. На цій основі обґрунтовано доцільність використання інтегрованого підходу як найбільш універсального та придатного для аналізу сервісів обміну повідомленнями в CPaaS-екосистемах.

Запропоновано та теоретично обґрунтовано показник Price Delivery Gap (PDG), який кількісно відображає розрив між очікуваною та фактичною ефективністю доставки повідомлень у вартісному вимірі. Визначено його зв'язок із параметрами QoS і QoE, а також показано місце PDG у сервісо-орієнтованій архітектурі 5G як інтегрального KPI для SLA-моніторингу та оптимізації маршрутизації.

Подальший розвиток концепції PDG реалізовано у вигляді інтегрованої моделі IGDP, яка поєднує технічну надійність доставки, користувацьке сприйняття та економічну ефективність у єдиному аналітичному представленні. Показано, що нормалізація складових моделі дозволяє застосовувати IGDP як цільову функцію в задачах багатокритеріальної оптимізації доставки повідомлень.

У розділі також розглянуто методологічні основи моделювання процесу доставки SMS, зокрема стохастичні та імовірно-часові моделі, що забезпечують формальний зв'язок між навантаженням мережі, параметрами якості та економічними показниками ефективності.

Окрему увагу приділено впливу шахрайського трафіку, зокрема Artificially Inflated Traffic (AIT), на ефективність доставки SMS. Доведено, що AIT є суттєвим джерелом деградації показників IGDP та зростання PDG, а його виявлення та блокування мають розглядатися не лише як завдання безпеки, а як складова оптимізації економічної ефективності сервісу. Аналіз запропонованих методів протидії AIT показав їхню високу результативність для високоінтенсивних атак, водночас виявив обмеження при обробці легітимного масового трафіку.

У результаті обґрунтовано необхідність впровадження модуля категоризації повідомлень як проміжного рівня між прийманням трафіку та механізмами маршрутизації, фрод-детекції й оцінювання ефективності. Категоризація дозволяє адаптувати порогові значення, вагові коефіцієнти IGDP та стратегії доставки до конкретного типу повідомлень, що створює методологічну основу для побудови інтелектуальної системи оптимізації доставки, яка буде розглянута у розділі 4.

Список джерел

1. ITU-T E.800. Definitions of terms related to quality of service and network performance. – Geneva : ITU-T, 2021.
2. ITU-T P.10/G.1000. Vocabulary for performance and quality of service. – Geneva : ITU-T, 2022.
3. ETSI TR 103 559 V1.1.1. Messaging QoS Metrics Definition. – Sophia Antipolis : ETSI, 2022.
4. 3GPP TS 23.501 V17.7.0. System Architecture for the 5G System (5GS). – Sophia Antipolis : 3GPP, 2022.
5. ETSI TR 103 640 V1.1.1. Network slicing impacts on service quality. – Sophia Antipolis : ETSI, 2022.
6. ITU-T E.804. Quality of service management: Concepts and methods. – Geneva : ITU-T, 2022.
7. ETSI TS 102 250-2. Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services; Part 2: Definition of quality of service parameters and their computation. – Sophia Antipolis : ETSI, 2017.
8. ITU-T Y.1541. Network performance objectives for IP-based services. – Geneva : ITU-T, 2020.
9. 3GPP TS 23.040. Technical realization of the Short Message Service (SMS). – Sophia Antipolis : 3GPP, 2023.
10. Lee J., Kim S., Park J. Performance optimization of SMS delivery in large-scale messaging systems // *IEEE Transactions on Network and Service Management*. – 2022. – Vol. 19, No. 3. – P. 2875–2887.
11. GSMA. Interconnect Benchmarking Report. – London : GSMA, 2023.
12. ITU-D. ICT Price Trends. – Geneva : ITU, 2022.
13. Zhang Y., Oestreich S. Integrated performance evaluation of OTT and RCS messaging services // *IEEE Access*. – 2023. – Vol. 11. – P. 104221–104234.
14. ETSI EG 202 009-3. Speech and multimedia transmission quality; QoS measurement methodologies. – Sophia Antipolis : ETSI, 2019.

- 15.ITU-T G.1030. Estimating end-to-end performance for data applications over networks. – Geneva : ITU-T, 2019.
- 16.ETSI TS 102 250-2. Quality of Service parameters for messaging services. – Sophia Antipolis : ETSI, 2017.
- 17.ETSI TR 103 559. Messaging quality metrics. – Sophia Antipolis : ETSI, 2022.
- 18.3GPP TS 23.502. Procedures for the 5G System (5GS). – Sophia Antipolis : 3GPP, 2022.
- 19.3GPP TS 28.554. Performance measurements for the 5G System. – Sophia Antipolis : 3GPP, 2022.
- 20.ETSI GS NFV-IFA 010. Management and orchestration; Performance management specification. – Sophia Antipolis : ETSI, 2021.
- 21.GSMA. Fraud and Security Group: Artificially Inflated Traffic (AIT) Overview. – London : GSMA, 2022.
- 22.3GPP TS 23.501. System Architecture for the 5G System. – Sophia Antipolis : 3GPP, 2022.
- 23.ITU-T E.802. Framework and methodologies for QoS measurement. – Geneva : ITU-T, 2021.
- 24.ITU-T G.1011. Reference guide to quality of experience assessment methodologies. – Geneva : ITU-T, 2019.
- 25.ETSI TR 103 559. End-to-end quality metrics for messaging services. – Sophia Antipolis : ETSI, 2022.
- 26.ITU-D. Measuring the Information Society Report. – Geneva : ITU, 2022.
- 27.GSMA. Mobile Interconnection Benchmarking Report. – London : GSMA, 2023.
- 28.ETSI EG 202 009-3. QoS measurement methodologies for messaging and multimedia services. – Sophia Antipolis : ETSI, 2019.
- 29.Kleinrock L. Queueing Systems. Volume 1: Theory. – New York : Wiley-Interscience, 1975.

30. Deb K. *Multi-Objective Optimization Using Evolutionary Algorithms*. – Chichester : John Wiley & Sons, 2001.
31. ITU-T E.804. *Quality management and optimization methods*. – Geneva : ITU-T, 2022.
32. Dey S., Kalyani S., Katti S. QoS/QoE-aware optimization in 5G networks // *IEEE Communications Surveys & Tutorials*. – 2022. – Vol. 24, No. 4. – P. 2381–2412.
33. ITU-T. *QoS Benchmarking Guidelines*. – Geneva : ITU-T, 2023.
34. Juniper Research, *Mobile Messaging Market Trends and Enterprise Use Cases*, 2024.
35. GSMA Fraud and Security Group, *SMS Firewalls and Anti-Fraud Best Practices (FS.19), Version 2.0*, 2023.
36. Grand View Research, *A2P Messaging Market Size, Share & Trends Analysis Report, 2025–2030*, 2024.
37. Grand View Research, *North America A2P Messaging Market Outlook 2030*, 2024.
38. MarketsandMarkets, *A2P Messaging Market—Global Forecast to 2029*, 2024.
39. AdaptiveMobile Security, *Messaging Abuse and A2P Fraud Trends, Threat Report Series, 2022–2024*.

РОЗДІЛ 3.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ДОСТАВКИ SMS ТА ОЦІНЮВАННЯ ВПЛИВУ ПАРАМЕТРІВ ЕФЕКТИВНОСТІ НА PDG

3.1. Мета та логіка математичного моделювання

У попередньому розділі було розроблено інтегровану модель IGDP (Integrated Gap–Delivery–Performance), яка поєднує технічну (QoS), користувацьку (QoE) та економічну (PDG) складові ефективності телекомунікаційних сервісів. Модель IGDP дозволяє оцінювати ефективність роботи сервісів коротких повідомлень не лише з погляду технічних параметрів, а й з урахуванням економічної доцільності та впливу на загальну продуктивність платформи.

Водночас отримана модель має переважно аналітичний характер і потребує кількісного підтвердження на основі емпіричних даних. Тому в цьому розділі здійснено математичне моделювання процесу доставки SMS-повідомлень, яке дозволяє дослідити вплив основних технічних параметрів (затримки, навантаження, ймовірності успішної доставки) на показники PDG (Price Delivery Gap) та, відповідно, на інтегральну ефективність IGDP.

Згідно з положеннями міжнародних рекомендацій ITU-T E.802 [1] та ETSI EG 202 009-3 «QoS measurement methodologies» [2], якість сервісу в мережах нового покоління має розглядатися як функція часу, навантаження та ймовірності успіху. Процес доставки SMS можна представити як систему масового обслуговування типу M/M/1, у якій випадкові затримки, відмови та повтори визначають фактичну ефективність доставки P_{eff} [3]. Модель M/M/1 використовується як аналітична апроксимація, тоді як відхилення від експоненційного розподілу затримок враховуються на емпіричному рівні.

Метою цього розділу є побудова математичних моделей процесу доставки коротких повідомлень у мережах 4G/5G, які дозволяють кількісно оцінити вплив параметрів QoS на PDG та визначити межі застосовності класичних моделей у реальних CPaaS-середовищах. У межах дослідження розглянуто експоненційну та гіперболічну моделі доставки, які апроксимують різні режими роботи мережі – стабільну та перевантажену відповідно до [4].

У реальних CPaaS-платформах ефективність доставки SMS безпосередньо впливає на фінансові показники сервісу, зокрема на маржинальність та рівень виконання SLA. Тому PDG може бути використаний як ключовий операційний KPI у системах моніторингу якості та оптимізації маршрутів доставки.

Результати моделювання дозволяють визначити залежність між QoS-параметрами та економічними втратами, що відображаються через PDG. При цьому встановлено, що високі значення затримки та випадкові відмови призводять до нелінійного зростання PDG, що знижує інтегральну ефективність IGDP. Разом із тим, класичні математичні моделі не враховують інших факторів, які мають істотний вплив на ефективність сервісу, зокрема — тип трафіку (OTP, транзакційний, маркетинговий), особливості вмісту повідомлень, а також аномальні та штучно згенеровані сценарії трафіку (AIT, fraud, smishing) [5].

У наступному розділі буде розглянуто використання інтелектуальних методів класифікації трафіку та виявлення аномалій з метою розширення можливостей оцінки ефективності сервісу. На основі цих підходів у четвертому розділі буде розроблено інтегрований фреймворк IEMOF (Intelligent Efficiency Management and Optimization Framework), який поєднає математичне моделювання, аналітику на основі AI/ML та адаптивну маршрутизацію для комплексного підвищення ефективності доставки SMS-повідомлень.

3.2. Ключові параметри для визначення надійності послуги SMS

З огляду на широкий спектр випадків використання SMS та те, що вони є невід'ємною частиною набору GSM, якість послуг регулюється стандартом, встановленим Європейським інститутом телекомунікаційних стандартів (ETSI) у його останній технічній специфікації ETSI TS 102 250-2 v.2.7.1. Таблиця огляду параметрів із тригерними точками наведена на рис. 3.1.

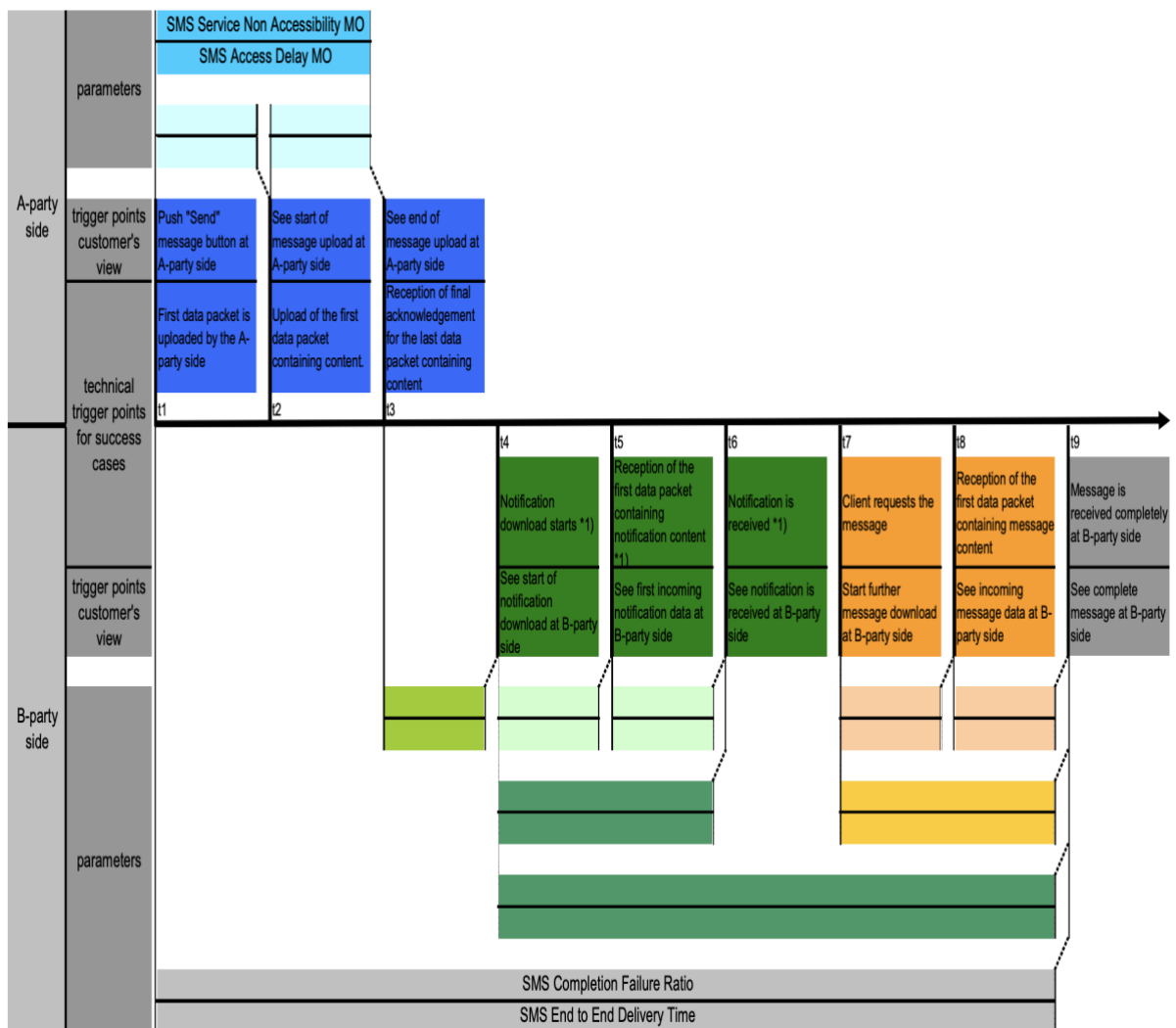


Рис. 3.1. Огляд параметрів SMS із тригерними точками.

Визначено наступний набір параметрів для оцінки якості обслуговування [6]:

1. Недоступність послуги SMS [%] – позначає ймовірність того, що кінцевий користувач не зможе отримати доступ до послуги коротких

повідомлень на запит, коли вона пропонується шляхом відображення індикатора мережі на користувацькому обладнанні (UE).

2. Затримка доступу до SMS [с] – означає проміжок часу між відправленням короткого повідомлення в мережу та отриманням підтвердження про відправлення від мережі на стороні відправника.

3. Коефіцієнт невдалого завершення SMS [%] – відповідає співвідношенню невдало отриманих та відправлених повідомлень від одного UE до іншого UE, за винятком дубльованих отриманих та пошкоджених повідомлень.

4. Час доставки SMS від кінця до кінця [с] – відповідає періоду часу між відправленням короткого повідомлення в мережу та отриманням того самого короткого повідомлення на іншому UE.

5. Коефіцієнт невдач підтвердження отримання SMS [%] – позначає ймовірність того, що підтвердження отримання відправленого повідомлення не буде отримано UE-відправником, незважаючи на запит.

6. Час підтвердження отримання SMS [с] – відповідає періоду часу між відправленням короткого повідомлення в мережу та отриманням підтвердження отримання цього повідомлення з мережі.

7. Коефіцієнт невдач підтвердження доставки SMS [%] – позначає ймовірність того, що підтвердження доставки для надісланої спроби не буде отримано вихідним UE, незважаючи на запит.

8. Час підтвердження доставки SMS [с] – відповідає періоду часу між надсиланням короткого повідомлення до мережі та отриманням підтвердження доставки від мережі.

Для оцінювання фактичної ефективності доставки коротких повідомлень доцільно розглядати сукупність технічних параметрів, які безпосередньо впливають на надійність сервісу. До них належать:

- P_s – імовірність успішної доставки;
- T_d – середній час затримки доставки;
- R_r – коефіцієнт повторних спроб передачі;

- P_f – частка відмов чи втрат;
- λ – інтенсивність надходження повідомлень до системи;
- μ – інтенсивність їх оброблення мережею;
- $\rho = \lambda/\mu$ – коефіцієнт завантаження системи, який відображає співвідношення між потоком вхідних і доставлених повідомлень.

Оскільки використання P2P SMS за останнє десятиліття зменшується, а використання A2P SMS, M2M, IoT та OTA SMS розширюється, такі параметри, як коефіцієнт невдач завершення SMS, час доставки SMS від кінця до кінця, час підтвердження отримання SMS стають ключовими параметрами, оскільки вони впливають на SMS (MT).

Потік транзакцій для сценарію MT SMS показаний на рис. 3.2.

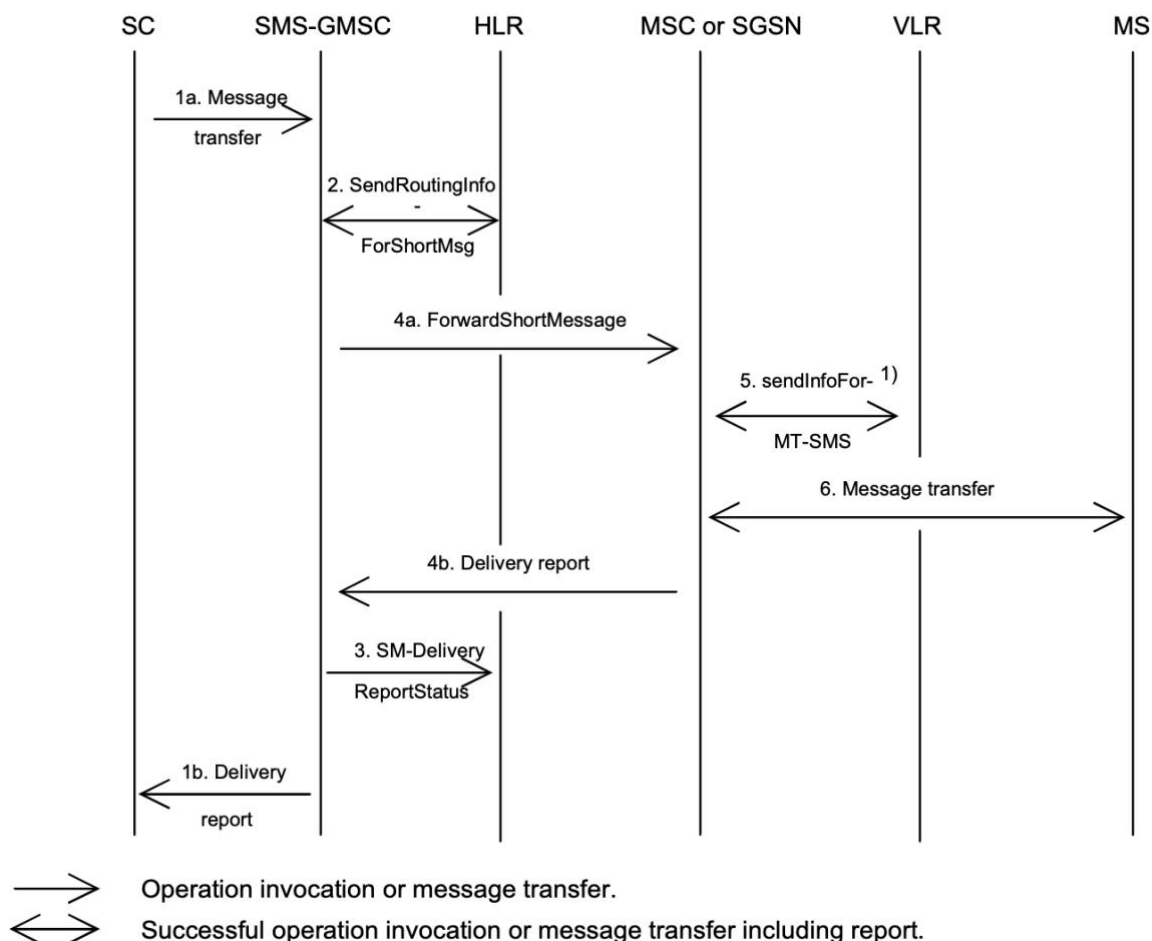


Рис. 3.2. Передача СМС МТ.

Коефіцієнт невдалої доставки SMS розраховується за такою формулою [6]:

$$\text{SMS Completion Failure [\%]} = \frac{\text{unsuccessfully received SMS}}{\text{all SMS service attempts}} \times 100. \quad (3.1)$$

Час доставки SMS від початку до кінця розраховується за такою формулою:

$$\text{SMS End-to-End Delivery Time [s]} = (t_{B, \text{receive}} - t_{A, \text{send}})[s]. \quad (3.2)$$

Час підтвердження отримання SMS розраховується за такою формулою:

$$\text{SMS Receive Confirmation Time [s]} = (t_{A, \text{receive confirmation}} - t_{A, \text{send}})[s]. \quad (3.3)$$

На ці параметри можуть впливати різні етапи доставки повідомлень: перевантаження в системі SMS (між відправником SMS і мережею мобільного оператора), перевантаження в сигнальних каналах мережі мобільного оператора або в міжмережевих з'єднаннях (з'єднання між мобільними операторами), недоступність послуги в певних географічних районах.

Ризик невдалої доставки для ринку A2P SMS можна розглядати як коефіцієнт невдалої доставки SMS. Нехай цей ризик дорівнює q , тоді $p = 1 - q$ – це ймовірність успішної доставки SMS.

Узагальнюючи наведені характеристики, можна визначити фактичну ефективність доставки повідомлень, яка враховує вплив імовірності доставки, повторних спроб, втрат і затримок:

$$P_{eff} = P_s \cdot (1 - P_f) \cdot e^{-kT_d}, \quad (3.4)$$

де k – емпіричний коефіцієнт чутливості ефективності до затримки.

У більш загальному вигляді ефективність доставки описується функцією:

$$P_{eff} = f(P_s, T_d, R_r, P_f, \rho), \quad (3.5)$$

що відображає залежність ефективності сервісу від сукупності параметрів якості обслуговування (QoS).

Отримана інтегральна ймовірність доставки $P_{eff}(t)$ у подальшому використовується як кількісна характеристика фактичної надійності доставки повідомлень і є складовою оцінки економічної ефективності доставки, що враховується при розрахунку показника PDG.

$$PDG = \frac{P_{exp} - P_{eff}}{P_{exp}}, \quad (3.6)$$

де P_{exp} – очікувана ефективність сервісу, встановлена відповідно до нормативів SLA чи технічних вимог оператора.

3.3. Побудова математичної моделі для параметра доставки SMS

У попередньому підрозділі було визначено набір параметрів, що формують ефективність доставки повідомлень P_{eff} як функцію якості обслуговування. Для подальшого аналітичного дослідження доцільно перейти до спрощеного представлення цієї залежності у часовій області, де ефективність доставки розглядається як функція часу, що враховує зростання затримок і зниження ймовірності доставки при зміні навантаження.

Таким чином, у цьому підрозділі моделюється динаміка $P_{eff}(t)$ за допомогою двох апроксимаційних залежностей – експоненційної та гіперболічної, які відповідають стабільному та перевантаженому режимам роботи мережі відповідно.

Розглянемо реальний приклад даних про ефективність доставки SMS-повідомлень.

У табл. 3.1 наведено статистичні дані про кількість SMS-повідомлень, надісланих усіма операторами в країні А в період з 15.10.22 по 31.10.22 одним із постачальників CPaaS.

Статистика доставки СМС повідомлень в період між 15.10-31.10.22

Параметр доставки повідомлень СМС МТ	Кількість повідомлень СМС МТ
Всього	2492766
Доставлено	2196792
Стік час очікування доставки	159584
Неможливо доставити	136390
Доставлено протягом 10 сек	1811373
Доставлено протягом 30 сек	248302
Доставлено протягом 1 хв	36723
Доставлено протягом 5 хв	55041
Доставлено протягом 15 хв	11394
Доставлено протягом 30 хв	6490
Доставлено протягом 1 год	7329
Доставлено протягом 3 год	6616
Доставлено протягом 6 год	3450
Доставлено протягом 12 год	2931
Доставлено протягом 24 год	3870
Доставлено протягом 48 год	3273

Дані з табл. 3.1 показують, що в цьому випадку 82,46% всіх доставлених SMS МТ доставляються протягом 10 секунд, що є прийнятним значенням. Всі більш тривалі терміни доставки вважаються надмірними (підлягають перевірці).

Спостережувані дані можна розглядати як часовий ряд з дискретним часом і вимірами в часових моментах $t = \{10, 30, 60, 300, 900, 1800, 3600, 10800, 21600, 43200, 86400, 172800\}$ [с]. Для побудови математичної моделі таких часових рядів можна використовувати загальні методи апроксимації, наприклад, описані в [7]. Кількість доставлених повідомлень зменшується з часом. Діапазон змін цієї кількості є значним.

Для кращого представлення даних для візуалізації краще використовувати логарифмічну шкалу для кількості та часу. Крім того, можна також використовувати натуральні логарифми для сукупної кількості доставлених SMS. Для оцінки ризику невдалої доставки обчислюються поточні статистичні ймовірності успішної доставки p_{si} та відповідні сукупні ймовірності $p_{\Sigma i}$.

На рис. 3.3 показано початкові залежності для спостережуваних даних, представлених у табл. 3.1.

Візуальний аналіз залежностей, представлений на рисунку 3.3, дозволяє зробити висновок, що кумулятивні криві більш підходять для оцінки ризику несправностей доставки.

Розглянемо дві альтернативні моделі оцінки ризику.

1. Експоненціальна модель.

Ця модель визначається наступним рівнянням:

$$f(t) = a - ae^{-bt}, \quad (3.7)$$

де a та b – коефіцієнти моделі, які необхідно оцінити.

Слід зазначити, що в разі аналізу даних кумулятивних ймовірностей оцінка коефіцієнта моделі a підходить для оцінки ймовірності правильної доставки SMS. Отже, ризик невдалої доставки можна визначити як $q = 1 - a$.

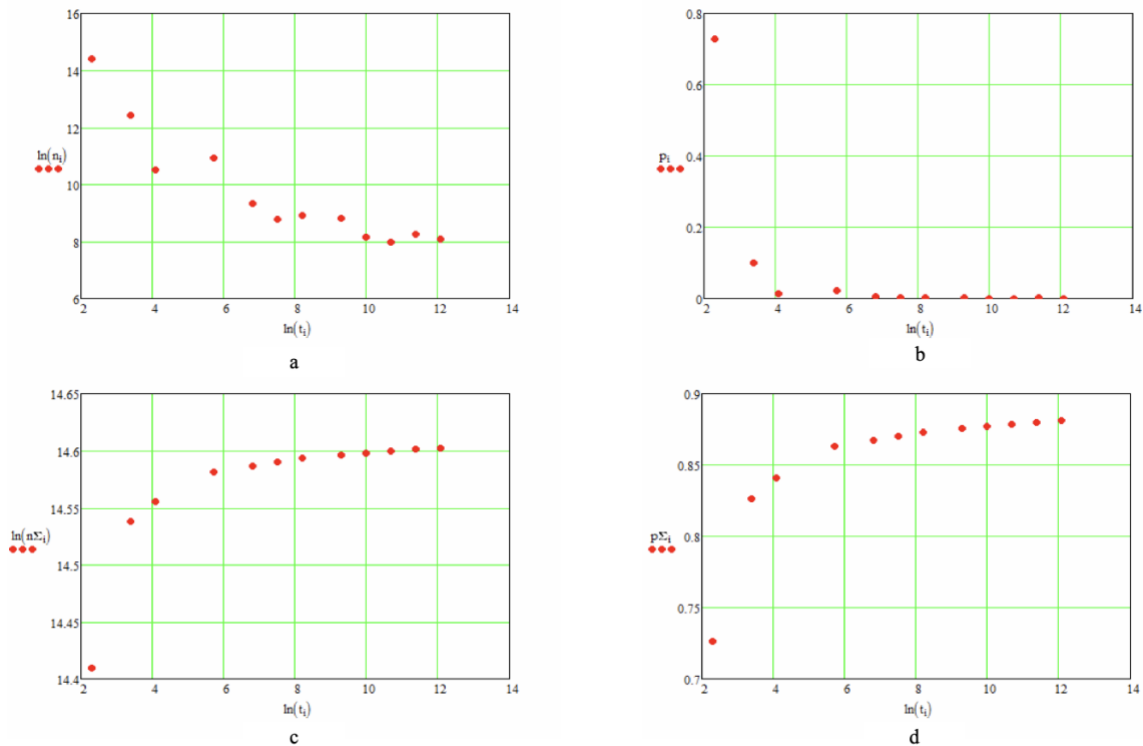


Рисунок 3.3. Спостережувані часові ряди:

a – залежність кількості доставлених повідомлень від часу;

b – залежність статистичних ймовірностей успішної доставки від часу;

c – залежність сукупної кількості доставлених повідомлень від часу;

d – залежність сукупних ймовірностей успішної доставки від часу.

Загальний підхід до оцінки коефіцієнта моделі полягає у використанні методу звичайних найменших квадратів після прийняття припущення про нормальний розподіл похибки. У цьому випадку необхідно вирішити систему рівнянь

$$\begin{cases} \sum_{i=1}^N f_i(e^{-bt_i} - 1) - a \sum_{i=1}^N (e^{-bt_i} - 1)^2 = 0, \\ \sum_{i=1}^N f_i t_i e^{-bt_i} - a \sum_{i=1}^N t_i e^{-bt_i} + a \sum_{i=1}^N t_i e^{-2bt_i} = 0. \end{cases} \quad (3.8)$$

Розв'язання такої системи рівнянь є складним завданням.

Розглянемо альтернативний підхід. На першій ітерації ми можемо обчислити невідомі коефіцієнти, припустивши, що експоненціальна крива (3.7) містить перший і останній зразки початкового набору даних. У цьому випадку необхідно розв'язати систему рівнянь

$$\begin{cases} f_1 = a(1 - e^{-bt_1}), \\ f_N = a(1 - e^{-bt_N}). \end{cases} \quad (3.9)$$

Після математичного спрощення ми можемо отримати

$$a = \frac{f_1}{1 - e^{-bt_1}}. \quad (3.10)$$

Коефіцієнт моделі b можна знайти як ненульове рішення рівняння

$$\frac{f_1}{f_N} - 1 + e^{-bt_1} - \frac{f_1}{f_N} e^{-bt_N} = 0. \quad (3.11)$$

Для розв'язання цього рівняння можна використати один із чисельних методів.

Друга ітерація пов'язана з мінімізацією стандартної похибки моделі. Для цього можна застосувати оптимізаційний параболоїд. Загальна методологія зазначеної оптимізації розглядається в [7].

Щоб оцінити ризик несправності доставки за допомогою повних статистичних даних, нам потрібно проаналізувати дані, отримані протягом

двох днів. Використання техніки побудови експоненціальної моделі (3.7) дає можливість створити онлайн-платформу для прийняття рішень з прогнозною здатністю для швидкої оцінки ризику невдач доставки. Наприклад, ми можемо виміряти дані протягом 5 хвилин і для цих даних оцінити коефіцієнти моделі з подальшим перерахунком одного з цих коефіцієнтів у ризику невдач доставки. Блок-схема відповідного методу показана на рисунку 3.4.

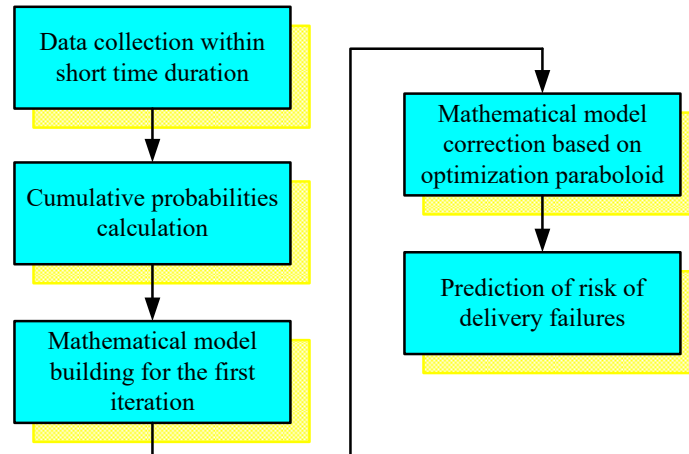


Рисунок 3.4. Схема оцінки ризиків доставки.

Розглянемо приклад побудови моделі для даних, показаних на рисунку 3d. Перша ітерація дає наступні оцінки коефіцієнтів моделі: $a=0.881$ та $b=0.756$. Математична модель кумулятивних ймовірностей буде такою:

$$p_{\Sigma}(\ln t) = 0.881 - 0.881e^{-0.756 \ln t}. \quad (3.12)$$

Для корекції математичної моделі ми розглянули 25 варіантів коефіцієнтів моделі (п'ять можливих значень кожного коефіцієнта) і обчислили стандартні відхилення. Результати обчислень наведено в табл. 3.2.

Таблиця 3.2

Результати обчислення стандартних відхилень

$S(a, b)$	$b = 0.656$	$b = 0.706$	$b = 0.756$	$b = 0.806$	$b = 0.856$
$a = 0.861$	0.03308	0.02535	0.0197	0.01639	0.01551
$a = 0.871$	0.02489	0.0164	0.009999	0.007296	0.009344
$a = 0.881$	0.01918	0.01099	0.006288	0.008203	0.01285
$a = 0.891$	0.01843	0.01409	0.01438	0.01763	0.02177
$a = 0.901$	0.02312	0.02241	0.02454	0.028	0.03183

Дані з табл. 3.2 були апроксимовані за допомогою параболоїда та методу найменших квадратів. Розрахунок дає наступну модель параболоїда:

$$S(a, b) = 26.079 - 55.695a + 30.091a^2 - 4.173b + 0.591b^2 + 3.692ab \quad (3.13)$$

Мінімум параболоїда визначає найкращі коефіцієнти моделі: $a_{\text{opt}} = 0.877$ та $b_{\text{opt}} = 0.791$. Отже, модель (3.12) після корекції набуває оптимальної форми:

$$p_{\Sigma \text{opt}}(\ln t) = 0.877 - 0.877e^{-0.791 \ln t}. \quad (3.14)$$

Результат апроксимації за допомогою моделі (3.14) показано на рис. 3.5.

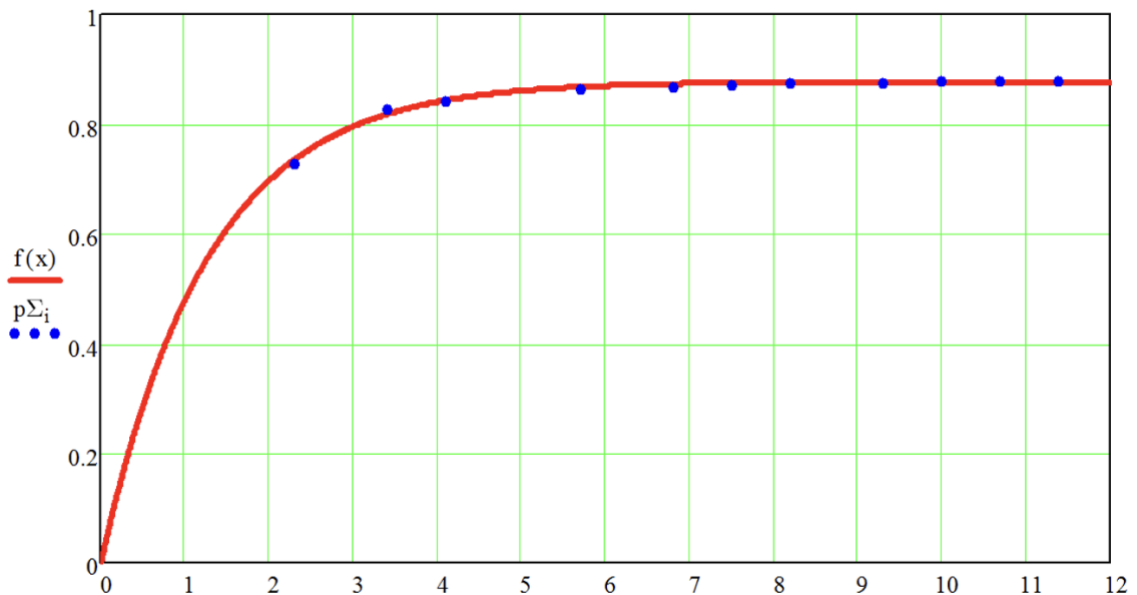


Рис. 3.5. Експоненціальна модель для кумулятивних ймовірностей успішної доставки.

Отже, ризик невдалої доставки для моделі (3.14) становить $q = 1 - a_{\text{opt}} = 0.123$.

2. Гіперболічна модель.

Ця модель визначається за наступним рівнянням:

$$f(t) = \frac{at}{1+bt}. \quad (3.15)$$

де a і b – коефіцієнти моделі, які необхідно оцінити.

Щоб визначити ризик невдач при доставці, ми можемо проаналізувати кумулятивні ймовірності, подібні до моделі (3.7). Але в цьому випадку ризик невдач при доставці можна приблизно визначити як $q \approx 1 - a/b$.

Точні значення коефіцієнтів моделі є розв'язками системи двох рівнянь, отриманих після застосування методу найменших квадратів:

$$\begin{cases} \sum_{i=1}^N \frac{f_i t_i}{1+bt_i} - a \sum_{i=1}^N \frac{t_i^2}{(1+bt_i)^2} = 0, \\ \sum_{i=1}^N \frac{f_i t_i^2}{(1+bt_i)^2} - a \sum_{i=1}^N \frac{t_i^3}{(1+bt_i)^3} = 0. \end{cases} \quad (3.16)$$

Для спрощення оцінки коефіцієнтів моделі можна використовувати підхід, описаний для моделі (3.7). Тому на першій ітерації припускаємо, що гіперболічна крива (3.15) містить перший і останній зразки початкового набору даних. У цьому випадку необхідно вирішити систему рівнянь

$$\begin{cases} f_1(1 + bt_1) = at_1, \\ f_N(1 + bt_N) = at_N. \end{cases} \quad (3.17)$$

Після математичного спрощення, можемо отримати:

$$a = \frac{f_1 f_N (t_1 - t_N)}{t_1 t_N (f_1 - f_N)} \text{ та } b = \frac{t_1 f_N - t_N f_1}{t_1 t_N (f_1 - f_N)} \quad (3.18)$$

На другому ітераційному циклі модель коригується за допомогою оптимізаційного параболоїда.

Розглянемо приклад побудови гіперболічної моделі для даних, показаних на рис. 3.3d. Після першого ітераційного циклу були отримані початкові оцінки коефіцієнтів моделі: $a = 1.455$ та $b = 1.568$. Тоді модель (3.15) буде:

$$p_{\Sigma}(\ln t) = \frac{1.455 \ln t}{1 + 1.568 \ln t}. \quad (3.19)$$

На другому ітераційному циклі було проведено корекцію математичної моделі для 25 варіантів коефіцієнтів моделі (п'ять можливих значень кожного коефіцієнта) на основі розрахунку стандартних відхилень. Результати розрахунків наведено в табл. 3.3.

Результати обчислення стандартних відхилень

$S(a, b)$	$b = 1.508$	$b = 1.538$	$b = 1.568$	$b = 1.598$	$b = 1.628$
$a = 1.395$	0.0277	0.04147	0.05606	0.0706	0.08485
$a = 1.425$	0.01738	0.02504	0.03828	0.05257	0.06689
$a = 1.455$	0.02475	0.01682	0.02262	0.03524	0.04922
$a = 1.485$	0.04123	0.02622	0.01675	0.02044	0.03234
$a = 1.515$	0.05969	0.04291	0.02779	0.01711	0.01856

Стандартні відхилення можна визначити як залежність від коефіцієнтів моделі за допомогою методу найменших квадратів:

$$S(a, b) = -0.974 - 0.782a + 4.112a^2 + 2.044b + 2.758b^2 - 7.256ab. \quad (3.20)$$

Мінімум параболоїда визначає найкращі коефіцієнти моделі: $a_{\text{opt}} = 1.444$ та $b_{\text{opt}} = 1.529$. Отже, модель (3.19) після корекції набуває оптимальної форми:

$$p_{\Sigma}(\ln t) = \frac{1.444 \ln t}{1 + 1.529 \ln t}. \quad (3.21)$$

Результат апроксимації за допомогою моделі (3.21) показано на рис. 3.6.

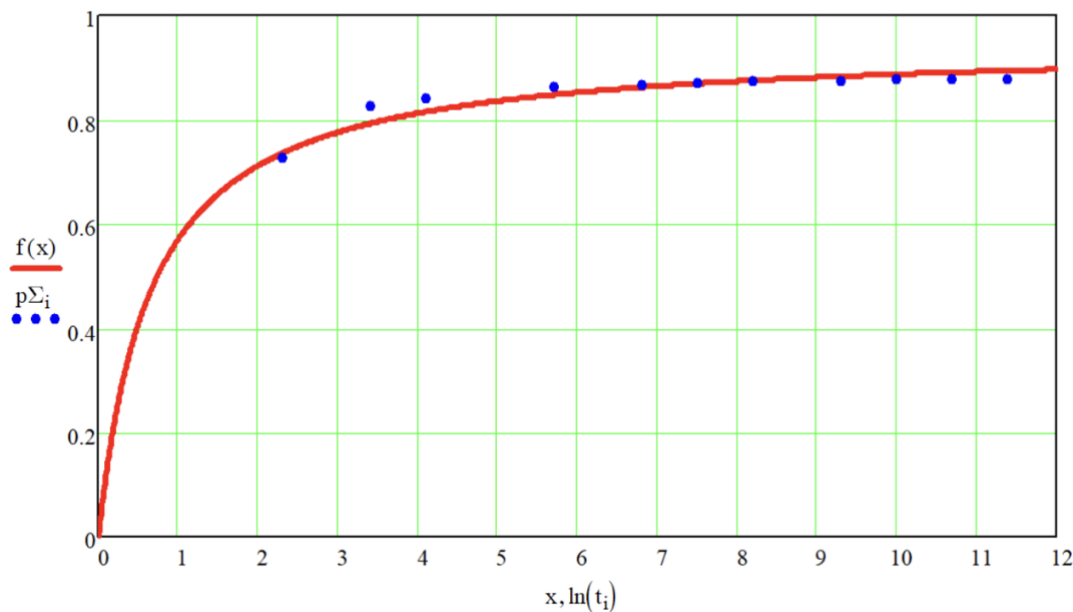


Рис. 3.6. Гіперболічна модель для кумулятивних ймовірностей успішної доставки.

Отже, ризик невиконання поставки для моделі (3.21) становить $q = 0.104$.

Розглянемо числові приклади оцінки ризику для двох наборів даних про доставку SMS МТ. Початкові дані наведено в табл. 3.4.

Аналіз даних для першого та другого наборів даних дає можливість зробити висновок, що в обох випадках реальна вартість ризику невиконання зобов'язань з доставки становить $q = 0.115$.

Розглянемо результати оцінки ризику з використанням запропонованої методології та експоненціальної моделі (3.7). Для прийняття рішення щодо значення ризику ми використовуємо лише чотири значення наборів даних, виміряних протягом 5 хвилин спостереження.

Таблиця 3.4

Статистика доставки SMS МТ між 31/10/22 та 06/11/22 (набір даних 1) та між 07/11/22 та 13/11/22 (набір даних 2)

Параметри доставки повідомлень СМС МТ	Кількість повідомлень (набір даних 1)	Кумулятивна вірогідність	Кількість повідомлень (набір даних 2)	Кумулятивна вірогідність
Всього	1020320	–	890677	–
Доставлено	902546	–	788318	–
Доставлено протягом 10 сек	740181	0.725	657701	0.738
Доставлено протягом 30 сек	101664	0.825	73300	0.821
Доставлено протягом 1 хв	12892	0.838	12076	0.834
Доставлено протягом 5 хв	22822	0.86	22637	0.86
Доставлено протягом 15 хв	7920	0.868	6999	0.868
Доставлено протягом 30 хв	3222	0.871	2562	0.87
Доставлено протягом 1 год	3942	0.875	3429	0.874
Доставлено протягом 3 год	3152	0.878	3017	0.878
Доставлено протягом 6 год	1739	0.88	1427	0.879
Доставлено протягом 12 год	1458	0.881	1452	0.881
Доставлено протягом 24 год	1930	0.883	1927	0.883
Доставлено протягом 48 год	1624	0.885	1791	0.885

Під час першої ітерації було отримано такі моделі для двох наборів даних:

$$p_{\Sigma}(\ln t / \text{dataset1}) = 0.870 - 0.870e^{-0.779 \ln t}. \quad (3.22)$$

$$p_{\Sigma}(\ln t / \text{dataset2}) = 0.867 - 0.867e^{-0.828 \ln t}. \quad (3.23)$$

Друга ітерація дає виправлені результати:

$$p_{\Sigma \text{opt}}(\ln t / \text{dataset1}) = 0.872 - 0.872e^{-0.797 \ln t}. \quad (3.24)$$

$$p_{\Sigma \text{opt}}(\ln t / \text{dataset2}) = 0.866 - 0.866e^{-0.838 \ln t}. \quad (3.25)$$

Результати апроксимації з використанням експоненціальної моделі для першого та другого наборів даних показано на рис. 3.7.

Оцінки ризику невиконання зобов'язань для першого та другого наборів даних становлять відповідно 0,128 та 0,134. На рис. 3.7 показано хорошу збіжність даних з експоненційною моделлю та достатню точність оцінки ризику в завданні довгострокового прогнозування. Водночас експоненційна модель дає дещо завищені оцінки ризику.

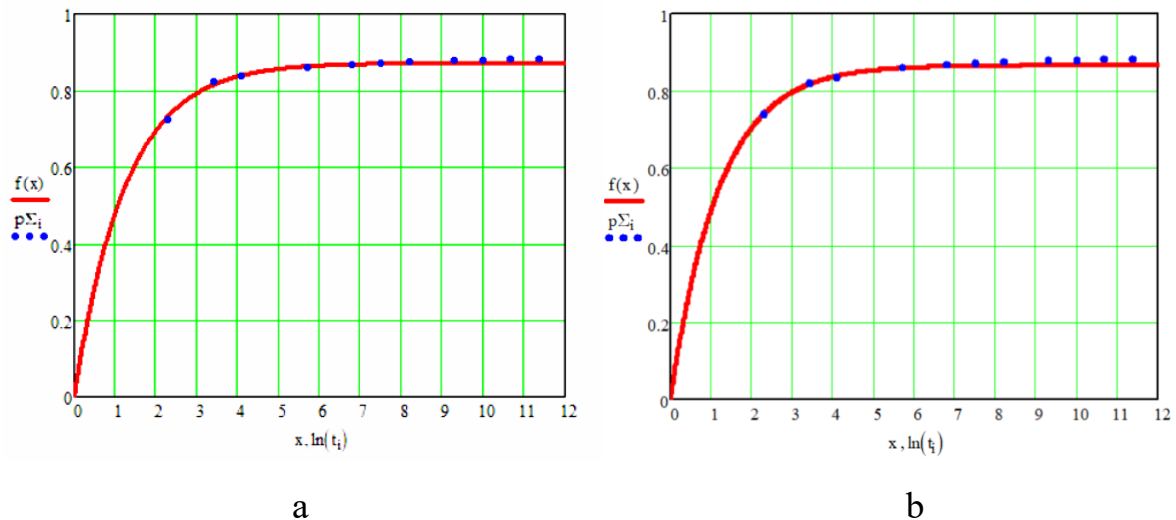


Рис. 3.7. Експоненціальна модель для першого (а) та другого (б) наборів даних.

Розглянемо результати оцінки ризику з використанням запропонованої методології та гіперболічної моделі (3.15) для даних, зібраних протягом 5 хвилин спостереження. На першій ітерації були отримані наступні моделі для двох наборів даних:

$$p_{\Sigma}(\ln t / \text{dataset1}) = \frac{1.2 \ln t}{1 + 1.22 \ln t}. \quad (3.26)$$

$$p_{\Sigma}(\ln t / \text{dataset2}) = \frac{1.356 \ln t}{1 + 1.402 \ln t}. \quad (3.27)$$

Друга ітерація дає скориговані моделі:

$$p_{\Sigma\text{opt}}(\ln t / \text{dataset1}) = \frac{1.147 \ln t}{1+1.202 \ln t}. \quad (3.28)$$

$$p_{\Sigma\text{opt}}(\ln t / \text{dataset2}) = \frac{1.327 \ln t}{1+1.392 \ln t}. \quad (3.29)$$

Оцінки ризику несправності доставки для першого та другого наборів даних становлять відповідно 0,111 та 0,1. Гіперболічна модель має достатню точність і дає дещо підвищені оцінки ризику.

Для підвищення точності оцінки ризиків бажано організувати збір даних з достатньо високою частотою вибірки протягом 5 хвилин спостереження.

Запропоновані аналітичні залежності описують часову поведінку ефективності $P_{eff}(t)$ і дозволяють на основі отриманих кривих оцінювати економічну ефективність за допомогою показника PDG. Таким чином, математичні моделі експоненційного та гіперболічного типу можна розглядати як динамічні апроксимації фактичної ефективності доставки, визначеної у попередньому підрозділі.

Проведене моделювання продемонструвало, що обидві апроксимаційні моделі забезпечують достатню точність прогнозування ризику невдач доставки на основі коротких часових вибірок. Це створює передумови для практичної реалізації в CPaaS-системах, де показник PDG може розраховуватися в реальному часі як функція ефективності доставки.

Водночас отримані моделі не враховують змістовні характеристики трафіку та його аномалії, які також впливають на ефективність сервісу. Тому в наступному розділі буде розглянуто методи інтелектуальної класифікації та виявлення fraud-трафіку, що доповнюють результати математичного моделювання.

3.4. Регресійні методи оцінювання ефективності доставки SMS в умовах гетероскедастичності

3.4.1. Загальні положення

Запропоновані в параграфі 3.3 експоненційна та гіперболічна моделі дозволяють описати часову динаміку ефективності доставки SMS. Водночас у реальних CPaaS-середовищах часові ряди параметрів доставки характеризуються високою нестабільністю, шумами та гетероскедастичністю, що ускладнює коректну оцінку параметрів моделей і, як наслідок, показника PDG. Тому в цьому підрозділі розглянуто регресійні методи, спрямовані на підвищення стійкості оцінювання ефективності доставки в умовах нестабільних даних.

Для апробації регресійних методів оцінювання параметрів у умовах гетероскедастичності використано реальні набори даних із прикладної галузі, що характеризуються високою варіативністю, нерівномірною дисперсією та наявністю шумів. Зокрема, у дослідженні застосовано медичні часові ряди, які використовуються виключно як репрезентативний приклад нестабільних даних. Предметом аналізу є математичні властивості часових рядів та ефективність регресійних методів, а не специфіка медичної предметної області.

Обрані набори даних не впливають на загальність запропонованих методів, оскільки аналіз гетероскедастичності, кластеризації та стиснення даних базується на статистичних характеристиках часових рядів і не залежить від їх предметної інтерпретації. Аналогічні властивості притаманні й телекомунікаційним даним доставки SMS у CPaaS-середовищах.

3.4.2. OLS як базовий підхід та обмеження

Розглянемо приклад статистичних даних, на якому буде пояснений метод обліку гетероскедастичності. Первинні дані містять медичну інформацію і представлені у вигляді табл. 3.5 [8].

Дані в таблиці 5 пов'язані з проблемами невеликого розміру вибірки. Щоб підтвердити доцільність використання лінійної регресії, спочатку було проведено тест на лінійність даних [9]. Тест включав такі обчислення: знаходження кумулятивної функції залишків, визначення максимального та мінімального значень кумулятивної функції, обчислення діапазону варіації як різниці між максимальним і мінімальним значеннями та знаходження відношення діапазону варіації до стандартного відхилення.

Кумулятивна крива залишків була обчислена для простої лінійної регресії, отриманої за допомогою OLS. У цьому випадку регресія Y на X описується рівнянням $y(x) = 2.7864 + 0.8805x$. Для цього рівняння діапазон варіації кумулятивної функції становив 12,3994, а стандартне відхилення – 3,072. Значення критичного статистичного параметра становить 4,0362, що означає, що з імовірністю 0,995 досліджувані дані можна описати лінійною функцією.

Таблиця 3.5

Датасет

№	x	y	№	x	y
1	23	25.2	11	26.4	24.8
2	33.2	26	12	21.8	26.8
3	16.6	16.3	13	14.9	15.4
4	26.3	27.2	14	17.4	14.9
5	20	23.2	15	20	18.1
6	20	18.1	16	13.2	16.3
7	20.6	22.2	17	28.4	31.3
8	18.9	17.2	18	25.9	31.2
9	17.8	18.8	19	18.9	18
10	20	16.4	20	13.8	15.6

Для тих самих даних можна знайти регресію X на Y на основі OLS, яка буде описана формулою $(y) = 4.2101 + 0.7870y$, що може бути представлено у звичайній формі $y(x) = -5.3496 + 1.2707x$.

Регресії Y на X і X на Y можна об'єднати, використовуючи такий підхід:

- знаходження точки перетину ліній, тобто центру ваг досліджуваних даних у декартовій системі координат;

- знаходження дотичної до нахилу загальної прямої лінії як середнього геометричного дотичних до нахилів двох регресій;

- знаходження загального рівняння з обчисленою дотичною до нахилу через точку центру ваг.

У результаті можна отримати рівняння:

$$y(x) = -0.9097 + 1.0578x. \quad (3.30)$$

Досліджувані дані містять одну точку, яка, ймовірно, є винятковою. Автори публікації [8] відкинули цю точку з розрахунків і далі побудували модель, використовуючи OLS і лінійну регресію. Однак обґрунтування для неврахування цієї точки було недостатнім.

Тому було проведено приблизний статистичний аналіз наявності винятків у первинному наборі даних. Для цього були розраховані межі діапазонів варіації поля даних через крайні точки знизу і зверху. У цьому випадку було використано три точки. Результати представлені на рис. 3.8.

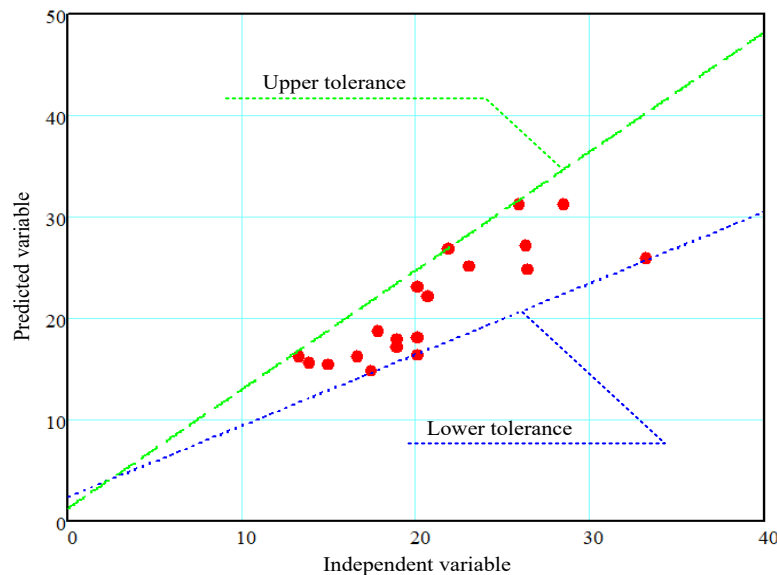


Рис. 3.8. Межі діапазонів варіації для досліджуваного набору даних.

Точка, яка підозрюється в тому, що є винятковою, входить в межі діапазонів варіації, а тому її слід враховувати при побудові математичної моделі. У цій роботі врахування гетероскедастичності розглядається як джерело для отримання надійної моделі. Друга точка набору даних є основним джерелом гетероскедастичності, оскільки вона має найбільший внесок у зміну дисперсії для великих значень залежної змінної. Критичність включення другої точки в розрахунок пояснюється невеликим розміром вибірки. Якби набір даних містив приблизно 100 вибірок, то, ймовірно, було б два або три пункти, подібні до другого, які були б нормальними, а не винятковими. Для побудови математичної моделі в умовах гетероскедастичності необхідно додатково обчислити коефіцієнти ваги гетероскедастичності. Для цього необхідно отримати рівняння гетероскедастичності.

Метод зважених найменших квадратів (WLS) використовує вагові коефіцієнти $w(y_i) = 1/x_i$. В результаті обчислення можна отримати регресії Y на X і X на Y у таких формах: $y(x) = 2.3980 + 0.8992x$ та $x(y) = 3.8056 + 0.8061y$. Другий можна записати у вигляді $y(x) = -4.7209 + 1.2405x$. Отже, середня геометрична регресія WLS буде такою:

$$y(x) = -0.8757 + 1.0561x. \quad (3.31)$$

Іншим класичним підходом до врахування гетероскедастичності є розрахунок за методом Уайта. Застосування цього методу дає наступні рівняння для регресій Y на X і X на Y : $y(x) = 2.7864 + 0.8805x$ та $y(x) = -5.3496 + 1.2707x$. Ці рівняння збігаються з регресіями OLS. Отже, середня геометрична регресія для методу Уайта буде такою самою, як і для моделі OLS (3.30).

3.4.3. Аналіз гетероскедастичності на основі методу рухомого трикутника

У цьому дослідженні запропоновано два нових показники для врахування гетероскедастичності. Перший показник є узагальненням методу ковзної середньої. Для реалізації цього підходу рухомий трикутник послідовно проходить через весь набір даних. Ширина ковзного вікна становить три точки, на основі яких будуються трикутники. Показником, що характеризує гетероскедастичність, є довжина відрізка, що з'єднує верхню точку трикутника з протилежною стороною трикутника. У цьому випадку пряма, що містить цей відрізок, є паралельною осі ординат (рис. 3.9).

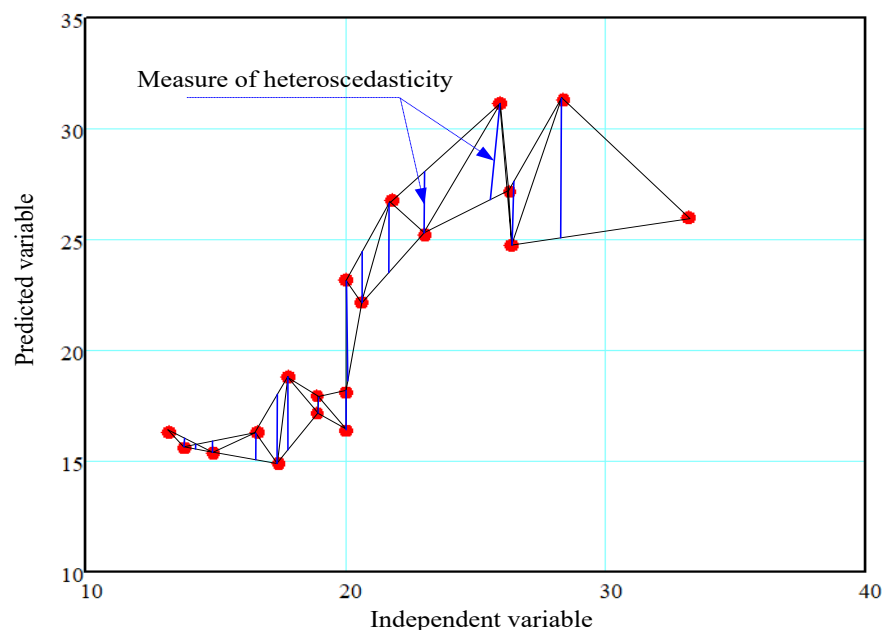


Рис. 3.9. Принцип побудови трикутників.

У цьому випадку рівнянням гетероскедастичності є залежність показника гетероскедастичності від середнього арифметичного координат вершин трикутника вздовж осі абсцис. В результаті використання OLS було отримано таке рівняння гетероскедастичності: $H(y) = -0.7383 + 0.1405y$. Зміна параметра гетероскедастичності для крайніх лівого та правого точок становить 2,1036. Для знаходження вагових коефіцієнтів використовується таке рівняння:

$$w(y_i) = \left(\frac{\bar{y}}{H(y_i)} \right)^2, \quad (3.32)$$

де \bar{y} є середнім значенням прогнозованої змінної.

Для цих ваг (3.32), заснованих на методі зважених найменших квадратів (WLS), регресію Y на X можна знайти за рівнянням $y(x) = -0.5774 + 1.0098x$. Після аналогічних обчислень можна знайти рівняння регресії X на Y , і це рівняння має вигляд $x(y) = 0.3283 + 1.1675y$ і може бути представлений у звичайній формі $(x) = -0.0191 + 0.9996x$. Регресії Y на X і X на Y можна об'єднати, аналогічно до підходу, описаного вище, використовуючи середнє геометричне:

$$y(x) = 0.1943 + 1.0047x. \quad (3.33)$$

3.4.4. Аналіз гетероскедастичності на основі інтервальної функції

Розглянемо другий підхід до побудови рівняння гетероскедастичності. Цей підхід базується на інтервальній функції, яка є залежністю різниці двох сусідніх значень порядкової статистики від їхньої напівсуми, тобто

$$y_{int} = \varphi(x_{int}), \quad (3.34)$$

$$y_{int\ i} = \left| Y_{i-1}^{(order)} - Y_i^{(order)} \right|, \quad x_{int} = 0.5 \left(Y_{i-1}^{(order)} + Y_i^{(order)} \right), \quad (3.35)$$

де $Y_{i-1}^{(order)}$ та $Y_i^{(order)}$ є значеннями статистичного порядку.

У цьому випадку рівнянням гетероскедастичності є інтервальна функція. В результаті використання OLS можна отримати таке рівняння: $H(y) = -0.6341 + 0.0711y$.

Вагові коефіцієнти обчислюються за формулою (3.32). Після цього регресія Y на X може бути знайдена у наступній формі $(x) = -0.9746 + 1.0129x$. Після аналогічних обчислень можна знайти рівняння регресії X на Y , і це рівняння має вигляд $x(y) = 0.015 + 0.9793y$ і може бути представлено у звичайній формі $y(x) = -0.0154 + 1.0211x$. Середня геометрична регресія становить:

$$y(x) = 0.106 + 1.017x. \quad (3.36)$$

3.4.5. Регресія на основі кластерного підходу

Усереднення даних за допомогою ковзної середньої або центроїдів може згладити часовий ряд, одночасно зменшивши варіацію дисперсії [8, 10]. Для даних з табл. 3.5 ми застосовуємо метод k -середніх кластерів для 4 кластерів. В результаті обчислень можна отримати кластери, як показано на рис. 3.10.

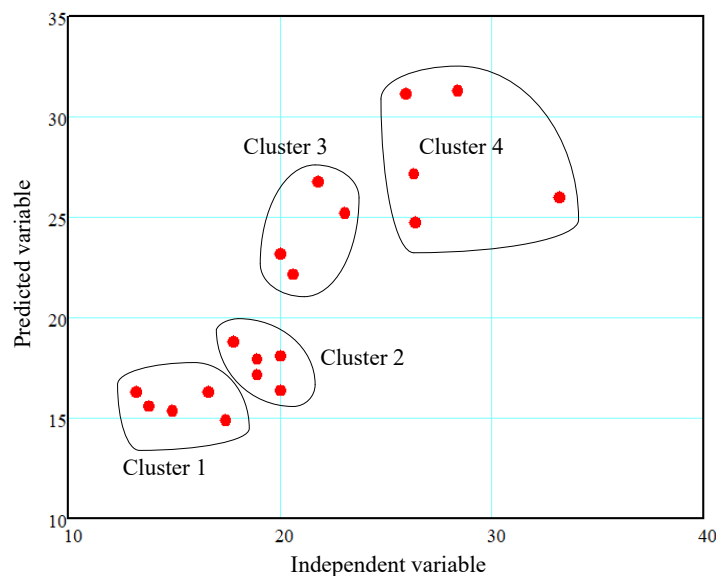


Рис. 3.10. Результати кластеризації за допомогою k -середніх.

Використовуючи координати чотирьох центроїдів кластерів, будується лінійна регресія на основі OLS, в результаті чого можна отримати рівняння регресії Y на X і X на Y : $y(x) = 0.2477 + 1.013x$ та $x(y) = 1.9837 + 0.8834y$. Друга регресія може бути представлена у виді $y(x) = -2.2454 + 1.1319x$. Середня геометрична регресія становить

$$y(x) = -1.1856 + 1.0708x. \quad (3.37)$$

3.4.6. Регресія на основі стиснення даних

Стиснемо поле кореляції. Для цього будемо варіаційні ряди для залежних і незалежних змінних. Отримані значення можна об'єднати в пари від найменшого до найбільшого значення. У цьому випадку отримуємо нову залежність між Y і X . В результаті цієї операції досягається збільшення коефіцієнта кореляції. Для вихідних даних він дорівнював 0,8325, а після стиснення збільшився до 0,9465. Результат стиснення показано на рис. 3.11.

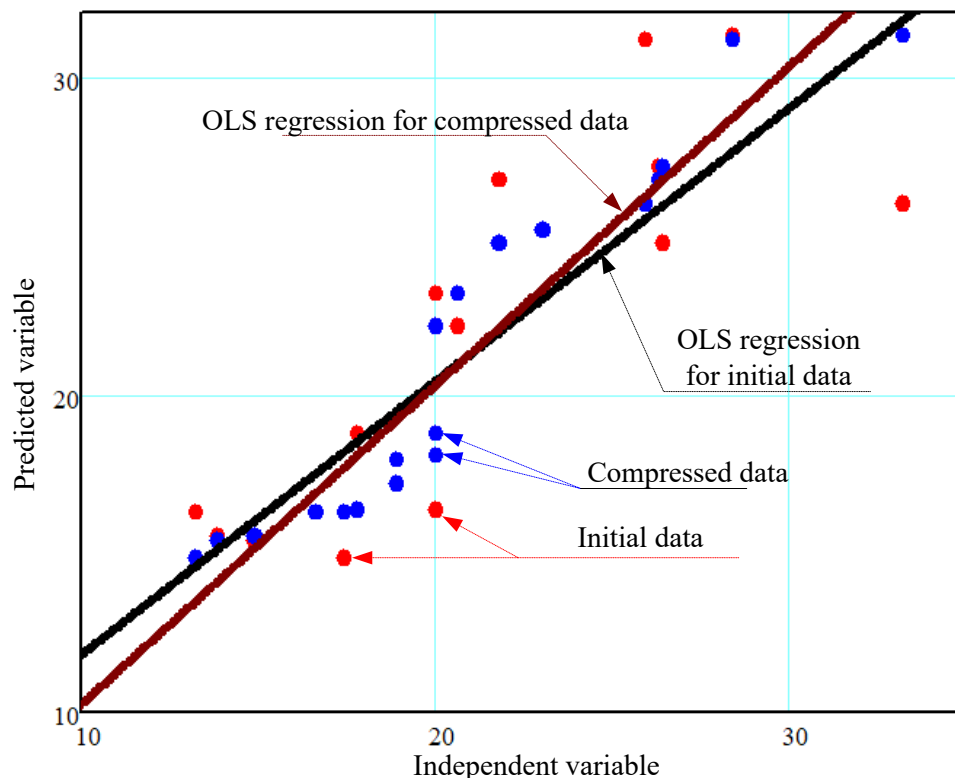


Рис. 3.11. Початкові та стиснені дані.

Для стиснених даних визначимо регресії Y на X і X на Y . У цьому випадку вони є такими: $(x) = 0.2641 + 1.0013x$ та $y(x) = -2.1639 + 1.1177x$.

Середня геометрична регресія становить:

$$y(x) = -0.9165 + 1.0579x. \quad (3.38)$$

3.4.7. Порівняльний аналіз розроблених моделей

На рис. 3.12 наведено результати апроксимації досліджуваного набору даних. Візуально побудовані моделі майже збігаються між собою. Однак моделі дещо відрізняються.

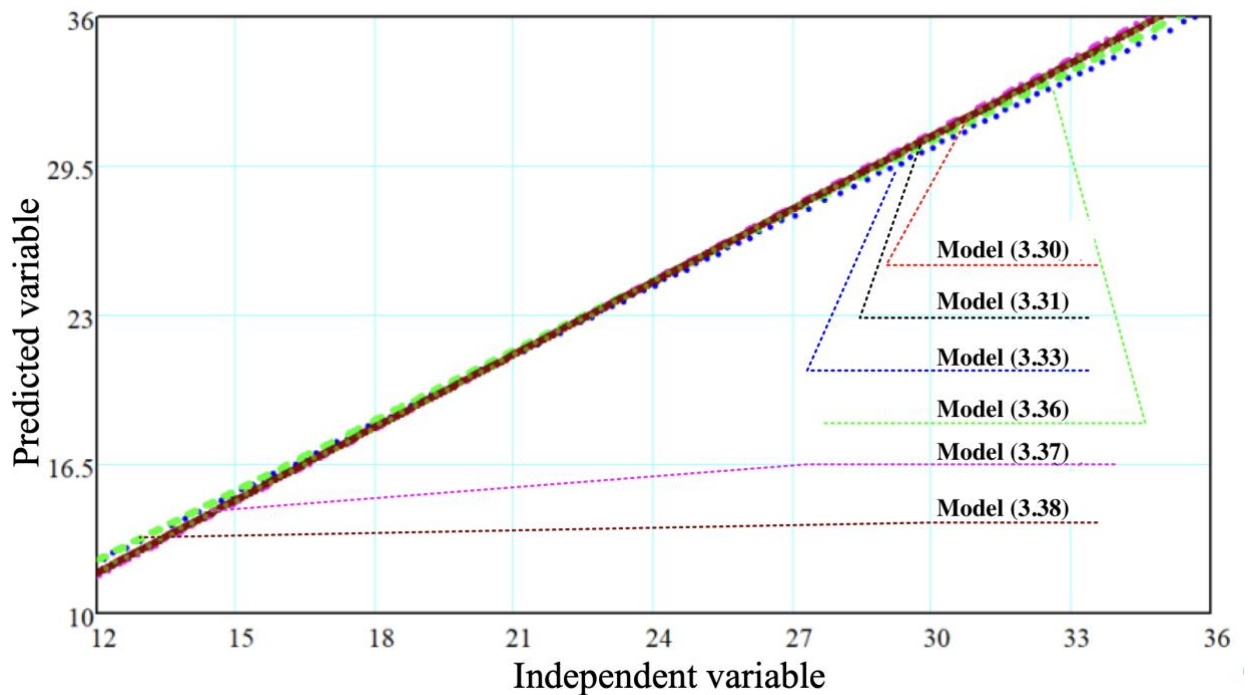


Рис. 3.12. Порівняння моделей.

Для вибору найкращої моделі було розраховано стандартне відхилення (SD) та коефіцієнт детермінації (DC). Додатковими параметрами є вільний та кутовий коефіцієнти лінійної регресії. В ідеалі, виходячи з характеру даних, цей коефіцієнт повинен бути додатним. Результати порівняння точності моделей наведено в табл. 3.6. Можна побачити, що модель (3.33) виявилася найкращою для досліджуваних даних. Модель (3.36) є дещо гіршою. Загалом,

врахування гетероскедастичності покращило точність моделі як з точки зору мінімізації стандартного відхилення, максимізації коефіцієнта детермінації, так і з точки зору отримання позитивного значення вільного коефіцієнта лінійної регресії. Врахування гетероскедастичності дає можливість довести неможливість ігнорування другої точки (випадкового значення) під час обчислення та отримати більш правильну математичну модель.

Таблиця 3.6

Параметри точності для різних моделей

Модель	SD	DC	Кутовий коефіцієнт	Вільний коефіцієнт
Модель (3.30)	3.1236	0.66495	1.0578	-0.9097
Модель (3.31)	3.1211	0.6655	1.0561	-0.8757
Модель (3.33)	3.0563	0.6792	1.0047	0.1943
Модель (3.36)	3.0747	0.6753	1.0170	0.106
Модель (3.37)	3.1435	0.6607	1.0708	-1.1856
Модель (3.38)	3.1238	0.6649	1.0579	-0.9165

Для визначення найкращої моделі було проведено їх порівняльний аналіз шляхом обчислення стандартного відхилення та коефіцієнта детермінації. Згідно з критерієм мінімального стандартного відхилення та максимального коефіцієнта детермінації, найкращими виявилися моделі (3.33) та (3.36). Отже, врахування гетероскедастичності є важливим етапом у побудові математичних моделей.

Надійність регресійної моделі була досягнута шляхом введення правильних коефіцієнтів зважування на основі рівняння гетероскедастичності.

3.5. Висновки до третього розділу

У третьому розділі дисертації виконано комплексне математичне моделювання процесу доставки коротких повідомлень (SMS) у мережах

4G/5G з метою кількісного оцінювання впливу параметрів якості обслуговування (QoS) на економічну ефективність сервісу, виражену через показник Price Delivery Gap (PDG).

На основі рекомендацій ETSI та ITU-T систематизовано ключові параметри надійності доставки SMS, зокрема ймовірність успішної доставки, часові затримки, частку втрат, повторні спроби передачі та коефіцієнт завантаження системи. Показано, що саме сукупна дія цих параметрів визначає фактичну ефективність доставки та формує економічні втрати, які відображаються через PDG.

Розроблено експоненційну та гіперболічну математичні моделі доставки SMS, які апроксимують часову динаміку ефективності у стабільному та перевантаженому режимах роботи мережі відповідно. Проведене моделювання підтвердило, що зростання навантаження та затримок призводить до нелінійного зниження ймовірності успішної доставки та відповідного зростання PDG, що знижує інтегральну ефективність IGDP.

На основі емпіричних даних CPaaS-провайдер здійснено оцінювання ризику невдач доставки з використанням запропонованих моделей. Отримані числові значення ризику в діапазоні 0,10–0,13 умовного ризику невдач доставки протягом нормативного часового вікна для агрегованого A2P-трафіку узгоджуються з реальними показниками ринку A2P SMS і підтверджують практичну придатність запропонованої методології для моніторингу та прогнозування ефективності доставки.

Показано, що реальні часові ряди параметрів доставки характеризуються високою нестабільністю та гетероскедастичністю, що обмежує застосовність класичних регресійних моделей. У зв'язку з цим у розділі розглянуто регресійні методи з урахуванням гетероскедастичності, зокрема аналіз на основі рухомого трикутника та інтервальної функції, які інтерпретуються як індикатори нестабільності доставки та дозволяють підвищити надійність оцінювання параметрів ефективності.

Запропоновано підходи кластеризованої регресії та регресії на основі стиснення даних, які забезпечують додаткове зменшення впливу шумів і неоднорідності даних. Показано, що кластеризована регресія дозволяє неявно враховувати різні режими роботи системи доставки, які в практичних SaaS-сценаріях корелюють з типами трафіку (транзакційний, маркетинговий, аномальний), навіть за відсутності явної категоризації повідомлень.

Отримані результати демонструють, що поєднання апроксимаційних моделей доставки з регресійними методами, адаптованими до гетероскедастичних даних, створює надійну математичну основу для коректного обчислення PDG у реальних умовах експлуатації SaaS-платформ. Водночас встановлено, що розглянуті моделі не враховують семантичні та поведінкові характеристики трафіку, а також штучно згенеровані аномалії (AIT, spam, smishing), що обмежує точність оцінювання інтегральної ефективності.

Оцінені регресійними моделями значення параметрів доставки використовуються як вхідні величини для розрахунку фактичної ефективності доставки, що безпосередньо входить до формули PDG.

Ці обмеження обґрунтовують необхідність подальшого розвитку моделей у напрямку інтелектуальної класифікації та аналізу трафіку. У наступному розділі дисертації буде розглянуто методи категоризації повідомлень і виявлення аномальних сценаріїв доставки, які доповнюють результати математичного моделювання та забезпечують перехід до побудови інтегрованого фреймворку IEMOF (Intelligent Efficiency Management and Optimization Framework).

Список джерел

1. ITU-T E.802. Framework and methodologies for service quality and Quality of Experience (QoE) assessment. – Geneva : ITU-T, 2021.
2. ETSI EG 202 009-3. Speech and multimedia transmission quality; QoS measurement methodologies. – Sophia Antipolis : ETSI, 2019.
3. Kleinrock L. Queueing Systems. Volume 1: Theory. – New York : Wiley-Interscience, 1975.
4. Kuzmin V., Zaliskyi M. Risk-oriented modeling of SMS delivery efficiency under varying load conditions // Proceedings of the International Conference on Information and Telecommunication Technologies and Radio Electronics (ICTM 2025). – 2025.
5. GSMA Fraud and Security Group. Artificially Inflated Traffic (AIT): Detection and Mitigation Guidelines. – London : GSMA, 2022.
6. ETSI TS 102 250-2 v.2.7.1 (2019-11), Speech and multimedia Transmission Quality. QoS aspects for popular services in mobile networks. Part 2: Definition of Quality of Service parameters and their computation, 2019.
7. J. Al-Azzeh, A. Mesleh, M. Zaliskyi, R. Odarchenko and V. Kuzmin, "A Method of Accuracy Increment Using Segmented Regression." Algorithms, 15 (10): 378 (2022)
8. Draper, N.R., Smith, H.S.: Applied regression analysis. Wiley, New York (1998).
9. Kuzmin, V.N.: New statistical method for identification of nonlinearity of empirical data. In: Fifth International Conference on Computer Data Analysis and Modeling, Volume 1: A-M, pp. 159–164 (1998).
10. Lawrence, K. D., Arthur, J. L.: Robust Regression: Analysis and Applications. CRC Press, Boca Raton (2019).

РОЗДІЛ 4.

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ДОСТАВКИ SMS У МЕРЕЖАХ 4G/5G

4.1. Обмеження математичних моделей та необхідність інтелектуального аналізу процесу доставки SMS

Результати математичного моделювання, отримані в попередньому розділі, демонструють можливість кількісного оцінювання ефективності доставки SMS на основі параметрів якості обслуговування та часових характеристик процесу доставки. Запропоновані моделі дозволяють отримувати формалізовані оцінки фактичної ефективності доставки та показника PDG у різних режимах роботи системи. Водночас підходи, що базуються на агрегованих QoS-метриках, мають принципові обмеження, пов'язані з їх здатністю відображати реальну якість сервісу та економічну ефективність у складних і динамічних середовищах доставки повідомлень [1–3].

У сучасних CPaaS-системах у мережах 4G/5G процес доставки SMS характеризується високою неоднорідністю A2P-трафіку, наявністю різних сервісних сценаріїв та динамічною зміною режимів навантаження. Як зазначається в рекомендаціях ETSI та ITU-T, однакові значення класичних QoS-показників можуть по-різному впливати на сприйману якість та цінність сервісу залежно від контексту використання [2, 3]. Це означає, що математичні моделі, які оперують усередненими характеристиками доставки, не завжди дозволяють коректно оцінити фактичну ефективність сервісу.

Одним із ключових обмежень суто математичних моделей є їхня контекстна нечутливість. У попередніх розділах ефективність доставки оцінювалася на основі інтегральної ймовірності доставки, затримки та ризику

невдач. Проте з погляду реальних сервісних сценаріїв ці параметри мають різну вагу для різних типів повідомлень. Зокрема, затримка доставки в межах кількох секунд є критичною для транзакційних повідомлень (наприклад, одноразових паролів), тоді як для маркетингових повідомлень вона практично не впливає на споживчу цінність доставки [4, 5]. У межах універсальних математичних моделей така різниця не враховується, що призводить до спотворення оцінки економічної ефективності.

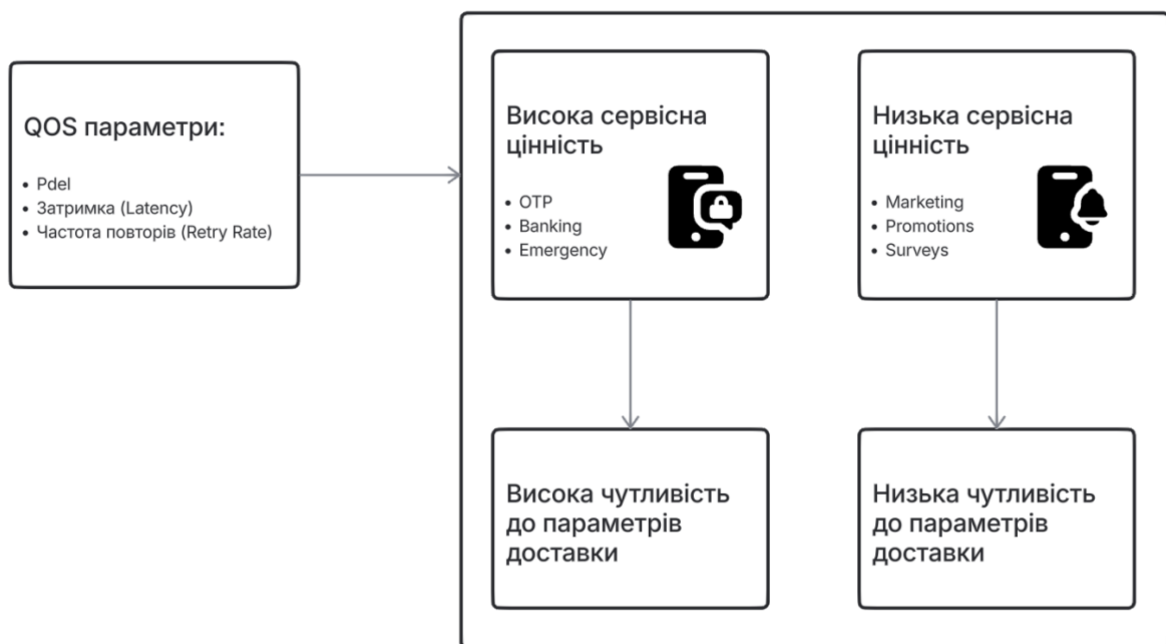


Рис. 4.1. Обмеження агрегованих QoS-моделей у змішаному A2P-трафіку

Додатковим фактором, який істотно обмежує застосовність класичних моделей, є наявність штучно згенерованого трафіку (Artificially Inflated Traffic, AIT). Як зазначається в рекомендаціях GSMA, AIT може формально відповідати допустимим технічним характеристикам доставки, зокрема мати високі показники успішності, але при цьому не створювати реальної бізнес-цінності та призводити до прямих фінансових втрат [6]. У таких умовах математичні моделі коректно відображають технічну успішність доставки, однак не здатні відрізнити легітимний трафік від штучно згенерованого, що призводить до викривлення оцінки PDG.

Результати розділу 3 також показали, що навіть із використанням регресійних методів з урахуванням гетероскедастичності параметри доставки залишаються чутливими до зміни латентних режимів роботи системи. Кластеризована регресія дозволяє частково враховувати цю неоднорідність, проте отримані кластери мають статистичну природу і не забезпечують явної інтерпретації з погляду типів трафіку або семантики повідомлень [7, 8]. Це суттєво обмежує можливість використання таких моделей для оперативного управління доставкою та прийняття рішень у реальному часі.

Таким чином, математичні моделі доставки SMS формують необхідну основу для кількісного оцінювання ефективності, проте не є достатніми для повноцінного управління процесом доставки в сучасних CPaaS-системах. Виявлені обмеження зумовлюють необхідність інтелектуального аналізу, який дозволяє враховувати контекст доставки, тип повідомлень, поведінкові особливості трафіку та аномальні сценарії, що виходять за межі класичних статистичних припущень [1, 6].

У наступних підрозділах цього розділу розглядаються інтелектуальні методи підвищення ефективності доставки SMS, зокрема категоризація повідомлень, виявлення штучно згенерованого трафіку та інтеграція цих підходів у єдиний фреймворк управління ефективністю доставки, що забезпечує перехід від пасивного оцінювання PDG до активного управління ним.

4.2. Категоризація повідомлень як основа адаптивного управління ефективністю доставки

Як показано в підрозділі 4.1, універсальні математичні моделі доставки SMS, що базуються на агрегованих показниках якості обслуговування, не здатні повною мірою враховувати контекст використання сервісу та семантичні відмінності між різними типами повідомлень. У сучасних A2P-

сценаріях доставка SMS охоплює широкий спектр сервісних задач – від критично важливих транзакційних повідомлень до масових інформаційних та маркетингових розсилок. Ігнорування цієї неоднорідності призводить до спотворення інтегральних показників ефективності доставки, зокрема PDG, і знижує практичну цінність отриманих оцінок [9, 10].

Згідно з рекомендаціями міжнародних стандартів та галузевих організацій, вимоги до якості доставки повідомлень суттєво залежать від їх призначення. Для транзакційних повідомлень, зокрема одноразових паролів, банківських та безпекових сповіщень, ключовими параметрами є мінімальна затримка доставки та висока ймовірність успішного отримання. Водночас для маркетингових або інформаційних повідомлень допустимі ширші часові вікна доставки без суттєвого впливу на сприйману цінність сервісу [11, 12]. Це обумовлює необхідність поділу A2P-трафіку на категорії з різними вимогами до ефективності доставки.

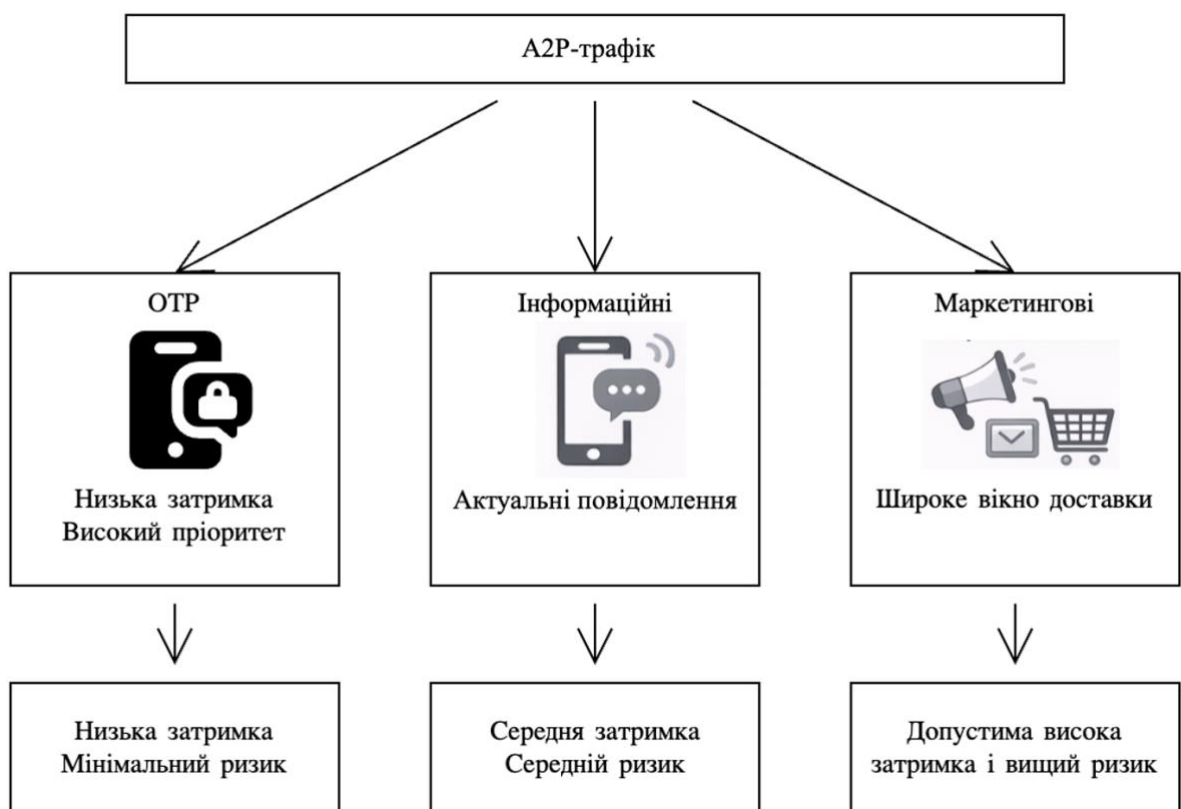


Рис.4.2. Категоризація A2P-повідомлень за сервісними вимогами

У межах запропонованої методології інтегрального оцінювання ефективності доставки категоризація повідомлень дозволяє інтерпретувати показник PDG не як універсальну метрику, а як контекстно залежну величину, значення якої формується з урахуванням типу повідомлення та його сервісної ролі. Таким чином, однакові значення ймовірності доставки або затримки можуть по-різному впливати на значення PDG для різних категорій повідомлень, що є принципово важливим з точки зору економічної ефективності сервісу [12].

Практична реалізація категоризації повідомлень може базуватися на аналізі метаданих, історичних характеристик доставки та поведінкових ознак трафіку. У цьому контексті результати кластеризованої регресії, отримані в розділі 3, свідчать про існування латентних режимів доставки, які відрізняються за параметрами затримки, стабільності та рівнем ризику невдач. Хоча такі кластери мають статистичну природу, вони можуть бути інтерпретовані як наближене відображення різних типів або підтипів трафіку, що створює основу для подальшої семантичної категоризації повідомлень.

Поєднання категоризації повідомлень із математичними моделями доставки дозволяє перейти від статичного оцінювання ефективності до адаптивного управління процесом доставки. Для кожної категорії можуть бути визначені власні цільові значення ефективності, допустимі часові вікна доставки та порогові значення ризику, які безпосередньо враховуються при розрахунку PDG. Такий підхід узгоджується з рекомендаціями щодо контекстно орієнтованого оцінювання якості телекомунікаційних сервісів, наведеними в стандартах ITU-T та ETSI [2, 6].

Додатковою перевагою категоризації є можливість локалізації негативного впливу аномальних сценаріїв доставки. Зокрема, відокремлення нетипового або штучно згенерованого трафіку в окрему категорію дозволяє уникнути викривлення інтегральних показників ефективності для легітимних повідомлень і забезпечити більш стабільні оцінки PDG, що є критично важливим для практичного управління доставкою [14].

Таким чином, категоризація повідомлень є ключовим елементом інтелектуального управління ефективністю доставки SMS. Вона забезпечує зв'язок між статистичними моделями доставки та реальними сервісними вимогами і створює підґрунтя для адаптивного прийняття рішень у CPaaS-системах мереж 4G/5G.

4.3. Виявлення штучно згенерованого трафіку (AIT) як чинник підвищення достовірності оцінювання ефективності доставки

Однією з ключових практичних проблем, що обмежують застосовність класичних моделей оцінювання ефективності доставки SMS, є наявність штучно згенерованого трафіку (Artificially Inflated Traffic, AIT). Згідно з визначенням GSMA, AIT являє собою трафік, який створюється з метою штучного збільшення обсягу повідомлень без реальної сервісної або бізнес-цінності, що призводить до фінансових втрат для операторів та постачальників CPaaS-послуг [13].

Особливістю AIT є те, що з технічної точки зору такий трафік часто відповідає формальним вимогам до якості доставки: повідомлення можуть успішно доставлятися, підтверджуватися мережею та мати допустимі значення затримки. Унаслідок цього показники, засновані виключно на QoS-метриках або інтегральній ймовірності доставки, не дозволяють коректно відрізнити легітимний A2P-трафік від штучно згенерованого [14]. Це призводить до викривлення оцінки фактичної ефективності сервісу та заниження значень показника PDG.

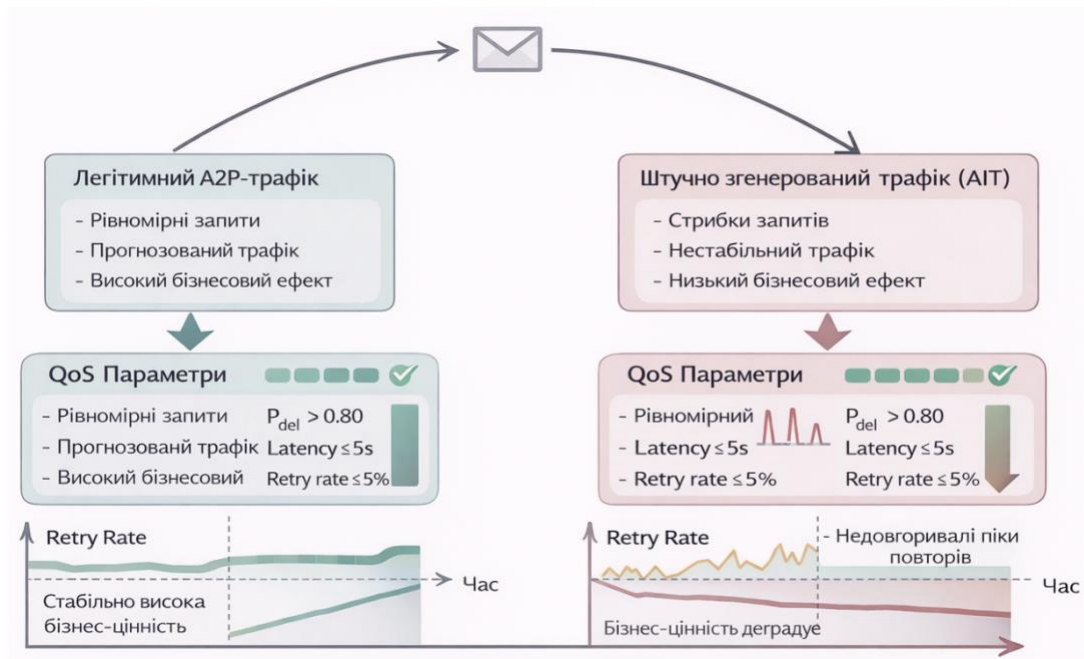


Рис. 4.3. Прояви штучно згенерованого трафіку в параметрах доставки SMS

На рис. 4.3 наведено порівняння проявів легітимного A2P-трафіку та АІТ у параметрах доставки SMS. Видно, що за дотримання допустимих значень технічних метрик доставки (P_{del} , latency, retry-rate) інтегральна економічна ефективність каналу може суттєво знижуватися.

У межах запропонованої методології інтегрального оцінювання ефективності АІТ розглядається не лише як фродова активність, а як фактор ризику, що безпосередньо впливає на економічну ефективність доставки повідомлень. Зростання частки АІТ у загальному трафіку призводить до збільшення витрат на доставку без відповідного зростання корисного результату, що відображається у негативній динаміці PDG навіть за стабільних технічних показників доставки.

Виявлення АІТ у системах доставки SMS ґрунтується на аналізі поведінкових та статистичних ознак трафіку, зокрема часових патернів, кореляцій між обсягом повідомлень та підтвердженнями доставки, а також нестабільності показників ефективності в окремих часових інтервалах. Як показано в авторській роботі [15], застосування методів аналізу часових рядів та гетероскедастичності дозволяє ідентифікувати аномальні режими трафіку,

характерні для АІТ, навіть у випадках, коли середні значення QoS-метрик залишаються в допустимих межах.

Важливою перевагою такого підходу є можливість інтеграції результатів детекції АІТ з методами кластеризації, розглянутими в розділі 3. Аномальні кластери, що характеризуються підвищеною нестабільністю параметрів доставки та непропорційним зростанням обсягів трафіку, можуть бути інтерпретовані як кандидати на АІТ та відокремлені від легітимних категорій повідомлень. Це дозволяє локалізувати негативний вплив АІТ і запобігти його поширенню на інтегральні показники ефективності для корисного трафіку.

Інтеграція механізмів виявлення АІТ у процес оцінювання ефективності доставки забезпечує перехід від пасивного моніторингу до активного управління ризиками. Зокрема, виключення або зниження ваги АІТ-трафіку при розрахунку PDG дозволяє отримувати більш достовірні оцінки фактичної ефективності сервісу та приймати обґрунтовані управлінські рішення щодо маршрутизації, тарифікації та застосування альтернативних каналів доставки [16-18].

У практичних системах доставки SMS результати виявлення штучно згенерованого трафіку можуть використовуватися не лише для аналітичних цілей, але й для адаптивного управління процесом доставки. Залежно від рівня впевненості детекції та сервісного контексту, реакція системи може включати обмеження швидкості доставки, зміну маршрутів, корекцію пріоритетів або тимчасове виключення відповідного трафіку з процесу розрахунку інтегральних показників ефективності.

Важливо зазначити, що блокування АІТ не розглядається як бінарне рішення, а реалізується у вигляді багаторівневої стратегії управління ризиками. На початкових етапах аномальний трафік може підлягати підвищеному моніторингу або зниженню ваги при обчисленні PDG, тоді як за підтверджених ознак штучної генерації можуть застосовуватися більш жорсткі обмеження доставки. Такий підхід дозволяє мінімізувати ризик

хибнопозитивних рішень і забезпечити баланс між захистом від економічних втрат та збереженням якості обслуговування легітимних клієнтів [15].

Інтеграція механізмів виявлення та керованого обмеження АІТ у загальний процес оцінювання ефективності доставки дозволяє перейти від пасивної фіксації фродової активності до проактивного управління ефективністю сервісу. У цьому випадку зменшення негативного впливу АІТ безпосередньо відображається у стабілізації значень PDG та підвищенні достовірності прийняття управлінських рішень щодо маршрутизації та тарифікації повідомлень.

Таким чином, виявлення штучно згенерованого трафіку є невід'ємним елементом інтелектуального управління ефективністю доставки SMS. Поєднання методів детекції АІТ з категоризацією повідомлень і математичними моделями доставки створює передумови для формування інтегрованого фреймворку управління ефективністю, який буде розглянуто в наступному підрозділі.

4.4. Інтегрований фреймворк інтелектуального управління ефективністю доставки SMS

Результати попередніх підрозділів показали, що ефективність доставки SMS у сучасних CPaaS-системах формується під впливом сукупності технічних, поведінкових та економічних чинників. Суто математичні моделі доставки дозволяють кількісно оцінювати параметри якості обслуговування, однак не забезпечують урахування контексту використання сервісу, типів повідомлень та аномальних сценаріїв трафіку. З іншого боку, інтелектуальні методи аналізу, зокрема категоризація повідомлень і виявлення штучно згенерованого трафіку, потребують формалізованої основи для прийняття узгоджених управлінських рішень. Це зумовлює необхідність побудови

інтегрованого фреймворку, який поєднує кількісне оцінювання ефективності з інтелектуальними механізмами управління.

Запропонований у межах цієї роботи фреймворк інтелектуального управління ефективністю доставки SMS (рис. 4.4) базується на показнику PDG як інтегральній метриці, що відображає економічну доцільність доставки повідомлень з урахуванням фактичної ефективності, ризиків та сервісного контексту. На відміну від традиційних підходів, PDG у межах фреймворку розглядається не лише як результат вимірювання, а як керований параметр, значення якого може цілеспрямовано змінюватися шляхом адаптації процесу доставки.

Як альтернативні канали доставки можуть використовуватись RCS та OTT-платформи (зокрема WhatsApp), що підтримують A2P-взаємодію відповідно до специфікацій GSMA та виробників платформ [16 – 18].

Структурно фреймворк складається з кількох взаємопов'язаних рівнів. На базовому рівні здійснюється збір та агрегація параметрів доставки, включаючи часові характеристики, показники успішності та стабільності доставки, що формують вхідні дані для математичних моделей, розглянутих у розділі 3. На цьому рівні обчислюються кількісні оцінки ефективності та ризику, які використовуються як формалізована основа для подальшого аналізу.

На наступному рівні реалізується категоризація повідомлень, що дозволяє інтерпретувати отримані оцінки ефективності з урахуванням типу трафіку та сервісних вимог. Для кожної категорії повідомлень визначаються власні цільові значення ефективності, допустимі часові вікна доставки та порогові значення ризику, що безпосередньо враховуються при обчисленні PDG. Таким чином, фреймворк забезпечує перехід від універсальних оцінок до контекстно залежного управління доставкою.

Окремим функціональним компонентом фреймворку є модуль виявлення та керованого обмеження штучно згенерованого трафіку (AIT). Результати детекції AIT використовуються для адаптації процесу доставки, зокрема шляхом корекції пріоритетів, зміни маршрутів або зниження ваги відповідного

трафіку при розрахунку інтегральних показників ефективності. Такий підхід дозволяє локалізувати негативний вплив АІТ та запобігти викривленню значень PDG для легітимного трафіку без застосування жорстких бінарних рішень.

Інтеграція зазначених компонентів забезпечує реалізацію замкненого контуру управління ефективністю доставки. На основі поточних значень PDG та результатів інтелектуального аналізу система формує управлінські дії, спрямовані на оптимізацію процесу доставки, після чого оновлені параметри доставки знову використовуються для оцінювання ефективності. Такий ітеративний підхід узгоджується з рекомендаціями міжнародних стандартів щодо управління якістю телекомунікаційних сервісів і дозволяє адаптувати систему доставки до динамічних змін навантаження та структури трафіку.

На відміну від існуючих підходів, запропонований фреймворк поєднує категоризацію повідомлень, АІТ-детекцію та адаптивну мультиканальну маршрутизацію в єдиній моделі з використанням показника PDG як критерію прийняття рішень

Запропонований інтегрований фреймворк не прив'язаний до конкретної реалізації мережевої інфраструктури або постачальника послуг і може бути застосований як у операторських мережах, так і в SPaaS-платформах. Його модульна структура забезпечує можливість розширення за рахунок додавання альтернативних каналів доставки, таких як OTT або RCS, а також адаптації до нових типів сервісів і сценаріїв використання. Це створює підґрунтя для практичного впровадження інтелектуального управління ефективністю доставки SMS у мережах 4G/5G.

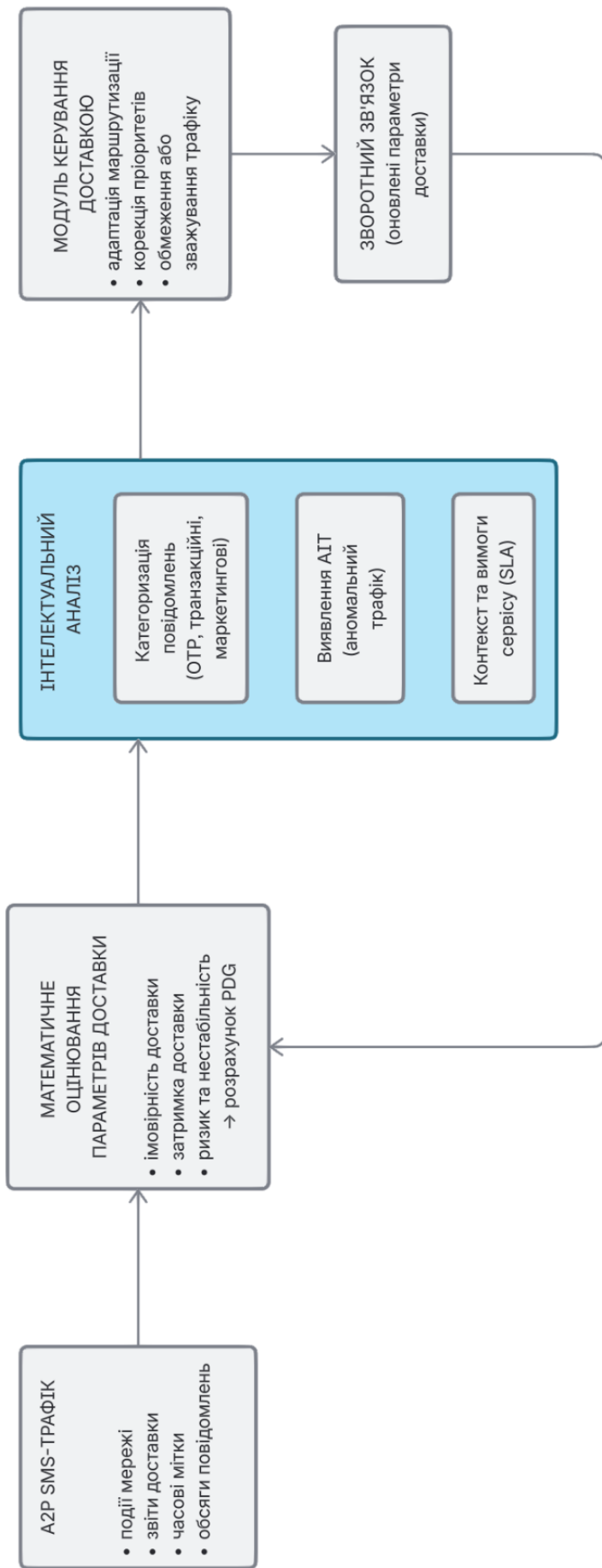


Рис. 4.4. Інтегрований фреймворк інтелектуального управління ефективністю доставки SMS

4.5. Сценарна оцінка ефективності інтегрованого фреймворку управління доставкою SMS

У цьому підрозділі проведено сценарну оцінку ефективності запропонованого інтегрованого фреймворку управління доставкою повідомлень у змішаному A2P-трафіку, що включає легальний OTP- та маркетинговий трафік за наявності домішки штучно згенерованого трафіку (AIT). Метою оцінки є кількісна демонстрація впливу категоризації повідомлень, виявлення AIT та адаптивного використання альтернативних каналів доставки на інтегральний показник ефективності PDG у реалістичних операторських сценаріях.

Оцінка проводиться для двох операторів (далі – Оператор А та Оператор В), які відрізняються тарифною політикою, характеристиками якості доставки та покриттям альтернативних каналів. Це дозволяє перевірити універсальність запропонованого підходу та уникнути прив'язки результатів до одного конкретного профілю мережі.

4.5.1. Вихідні умови та сценарії

Розглядається змішаний A2P-трафік зі сталою структурою:

- частка OTP-повідомлень – 30%;
- частка маркетингових повідомлень – 70%.

Штучно згенерований трафік присутній лише в OTP-сегменті, що відповідає практичним сценаріям фродової генерації OTP-запитів. Для маркетингового трафіку AIT не враховується.

Для оцінки ефективності розглянуто такі сценарії:

1. Сценарій 1 – SMS тільки, без фреймворку.

Увесь трафік доставляється через SMS без категоризації повідомлень, без виявлення AIT та без адаптивного управління.

2. Сценарій 2 – Мультиканальна доставка без інтелектуального управління.

Використовуються SMS та альтернативні канали (RCS, WhatsApp) за статичними правилами fallback без урахування типу повідомлень та ризику АІТ.

3. Сценарій 3 – Категоризація та АІТ-детекція (тільки в SMS).

ОТР- та маркетинговий трафік розділяються, АІТ у ОТР локалізується, проте альтернативні канали доставки не використовуються.

4. Сценарій 4 – Інтегрований фреймворк.

Застосовується категоризація повідомлень, виявлення АІТ та адаптивне використання SMS і альтернативних каналів доставки з урахуванням вартості, затримки, покриття та ризику доставки.

4.5.2. Методика розрахунку показника PDG

Для сценарного аналізу використано нормалізовану форму показника PDG, що дозволяє порівнювати відносну ефективність різних стратегій доставки за фіксованих вагових коефіцієнтів.

Ефективність доставки для типу повідомлень $k \in \{\text{ОТР, МКТ}\}$ через канал $c \in \{\text{SMS, RCS, WA}\}$ визначається як:

$$E_{k,c} = (P_c \cdot C_c) \cdot (1 - AIT_{k,c}) - \alpha \cdot Cost_{k,c} - \beta \cdot L_c,$$

де P_c – імовірність успішної доставки; C_c – частка абонентської бази, охоплена каналом; $AIT_{k,c}$ – частка штучно згенерованого трафіку; $Cost_{k,c}$ – вартість доставки одного повідомлення; L_c – середня затримка доставки; α, β – нормувальні коефіцієнти, однакові для всіх сценаріїв.

Якщо тип повідомлень доставляється через декілька каналів, ефективність агрегується як:

$$E_k = \sum_c \omega_{k,c} \cdot E_{k,c},$$

де $w_{k,c}$ – частка трафіку типу k , маршрутизована через канал c .

Загальна ефективність кампанії:

$$E_{total} = \lambda_{OTP} \cdot E_{OTP} + \lambda_{MKT} \cdot E_{MKT} ,$$

Для порівняння сценаріїв використано нормалізований показник:

$$PDG_s = \frac{E_{total,S}}{E_{total,S1}} ,$$

де сценарій 1 (S1) приймається за базовий.

4.5.3. Числові результати сценарної оцінки

На основі описаної методики виконано розрахунок нормалізованого показника PDG для сценаріїв S1–S4 для двох операторів (табл. 4.1 – 4.4). Сценарій S1 (SMS тільки, без фреймворку) прийнято за базовий, його значення PDG нормалізовано до 1.00 для кожного оператора окремо.

Таблиця 4.1

Вхідні параметри доставки (Оператор А)

Параметр	SMS	RCS	WhatsApp
Ціна за повідомлення (OTP)	0.035	0.030	0.030
Ціна за повідомлення (Marketing)	0.035	0.040	0.060
Coverage, %	99	70	50
Середня затримка, с	23	15	19
Успішність доставки, %	90	95	92
AIT у OTP, %	15	0	0

Таблиця 4.2

Сценарії для оператора А

Сценарій	Опис сценарію	E_{total}	Нормалізований PDG
S1	SMS only, без фреймворку	0.586	1.00
S2	Multi-channel без категоризації	0.557	0.95
S3	Категоризація + AIT (SMS)	0.732	1.25
S4	Інтегрований фреймворк	0.593	1.12

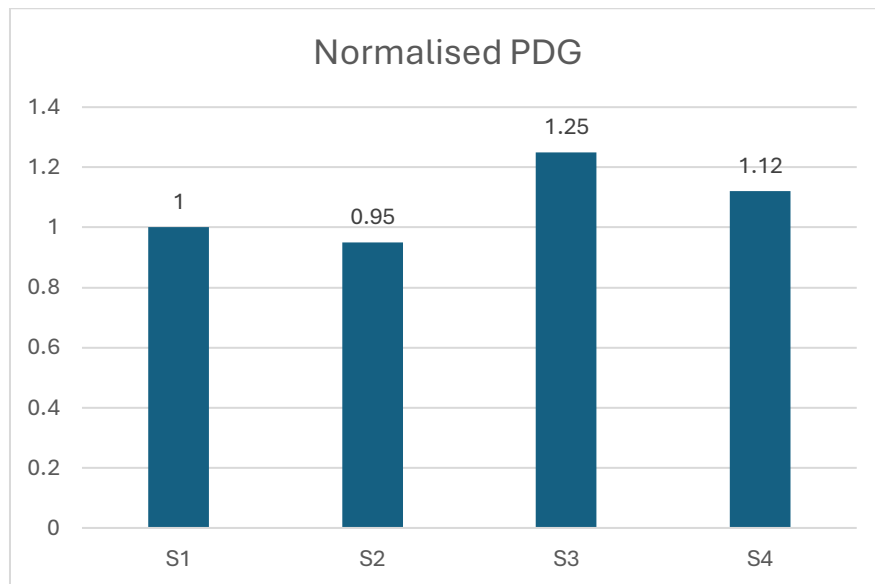


Рис. 4.5. Порівняння сценаріїв доставки за нормалізованим показником PDG для оператора А

Як видно з рис. 4.5, використання лише мультимедійної доставки без інтелектуального управління (S2) не забезпечує підвищення економічної ефективності порівняно з базовим сценарієм. Максимальне значення PDG досягається в сценарії S3, що відповідає локальній оптимізації доставки SMS з урахуванням категоризації та виявлення АІТ. У сценарії S4 застосовується глобальна оптимізація доставки з використанням альтернативних каналів, що призводить до дещо нижчого значення PDG, однак забезпечує підвищення надійності та покриття доставки в цілому.

Таблиця 4.3

Вхідні параметри доставки (Оператор В)

Параметр	SMS	RCS	WhatsApp
Ціна за повідомлення (ОТР)	0,025	0,025	0,03
Ціна за повідомлення (Marketing)	0,025	0,02	0,015
Coverage, %	99	80	60
Середня затримка, с	19	20	15
Успішність доставки, %	88	93	94
АІТ у ОТР, %	10	0	0

Сценарії для оператора В

Сценарій	Опис	E_{total}	Нормалізований PDG
S1	SMS only, без фреймворку	0.630	1.00
S4	Інтегрований фреймворк	0.623	0.99



Рис. 4.6. Порівняння сценаріїв доставки за нормалізованим показником PDG для оператора В

Для Оператора В сценарії S2 та S3 не наведено окремо, оскільки за рахунок нижчої вартості SMS-доставки та меншої частки АІТ базовий сценарій S1 уже забезпечує близьку до оптимальної ефективність. У таких умовах проміжні сценарії не призводять до суттєвих змін показника PDG і не дають додаткової інформації щодо впливу інтегрованого фреймворку. Тому для цього оператора проведено безпосереднє порівняння базового сценарію S1 та інтегрованого сценарію S4 з метою перевірки універсальності запропонованого підходу.

Для Оператора В базовий рівень ефективності SMS-доставки є вищим за рахунок нижчої вартості та кращих характеристик затримки, унаслідок чого

абсолютний виграш від застосування інтегрованого фреймворку є меншим. Водночас результати демонструють, що використання фреймворку не погіршує PDG, забезпечуючи додаткову стабільність оцінок і контроль ризиків АІТ.

4.5.4. Аналіз та інтерпретація результатів

Числові результати сценарної оцінки підтверджують, що застосування інтегрованого фреймворку управління доставкою дозволяє підвищити ефективність легального А2Р-трафіку в умовах наявності штучно згенерованого ОТР-трафіку. Для Оператора А сценарій S4 забезпечує зростання показника PDG на 12% порівняно з базовим сценарієм, при цьому основний внесок у покращення досягається не за рахунок покращення мережових QoS-параметрів, а завдяки локалізації впливу АІТ та контекстному використанню альтернативних каналів доставки.

Порівняння сценаріїв S2 та S4 є принципово важливим: воно показує, що просте додавання альтернативних каналів без категоризації повідомлень та оцінки ризиків не лише не гарантує підвищення ефективності, але може призводити до її зниження. Натомість інтеграція категоризації, АІТ-детекції та адаптивної маршрутизації дозволяє перетворити багатоканальну доставку на інструмент підвищення PDG.

Для Оператора В результати підтверджують універсальність підходу: навіть у випадках, коли базова ефективність доставки є високою, інтегрований фреймворк не знижує PDG і забезпечує додаткову керованість процесу доставки. Це свідчить про можливість застосування запропонованого фреймворку в мережах з різними економічними та технічними характеристиками без ризику деградації ефективності.

4.5.5. Ключовий результат для сценарної оцінки ефективності інтегрованого фреймворку управління доставкою SMS

Основним результатом підрозділу є кількісне підтвердження того, що інтегрований фреймворк дозволяє підвищити або стабілізувати показник PDG для легального A2P-трафіку в умовах присутності АІТ, використовуючи інтелектуальне управління доставкою замість покращення мережевих параметрів.

4.6. Висновки до четвертого розділу

У четвертому розділі дисертаційної роботи розроблено та досліджено інтегрований фреймворк інтелектуального управління ефективністю доставки A2P-повідомлень у мережах 4G/5G. Основною ідеєю запропонованого підходу є перехід від пасивного вимірювання параметрів якості доставки до активного управління процесом доставки на основі інтегрального показника PDG, що враховує не лише технічні характеристики мережі, але й сервісний контекст та економічні чинники.

У розділі обґрунтовано доцільність використання категоризації повідомлень як базового механізму контекстної інтерпретації показників доставки. Показано, що розділення A2P-трафіку на категорії з різними вимогами до затримки, надійності та вартості доставки дозволяє коректніше оцінювати ефективність сервісу та уникати викривлення інтегральних показників у змішаних сценаріях використання.

Окрему увагу приділено впливу штучно згенерованого трафіку (АІТ) на ефективність доставки критичних повідомлень, зокрема ОТР. Показано, що АІТ може істотно знижувати економічну ефективність доставки навіть за формально задовільних показників QoS. У межах запропонованого фреймворку АІТ розглядається як фактор ризику, результати виявлення якого

використовуються для корекції процесу доставки та розрахунку показника PDG, що дозволяє локалізувати негативний вплив неефективного трафіку.

Розроблений інтегрований фреймворк поєднує математичні моделі доставки, категоризацію повідомлень, механізми виявлення АІТ та адаптивне використання альтернативних каналів доставки (SMS, RCS, OTT) у замкнений контур управління. Така структура забезпечує можливість цілеспрямованого впливу на значення PDG шляхом прийняття управлінських рішень щодо маршрутизації, пріоритетів доставки та використання каналів зв'язку.

Сценарна оцінка ефективності запропонованого підходу, виконана для двох операторів з різними тарифними та мережевими характеристиками, підтвердила практичну цінність фреймворку. Для оператора з вищою вартістю доставки та більшою часткою АІТ застосування інтегрованого фреймворку забезпечує зростання показника PDG порівняно з базовим сценарієм, тоді як для оператора з уже високою базовою ефективністю доставка не зазнає погіршення, що свідчить про універсальність та стабільність запропонованого рішення.

Ключовим результатом четвертого розділу є демонстрація того, що підвищення ефективності доставки А2Р-повідомлень може досягатися без покращення мережових QoS-параметрів, виключно за рахунок інтелектуального управління процесом доставки, урахування контексту повідомлень та ризиків, пов'язаних із штучно згенерованим трафіком. Це підтверджує доцільність застосування запропонованого фреймворку в практичних CPaaS-системах і створює підґрунтя для подальшого розвитку методів оцінювання та оптимізації доставки повідомлень у мережах наступних поколінь.

Список джерел

1. ITU-T Recommendation E.800. Definitions of terms related to quality of service. — International Telecommunication Union, Geneva.
2. ITU-T Recommendation E.802. Framework and methodologies for the determination and measurement of QoS parameters. — International Telecommunication Union, Geneva.
3. ITU-T Recommendation G.1011. Reference guide to quality of experience assessment methodologies. — International Telecommunication Union, Geneva.
4. ETSI EG 202 009-1. Speech and multimedia transmission quality (STQ); QoS aspects for popular services. — European Telecommunications Standards Institute.
5. ETSI EG 202 009-3. Speech and multimedia transmission quality (STQ); QoS aspects for messaging services in mobile networks. — European Telecommunications Standards Institute.
6. ETSI TR 103 559. Speech and multimedia transmission quality (STQ); End-to-end QoS for messaging services. — European Telecommunications Standards Institute.
7. ETSI EG 203 215. Quality of Experience (QoE) requirements for messaging services. — European Telecommunications Standards Institute.
8. 3GPP TS 23.040. Technical realization of the Short Message Service (SMS).
9. 3GPP TS 23.682. Architecture enhancements to facilitate communications with packet data networks.
10. 3GPP TS 29.118. SMS over IP (SoIP).
11. 3GPP TS 23.501. System Architecture for the 5G System (5GS).
12. GSMA. A2P Messaging Guidelines and Best Practices. — GSM Association.
13. GSMA Fraud and Security Group. Artificially Inflated Traffic (AIT): Definition, Detection and Mitigation. — GSM Association.
14. GSMA Fraud and Security Group. Messaging Fraud and Security Guidelines. — GSM Association.

15. Mykhailo Odarchenko, Maksym Zaliskiy (2026). “DETECTION OF ARTIFICIALLY INFLATED TRAFFIC IN A2P SMS USING PREFIX-BASED TIME-WINDOW ANALYSIS”. *Science-Based Technologies* 68 (1).
16. GSMA. RCS Universal Profile Specification. — GSM Association.
17. Google. RCS Business Messaging: Technical Overview. — Google Jibe.
18. Meta Platforms Inc. WhatsApp Business Platform Documentation.

ЗАГАЛЬНІ ВИСНОВКИ

1. У дисертаційній роботі вирішено актуальну науково-практичну задачу підвищення ефективності надання сервісу коротких повідомлень (SMS) у мережах 4G/5G шляхом інтеграції технічних, користувацьких та економічних факторів оцінювання, а також мінімізації впливу штучно згенерованого трафіку (AIT) на показники якості та вартості доставки.

2. Проведений аналіз еволюції архітектури SMS від 2G до 5G SA показав, що перехід до Service-Based Architecture та впровадження функції SMSF призвели до зростання варіативності параметрів доставки та появи економічного розриву між номінальною та фактичною ефективною вартістю повідомлення, що потребує нових інтегрованих підходів до оцінювання ефективності.

3. Обґрунтовано доцільність використання інтегрального показника Price Delivery Gap (PDG) як універсальної техніко-економічної метрики, що відображає розрив між очікуваною тарифною вартістю доставки та фактичною ефективною вартістю з урахуванням затримок, повторних спроб і втрат повідомлень. Показано, що PDG є кількісним індикатором операційної ефективності сервісу SMS у сучасних A2P-сценаріях.

4. Розроблено інтегровану модель IGDP (Integrated Gap-Delivery-Performance), яка поєднує показники QoS, QoE та PDG в єдину систему оцінювання ефективності. Запропонована модель дозволяє формалізувати взаємозв'язок між технічними параметрами мережі, сприйняттям користувачем якості сервісу та економічною доцільністю доставки повідомлень.

5. Побудовано математичні та регресійні моделі процесу доставки SMS, які враховують гетероскедастичність даних. Порівняльний аналіз підходів (OLS, кластерні методи, методи стиснення даних, інтервальні функції) показав, що використання адаптивних моделей дозволяє підвищити

точність оцінювання ефективності доставки та зменшити похибки прогнозування PDG в умовах змінного мережевого навантаження.

6. Досліджено вплив параметрів QoS та QoE на значення PDG та встановлено, що затримка доставки, коефіцієнт повторних спроб (retry-rate) та нерівномірність маршрутизації мають статистично значущий вплив на економічну ефективність SMS-сервісу, особливо у сегменті A2P.

7. Розроблено метод виявлення штучно згенерованого трафіку (AIT) у A2P-SMS на основі префіксно-часового аналізу та статистичної оцінки аномалій. Показано, що своєчасна ідентифікація AIT дозволяє зменшити викривлення показників ефективності та підвищити достовірність оцінювання PDG.

8. Запропоновано інтегрований фреймворк інтелектуального управління ефективністю доставки SMS, який включає категоризацію повідомлень, адаптивну маршрутизацію, контроль SLA та механізми RCS/OTT-fallback. Сценарне моделювання продемонструвало підвищення загальної ефективності доставки та зменшення негативних значень PDG у змішаному трафіку.

9. Результати сценарної оцінки підтвердили, що інтеграція технічного контролю QoS, економічного моніторингу PDG та механізмів виявлення AIT забезпечує зниження операційних втрат і підвищення стабільності функціонування SMS-сервісу в мережах 4G/5G.

10. Отримані результати мають теоретичне та практичне значення. Теоретично розширено підходи до оцінювання ефективності телекомунікаційних сервісів шляхом формалізації інтегральної техніко-економічної метрики. Практично розроблені моделі та алгоритми можуть бути використані операторами мобільного зв'язку та CPaaS-платформами для оптимізації маршрутизації SMS-трафіку, формування SLA-політик та зниження впливу шахрайського трафіку.

Таким чином, поставлена мета дисертаційної роботи досягнута, а сформульовані завдання повністю вирішені.

Sinch AB

CERTIFICATE OF IMPLEMENTATION

This is to certify that the research results obtained by Mykhailo Odarchenko within the dissertation study on improving the efficiency of SMS delivery in 4G/5G networks have practical significance and are recommended for implementation at Sinch AB.

In particular, the following research outcomes are subject to implementation in the company's messaging products and operational processes:

Adaptive routing method for SMS traffic, aimed at improving delivery efficiency by dynamically selecting routing paths based on message type, quality parameters, network conditions, and cost factors.

Enhanced Artificially Inflated Traffic (AIT) detection method, designed to improve the identification and prevention of fraudulent traffic patterns in messaging services and to reduce financial losses associated with artificial traffic generation.

The implementation of these results is expected to contribute to:

- improved delivery performance and routing efficiency;
- better traffic quality management;
- stronger fraud prevention capabilities;
- enhanced operational and economic effectiveness of messaging services.

This certificate is issued for submission as part of the dissertation documentation.

Sinch AB

Jonathan Campbell
VP Messaging Products

Date 18th March 2026

Signature 

