МІНІСТЕРСТВО ОСВІТИ і НАУКИ УКРАЇНИ

ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ЗВ'ЯЗКУ ім. О.С. ПОПОВА

А.Г. Ложковський

НОВІ МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА

Навчальний посібник

Для студентів закладів вищої освіти,
які навчаються за спеціальністю 172 „Телекомунікації та радіотехніка”

Одеса 2018

План НМВ 2018 р.

УДК 621.39

ББК 32.81

Л71

Рецензент – Лісовий І. П.

Ложковський А.Г.

Л71 Нові методи теорії телетрафіка /

Ложковський А.Г. – Одеса: ОНАЗ ім. О.С. Попова, 2018. – 80 с.: іл.

Викладено основні положення та методи аналізу теорії телетрафіка, на яких базуються процедури проектування телекомунікаційних систем і мереж. Розглянуто математичні моделі систем розподілу інформації з втратами, з чергою та з пріоритетами. Надано методи дослідження цих систем в умовах ідеалізованої моделі пуассонівського потоку та реальних непуассонівських потоків вимог мультисервісних мереж зв'язку з пакетною технологією передавання та комутації інформації.

Навчальний посібник призначено для студентів та аспірантів, які навчаються за спеціальністю 172 «Телекомунікації та радіотехніка», а також для фахівців, що займаються практичним застосуванням теорії телетрафіка.

Рекомендовано
до видання Вченою радою
ОНАЗ ім. О.С. Попова
Протокол № 10
від "27" червня 2018 р.

ББК 32.81

ЗМІСТ

ПЕРЕДМОВА 4
1 ОСНОВИ ТЕОРІЇ ТЕЛЕТРАФІКА 5
1.1 Задачі та методи теорії телетрафіка 5
1.2 Моделі систем розподілу інформації 6
1.3 Математична модель потоку вимог 7
1.4 Навантаження та його види 9
1.5 Характеристики якості обслуговування 11
1.5.1 Системи з втратами 11
1.5.2 Системи з чергами 11
1.5.3 Комбіновані системи (з чергами та втратами) 12
1.5.4 Пріоритетні системи 13
1.5.5 Пропускна здатність 13
2 КЛАСИЧНІ МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА 14
2.1 Модель пуассонівського потоку вимог 14
2.2 Система з втратами $M/M/m$, $M/G/m$ 15
2.3 Система з необмеженою чергою $M/M/m/\infty$ 17
2.4 Система з обмеженою чергою $M/M/m/r$ 20
2.5 Система з необмеженою чергою $M/G/1/\infty$ 23
2.6 Система з пріоритетами $M/G/1/\infty$ 25
2.6.1 Відносний пріоритет 25
2.6.2 Абсолютний пріоритет 26
3 МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА ДЛЯ МУЛЬТИСЕРВІСНИХ МЕРЕЖ ЗВ'ЯЗКУ 27
3.1 Модель трафіка мультисервісних мереж зв'язку 27
3.2 Апроксимація Хейворда 28
3.3 Функція розподілу станів системи з втратами $H/D/m$ 30
3.4 Імовірність втрат системи $H/D/m$ 34
3.5 Система з необмеженою чергою $H/D/m/\infty$ 37
3.6 Система з необмеженою чергою $GI/M/1/\infty$ 40
4 МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА ДЛЯ ПАКЕТНИХ МЕРЕЖ ЗВ'ЯЗКУ 46
4.1 Модель трафіка пакетних мереж зв'язку 46
4.2 Система з необмеженою чергою $fBM/D/1/\infty$ 51
4.3 Показник Херста трафіка з розподілами Парето та Вейбулла 56
4.4 Апроксимація станів системи $fBM/D/1/\infty$ 60
4.5 Розрахунок показника Херста методом R/S-аналізу 66
Контрольні запитання 74
Список літератури 75
Список позначень та скорочень 76
Предметний покажчик 77

ПЕРЕДМОВА

Головним змістом теорії телетрафіка є дослідження пропускної здатності телекомунікаційних систем та розробка нових науково обґрунтованих методів оцінки характеристик якості обслуговування. В математичних моделях теорії враховано вид вхідного потоку, схему системи та дисципліну обслуговування.

У теорії телетрафіка розроблено низку математичних моделей і методів розв'язання задач аналізу і синтезу телекомунікаційних систем для умов пуассонівської моделі трафіка. Проте набір цих методів є недостатньо повним з погляду структурних ознак реальних систем, дисциплін обслуговування й особливо характеру трафіка. Реальному трафіку мультисервісних мереж властивий значно більший рівень нерівномірності інтенсивності навантаження, ніж це передбачено класичною моделлю пуассонівського потоку. Для таких моделей трафіка ще не має відповідних методів розрахунку і на практиці оцінка характеристик якості обслуговування мультисервісних мереж зв'язку виконується наближеними методами та засобами імітаційного моделювання.

У першому розділі стисло викладено ключові положення та визначення теорії телетрафіка. Надано ключові формули розрахунку параметрів потоку вимог та створюваного ним навантаження, основні та допоміжні характеристики якості обслуговування для систем з втратами та чергами.

У другому розділі у стилі довідника надано головні результати теорії телетрафіка, які отримано методами Ерланга, що стали вже класичними для пуассонівської моделі трафіка, а також формули Поллачека-Хінчина.

У третьому розділі розглянуто модель мультисервісного трафіка, яка властива мережам зв'язку з технологією комутації каналів. Тут нерівномірність трафіка описується дисперсією інтенсивності навантаження, яка перевищує саму інтенсивність у межах від 2 до 15 разів. Математичною моделлю трафіка є гіперекспонентний розподіл інтервалу часу між вимогами потоку. Для такої моделі трафіка викладено нові методи визначення основних характеристик якості обслуговування у системах з втратами та нескінченною чергою.

У четвертому розділі розглянуто модель мультисервісного трафіка, яка властива мережам зв'язку з технологією комутації пакетів. Тут трафіку притаманна ще більша нерівномірність, за якої дисперсія інтенсивності навантаження перевищує саму інтенсивність від 15 і більше разів. Трафіку пакетних мереж зв'язку властиве так зване «пачкування», для якого методи гіперекспонентного потоку не застосовні. Вважається, що трафік – це самоподібний процес з певними значеннями показника Херста. Для моделей самоподібного трафіка з розподілами Парето та Вейбулла дано нові формули визначення показника Херста, який залежить від параметра форми цих розподілів. На основі апроксимації станів системи з самоподібним трафіком дано нові формули визначення імовірності очікування початку обслуговування. Для розрахунку показника Херста методом R/S-статистика надана спрощена методика.

1 ОСНОВИ ТЕОРІЇ ТЕЛЕТРАФІКА

1.1 Задачі та методи теорії телетрафіка

Математичні моделі телекомунікаційних систем ґрунтуються на основі теорії *систем масового обслуговування* (СМО). СМО обслуговують вимоги, що надходять до системи через випадкові інтервали часу, причому тривалість обслуговування також може бути випадковою. Методами теорії СМО досліджується вплив випадкових факторів на процеси функціонування системи.

Одним із класів СМО є *системи розподілу інформації* (СРІ). В них розподільчою мережею є телекомунікаційна мережа, вузли якої забезпечують з'єднання каналів передавання інформації. У вузлах за певними алгоритмами обслуговуються повідомлення телекомунікаційних служб мережі. Обслуговування повідомлення ототожнюється з *вимогою* на його передачу або обробку (виклики телефонної станції, пакети пакетного комутатора тощо).

Теорією телетрафіка досліджується кількісна сторона процесів обслуговування потоків вимог (трафіка) у СРІ. Предметом теорії телетрафіка є методи оцінки характеристик якості обслуговування (QoS). Задача теорії телетрафіка полягає у встановленні залежності між характером потоку вимог та пропускною здатністю системи з метою визначення параметрів схеми та найкращих шляхів керування процесами обслуговування трафіка.

Теорія телетрафіка – це набір імовірнісних методів аналізу, синтезу та оптимізації СРІ, а також проектування та експлуатації мереж зв’язку:

– *задача аналізу* – встановлення характеристик QoS і їх залежності від параметрів потоку вимог, схеми системи і дисципліни обслуговування;

– *задача синтезу* – визначення структурних параметрів мережі або системи (схеми), при заданих потоках, дисципліні і якості обслуговування. Задача синтезу певною мірою є зворотною до задачі аналізу;

– *задача оптимізації* є близькою до задач аналізу і синтезу та може бути комбінацією їх повторювання до досягнення мети за заданим критерієм. Іноді задача оптимізації подається як задача керування потоками вимог або структурою мережі для отримання найкращих показників QoS.

Основними методами розв’язання задач у теорії телетрафіка є аналітичний, числовий та метод статистичного моделювання:

– *аналітичні методи* дозволяють розв’язувати задачі теорії телетрафіка в тих випадках, коли структура системи, характеристики потоку і дисципліна обслуговування відносно прості;

– *числові методи* використовують спеціальні алгоритми, що дозволяють знаходити наближені рішення ітераційними або іншими методами. Вони застосовуються для складних систем, де кількість станів настільки велика, що розв'язати систему рівнянь статистичної рівноваги неможливо;

– *методи статистичного моделювання* є найбільш універсальними методами для розв’язання задач практично будь-якої складності. Метод полягає в побудові математичної моделі системи, реалізація якої здійснюється у виді програми імітаційного моделювання.

1.2 Моделі систем розподілу інформації

Теорія телетрафіка оперує не із самими системами розподілу інформації, а з їх математичними моделями. Для повного опису СРІ достатньо вказати імовірнісні процеси, що описують вхідний потік вимог, структуру системи та дисципліну обслуговування. Математична модель СРІ містить такі елементи:

1. *Вхідний потік вимог* **(***трафік***)** – класифікується за ознаками стаціонарності, ординарності та післядії. Основними параметрами потоку вимог є його інтенсивність та дисперсія інтенсивності.

2. *Схема системи розподілу інформації* – це інформація про кількість обслуговуючих каналів або серверів (пристроїв, що надають послуги) і черг та їх взаємне з'єднання і доступність для вхідних вимог.

3. *Дисципліна обслуговування потоку вимог* – характеризує взаємодію потоку вимог з СРІ та визначає спосіб (правила) обслуговування та долю вимог при їх надходженні до системи. Найпоширеніші такі способи обслуговування: з втратами, з чергами, комбіновані (з втратами та чергами), з пріоритетами.

Дисципліна (правило) обслуговування потоку вимог визначає назву або тип СРІ і тому розрізнюють такі СРІ:

1. *Системи з втратами* – вимоги, які при надходженні до системи не знаходять в ній вільного сервера, отримують відмову у сервісі та втрачаються.

2. *Системи з чергами* – вимоги, які не обслужені відразу через зайнятість усіх серверів системи, стають у чергу, і за допомогою деякої дисципліни обслуговування черги визначається, у якому порядку очікувані вимоги вибираються із черги для обслуговування за певним правилом (наприклад, *FIFO* – вимоги з черги обслуговуються в порядку їхнього надходження або *LIFO* – перевагу має вимога, що надійшла до черги останньою).

3. *Комбіновані системи* з чергами та втратами (системи з чергою при обмеженнях). Наприклад, очікувати може тільки кінцева кількість вимог, обумовлена кількістю місць очікування, меншою за нескінченність. Або вимога втрачається якщо час очікування у черзі або перебування у системі перевищує задані межі.

4. *Пріоритетні системи* – для вимог передбачено різні пріоритети в обслуговуванні. Якщо вимога, що надійшла, має високий пріоритет, а всі сервери зайняті, то вона або займає одне з перших місць у черзі, або тимчасово припиняє обслуговування вимоги низького пріоритету і займає її місце у сервері. При цьому можуть бути застосовані такі пріоритетні правила:

- абсолютний пріоритет (*pre-emptive discipline*) – вимога високого пріоритету перериває обслуговування вимоги низького пріоритету;
- відносний пріоритет (*head of the line priority discipline*) – вимога високого пріоритету посідає перше місце у черзі і переривань немає.

Структурні характеристики системи (схема) частково зумовлюють дисципліну обслуговування потоку вимог. Наприклад, за кількості місць очікування $r = 0$ буде система з втратами, при $0 < r < \infty$ – комбінована система з чергою та з втратами, а при $r = \infty$ – чиста система з чергою.

Для стислого запису Д. Кендаллом запропоновано спеціальне *умовне позначення базової моделі*, в якому літерами позначаються певні параметри математичної моделі СРІ: ***A*** **/** ***B*** **/** ***m*** **/** ***r***, а саме: *A* – вид імовірнісної функції розподілу інтервалів часу між сусідніми вимогами потоку; *B* – вид функції розподілу тривалості обслуговування; *m* – кількість обслуговуючих серверів та *r* (необов'язково) – кількість місць очікування у системі. Тобто *m* та *r* характеризують схему і опосередковано дисципліну обслуговування.

Для опису інтервалу часу між послідовними вимогами або тривалості обслуговування найбільш часто використовуються такі розподіли: *M* – експонентний (*M* – *Марковська* модель); *H* – гіперекспонентний (*Hyper-exponential*); *D* – детермінований (*Determined*); *U* – рівномірний (*Uniform*); *E* – розподіл Ерланга; *G* – довільний або узагальнений (*General*), *GI* – узагальнений рекурентний (*General Independent*). Будь-яка із цих та інших літер може бути застосована замість літер *A* і *B* у моделі Кендалла.

1.3 Математична модель потоку вимог

Сукупність (послідовність) подій надходження до системи в моменти $t_1 \ldots t_n$ вимог на обслуговування утворюють *потік вимог* (трафік). У *рекурентного потоку вимог* інтервали часу між подіями надходження вимог $z_n = t_n - t_{n-1}$ є стохастично незалежними та випадковими величинами. За незмінного інтервалу часу між вимогами z_n потік є *детермінованим*, однак в телекомунікаціях в основному потоки є *випадковими*.

Випадковий процес надходження у систему вимог характеризується законом розподілу, що встановлює зв'язок між значенням випадкової величини та імовірністю появи цього значення. Такий потік може бути описаний ймовірнісною функцією розподілу інтервалів часу між сусідніми вимогами або ймовірнісною функцією розподілу кількості вимог за умовну одиницю часу.

Математична модель випадкового потоку вимог може бути представленою за допомогою таких імовірнісних функцій:

1. Функція розподілу ймовірностей $P(z)$ інтервалів часу z між вимогами.

Основним параметром моделі є середнє значення інтервалів z, що для випадкової величини є математичним сподіванням $\bar{z}$. Параметр, зворотний до математичного сподівання $\bar{z}$, є *інтенсивністю потоку надходження* вимог λ за одиницю часу, в яких вимірюється $\bar{z}$:

$$\lambda = \frac{1}{\bar{z}}. \qquad (1.1)$$

У деяких моделях можуть використовуватись й додаткові параметри, такі як коефіцієнт варіації інтервалу z, показник самоподібності трафіка тощо.

Сукупність моментів завершення обслуговування вимог утворюють *потік звільнення* серверів системи з *інтенсивністю обслуговування вимог* μ:

$$\mu = \frac{1}{\bar{x}}, \qquad (1.2)$$

де $\bar{x}$ – середній час обслуговування однієї вимоги.

2. Функція розподілу ймовірностей $P_i(t)$ кількості вимог i за умовну одиницю часу t.

Основним параметром моделі є середнє значення випадкової величини i, що визначається за формулою Літтла як $\bar{i} = \lambda t$. В деяких розподілах можуть використовуватись й додаткові параметри, такі як дисперсія інтенсивності навантаження або моменти більш високих порядків.

Таким чином, математична модель потоку вимог відображається у два способи за допомогою імовірнісних функцій розподілу:

– інтервалів часу між сусідніми вимогами z;
– кількості вимог i за умовну одиницю часу t.

У першому випадку, як правило, застосовуються неперервні закони, а другому – дискретні.

Розвиток телекомунікаційних технологій, нові принципи побудови мереж зв'язку, зміна структурного складу абонентів і спектра надаваних послуг – все це збільшує нерівномірність інтенсивності потоків вимог, яка вимірюється дисперсією інтенсивності. Статистичні виміри параметрів трафіка з точки зору нерівномірності трафіка свідчать про три його типи, які надано в табл. 1.1

Таблиця 1.1 – Види трафіка телекомунікаційних мереж

Назва моделі трафіка	**Вид розподілу**	**Характеристики трафіка**	**Коефіцієнт скупченості навантаження**
Пуассонівський (моносервісний)	Інтервал часу між заявками: Z – експонентна ф-я Кількість вимог за одиницю часу: C – з-н Пуассона	Λ – інтенсивність навантаження	$S = \frac{D}{\Lambda} = 1$
Мультисервісний	Інтервал часу між заявками: Z – гіперекспонентна ф-я Кількість вимог за одиницю часу: C – з-н Гауса	Λ – інтенсивність навантаження, D – дисперсія інтенсивності навантаження	$S = 2…15$
Пакетний	Інтервал часу між заявками: Z – ф-я Парето або Вейбулла Кількість вимог за одиницю часу: C – з-н, схожий до з-ну Гауса	Λ – інтенсивність навантаження, k – коефіцієнт пачковості H – показник самоподібності Херста	$S > 15$ $H = 0{,}5…1$

Від адекватності обраної математичної моделі потоку вимог суттєво залежить точність розрахунку характеристик якості обслуговування *QoS*.

1.4 Навантаження та його види

Сумарний час обслуговування всіх вхідних вимог є *навантаженням* для серверів (ліній, каналів) СРІ. Чим більший цей час, тим більше навантаження «обслуговують» сервери СРІ. В теорії телетрафіка розрізняють вхідне (*traffic offered*), обслужене (*traffic carried*) та надлишкове (*overflow traffic*) навантаження. Надлишкове навантаження – це різниця між вхідним і обслуженим, що у системі з втратами є втраченим навантаженням.

Навантаження вимірюється в годино-заняттях. Параметр «навантаження» не дає чітких відомостей щодо напруженості роботи системи, та за який час воно виконане. Тому введено поняття *інтенсивності обслуженого навантаження Y*, яка визначається як приведений час навантаження. Він розраховується як сумарний час обслуговування всіх вимог x_i на деякому інтервалі часу $t_2 - t_1$, поділений на величину цього інтервалу ($t_2 - t_1$ може дорівнювати, наприклад, одній годині):

$$Y = \frac{\sum x_i}{t_2 - t_1}. \qquad (1.3)$$

Кількість серверів системи, що має бути залученою до обслуговування вимог вхідного потоку, теж є *навантаженням* для системи. Кожної миті кількість зайнятих серверів системи або кількість одночасно обслуговуваних вимог *C* визначає *миттєве значення* обслуговуваного навантаження. Через те що ця кількість є випадковою, то основною характеристикою обслуговуваного навантаження є її середнє значення, що так само, як і (1.3), називається *інтенсивністю обслуженого навантаження*

$$Y = \frac{1}{n}\sum_{i=1}^{n} C_i\,, \qquad (1.4)$$

яка однозначно збігається зі значенням, розрахованим за формулою (1.3).

Отже, *інтенсивність обслуженого навантаження* – це приведений час сумарного обслуговування всіх вимог або середнє значення кількості зайнятих серверів, яку ще називають *завантаженістю системи*.

За одиницю виміру інтенсивності навантаження прийнято 1 *Ерланг* (1 Ерл), яку названо на честь засновника теорії телетрафіка А.К. Ерланга.

Навантаження на сервери системи є результатом спільного процесу надходження та обслуговування вимог. Оскільки вимоги надходять до системи через випадкові інтервали часу (кількість вимог за умовну одиницю часу випадкова) і тривалість обслуговування також може бути випадковою, то й *навантаження є випадковою величиною*. Тому *інтенсивність вхідного навантаження* Λ нормується середнім часом обслуговування вимог $\bar{x}$. *Інтенсивність вхідного навантаження* Λ, вимірювана в *Ерлангах*, це середня кількість вимог, що надходить до системи за середній час обслуговування однієї вимоги:

$$\Lambda = \lambda \bar{x}\,. \qquad (1.5)$$

Тобто Λ – це є інтенсивність потоку надходження вимог λ, яка віднесена до середньої тривалості обслуговування $\bar{x}$.

Іноді використовується поняття *інтенсивність питомого навантаження* (навантаження на один сервер із m серверів):

$$\rho = \frac{\Lambda}{m}. \tag{1.6}$$

Інтенсивність Λ – це середня кількість вимог, що надходить до системи за середній час обслуговування вимог, а інтенсивність Y – це середня кількість зайнятих серверів. Дані інтенсивності є оцінками навантажень, які є випадковими величинами, і їх оцінюють математичним сподіванням.

Середню кількість вимог, що надходить до системи за середній час обслуговування вимоги або математичне сподівання середньої кількості вимог з функції розподілу $P_i(t)$ можна розрахувати так:

$$\Lambda = \sum_{i=0}^{\infty} iP_i , \tag{1.7}$$

де P_i – імовірність того, що за інтервал часу $t = \bar{x}$ до системи надійде точно i вимог, де $i = 0, \ldots, \infty$.

Середню кількість зайнятих серверів системи з функції розподілу її станів можна розрахувати так:

$$Y = \sum_{j=0}^{m} jp_j , \tag{1.8}$$

де P_j – імовірність того, що в довільний момент часу у системі з m серверів зайнято точно j серверів, де $j = 0, \ldots, m$.

Крім інтенсивності навантаження (математичного сподівання) важливою характеристикою випадкової величини навантаження є дисперсія інтенсивності, яка для вхідного та обслуговуваного навантаження відповідно визначиться так:

$$D_\Lambda = \sum_{i=0}^{\infty} (i - \Lambda)^2 P_i ; \qquad D_Y = \sum_{j=0}^{m} (j - Y)^2 P_j .$$

Якщо потік вимог є експонентним, то створюване ним навантаження, як випадкова величина, має розподіл Пуассона. При цьому характерний збіг дисперсії навантаження D_Λ з її математичним сподіванням Λ. Навантаження називається *пуассонівським*, і воно вважається мало нерівномірним.

Якщо $D_\Lambda < \Lambda$, то навантаження є згладженим, оскільки його відхилення від середнього значення будуть менші, ніж для пуассонівського навантаження.

Навантаження, у якого $D_\Lambda > \Lambda$ є надто нерівномірним, де на одних інтервалах часу кількість вхідних вимог мала, а на інших – їх кількість набагато більше. тобто вимоги «скупчуються».

Скупченість вхідного навантаження (пікфактор трафіка) S визначається як відношення дисперсії навантаження D_Λ до її математичного сподівання Λ:

$$S = \frac{D_\Lambda}{\Lambda}. \tag{1.9}$$

Коефіцієнт скупченості навантаження $S = 1$ для пуассонівського навантаження (мало нерівномірного), $S < 1$ для вирівняного (згладженого) навантаження та $S > 1$ для скупченого (надто нерівномірного) навантаження.

1.5 Характеристики якості обслуговування

В теорії телетрафіка якість обслуговування (*QoS – Quality of Service*) потоку вимог характеризується можливістю негайного обслуговування вимоги або очікування початку обслуговування. Ці можливості визначаються обраною дисципліною обслуговування вимог і для кожної з них властивий певний набір основних і допоміжних характеристик *QoS*.

1.5.1 Системи з втратами

У системі з втратами вимозі, яка надходить в момент відсутності вільних серверів, відмовляється в обслуговуванні і вона втрачається. Основною кількісною оцінкою якості обслуговування є імовірність втрати вимоги P_B (*B – blocking*, втрати). *Імовірність втрати вимоги* P_B визначається як відношення кількості втрачених за відрізок часу $(t_1,\ t_2)$ вимог C_B до загальної кількості вимог, що надійшли за той самий час до системи C:

$$P_B = \frac{C_B}{C}. \tag{1.10}$$

Допоміжними характеристиками *QoS* в цій системі є такі:

Середній час перебування вимоги у системі T збігається із середнім часом обслуговування вимоги $\bar{x}$, тобто $T = \bar{x}$.

Середня кількість вимог у системі N характеризує ступінь завантаженості системи і збігається з середньою кількістю зайнятих серверів, що є інтенсивністю обслуговуваного навантаження Y, тобто $N = Y$.

За формулою Літтла при інтенсивності навантаження Λ за T одиниць часу перебування вимоги у системі до неї надійде в середньому ΛT вимог:

$$N = \Lambda T. \tag{1.11}$$

Для систем з втратами інтенсивність обслуженого навантаження менша вхідного навантаження на величину ΛP_B, тобто

$$Y = \Lambda(1 - P_B). \tag{1.12}$$

Величина ΛP_B є інтенсивністю надлишкового навантаження і його потік за своєю структурою суттєво відрізняється від потоку вхідного навантаження більш нерівномірним характером.

1.5.2 Системи з чергами

Для кількісної оцінки якості обслуговування систем з чергою передбачено такі основні характеристики:

- імовірність очікування P_w (або середня частка затриманих вимог);
- середня довжина черги Q (або середня кількість вимог у черзі);
- середня тривалість очікування для затриманих вимог t_q;
- середня тривалість очікування для будь-якої вимоги W.

Імовірність очікування P_w визначається як відношення кількості вимог, що потрапили за відрізок часу $(t_1,\ t_2)$ у чергу C_Q до загальної кількості вимог, що надійшли за той самий час до системи C:

$$P_w = \frac{C_Q}{C}. \quad (1.13)$$

Середній час очікування у черзі t_q утворюється за рахунок затримки вимог саме у черзі. *Середній час очікування у системі W* являє собою середнє значення часу очікування, віднесене до всіх вимог – затриманих і незатриманих. Час очікування є випадковою величиною і тому визначається за правилом розрахунку математичного сподівання.

Середня довжина черги Q визначається середньою кількістю вимог, що очікують обслуговування. За формулою Літтла при інтенсивності навантаження Λ за W одиниць часу очікування до черги надійде в середньому ΛW вимог:

$$Q = \Lambda W . \quad (1.14)$$

Допоміжними характеристиками *QoS* є середня кількість вимог у системі N та середній час перебування вимоги у системі T.

Середня кількість вимог у системі N визначає ступінь завантаженості системи і за необмеженої черги складається з середньої кількості вимог, що надходять до системи Λ, та тих, що очікують у черзі Q:

$$N = \Lambda + Q . \quad (1.15)$$

Середній час перебування вимоги у системі T – це час, проведений однією вимогою у системі й усереднений за всіма вимогами (затриманим і незатриманим). Він складається із середнього часу обслуговування $\bar{x}$ і середнього часу очікування вимог у системі W:

$$T = \bar{x} + W . \quad (1.16)$$

Інтенсивність обслуженого навантаження Y менша за Λ і дорівнює:

$$Y = \Lambda(1 - P_w) . \quad (1.17)$$

Для кожної моделі потоку всі характеристики якості обслуговування перебувають у певній функціональній залежності.

1.5.3 Комбіновані системи (з чергами та втратами)

Система з чергою, в якій введено обмеження на максимальну кількість вимог, що можуть перебувати у черзі, або на максимальний час очікування початку обслуговування є системою з комбінованою дисципліною обслуговування. За обмеженої кількості місць очікування (максимальна довжина черги) у випадку надходження вимоги в момент, коли всі сервери і місця очікування зайняті попередніми вимогами, дана вимога втрачається. За обмеженого часу очікування, якщо вимога перебуває у черзі понад допустимий час, то їй відмовляється в обслуговуванні, і вона теж втрачається. Тому окрім характеристик *QoS* системи з чергами без обмежень розраховуються й такі:

- імовірність втрати вимоги P_B (через обмежену довжину черги);
- імовірність очікування понад припустимий час $P_{W>t}$.

Імовірність $P_{W>t}$ залежить від дисципліни обслуговування черги і найбільш простим для розрахунку є випадок упорядкованої черги *FIFO*. Цю імовірність ще називають *умовними втратами*, оскільки вимога, що очікує понад припустимий час t, може втратити свою актуальність для користувача.

1.5.4 Пріоритетні системи

Сучасні телекомунікаційні системи і мережі характеризуються наявністю пріоритетного обслуговування переданих і оброблюваних даних.

Комбіновані системи з обмеженнями на довжину черги та час очікування є найбільш поширеними в телекомунікаціях. В умовах різнорідного трафіка (мова, відео, дані) цю систему доповнюють механізмом пріоритетів, в якому всі вимоги поділяють на категорії і вимоги більш високої категорії при обслуговуванні мають певні переваги (пріоритети) перед вимогами більш низької категорії.

Для кількісної оцінки якості обслуговування систем із пріоритетами розраховуються такі самі характеристики, як і для системи з чергами, але для кожного з введених пріоритетів окремо. Наприклад, середній час очікування у черзі вимоги *k*-го пріоритету, середня кількість вимог у системі *k*-го пріоритету тощо.

1.5.5 Пропускна здатність

У СРІ якість обслуговування суттєво впливає на пропускну здатність і продуктивність системи. Критеріями якості обслуговування для систем із втратами є імовірність втрати вимоги, а для систем з чергами – імовірність очікування. Чим більше припустима норма втрат, тим менша якість обслуговування.

Пропускна здатність – це максимальна інтенсивність обслуженого навантаження при забезпеченні заданої якості обслуговування.

Інтенсивність обслуженого навантаження – це середня кількість зайнятих серверів і тому пропускна здатність системи не дорівнює загальній кількості серверів у ній. Через випадковість вхідних потоків навантаження середня кількість зайнятих серверів не досягає загальної їх кількості у системі. Задачею є лише наближення середньої кількості зайнятих серверів до загальної кількості серверів системи. Середня кількість зайнятих серверів у системі з втратами завжди більша, чим у системі з чергою за однакових Λ та m, отже й пропускна здатність системи з втратами вища.

Продуктивність – це гранична, статистично усереднена кількість вимог, які будуть обслужені системою за одиницю часу за заданої якості обслуговування. Ця характеристика використовується, як правило, для оцінки систем керування та керуючих пристроїв.

Пропускна здатність та продуктивність СРІ залежать не тільки від норми імовірності втрат, але й від структури системи (кількості серверів і схеми їх включення), дисципліни обслуговування і закону розподілу тривалості обслуговування. Крім того, на пропускну здатність та якість обслуговування суттєво впливає вид потоку вимог або його математична модель.

2 КЛАСИЧНІ МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА

2.1 Модель пуассонівського потоку вимог

Моделлю потоку у телефонних мережах зв'язку є *експонентний розподіл інтервалів часу між вимогами z* (викликами АТС) з інтенсивністю λ. З густини (щільності) цього розподілу розраховується імовірність будь-якої тривалості z_x випадкової величини *z* за заданої інтенсивності надходження вимог λ:

$$P(z) = \lambda e^{-\lambda z}. \qquad (2.1)$$

Середнє значення випадкової величини *z*, розподіленої за експонентним законом (2.1), дорівнює λ^{-1}, а параметр розподілу λ – це середня кількість вимог за одиницю часу, в яких вимірюємо $\bar{z}$. Коефіцієнт варіації $v_z = \sigma_z / \bar{z} \equiv 1$.

Якщо інтервал часу *z* розподілений за експонентним законом, то кількість подій *i* за умовну одиницю часу *t* обов'язково розподілена за законом Пуассона:

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}. \qquad (2.2)$$

Величина λ*t* є параметром розподілу Пуассона. За цим розподілом можна розрахувати імовірність надходження до системи точно *i* вимог за умовну одиницю часу тривалістю *t* за заданої інтенсивності надходження вимог λ. Якщо за одиницю часу *t* взято середній час обслуговування вимоги $\bar{x}$, то параметром розподілу (2.2) буде інтенсивність вхідного навантаження $\Lambda = \lambda \bar{x}$, причому дисперсія інтенсивності навантаження $\sigma^2 \equiv \Lambda$.

Отже, і розподіл Пуассона також є моделлю трафіка у телефонних мережах зв'язку, який часто називається *пуассонівським потоком*.

Середнє значення випадкової величини *i*, розподіленої за законом (2.2), визначається як $\bar{i} = \lambda t$. При зростанні *i* від нуля до $\bar{i}$ імовірність $P_i(t)$ поступово зростає, для значення $\bar{i}$ вона максимальна, а далі зі зростанням *i* ця імовірність поступово зменшується (при $\lambda t > 50$ розподіл майже симетричний).

Пуассонівський потік вимог має такі властивості:

- *стаціонарність* – імовірність кількості вимог на відрізку часу залежить тільки від тривалості цього відрізка і не залежить від того, де саме на осі часу він розміщений;
- *ординарність* – імовірність надходження >1 вимоги за нескінченно малий відрізок часу є нескінченно малою порівняно з імовірністю надходження точно однієї вимоги (вимоги надходять лише по одній);
- *без післядії* – для будь-яких відрізків часу, що не перетинаються, кількість вимог одного відрізка не залежить від того, скільки вимог надійшло на інший.

Вважається, що пуассонівський потік утворюється нескінченною кількістю джерел навантаження і для нього інтенсивність потоку λ, тобто середня кількість вимог за одиницю часу, є величина незмінна.

Для пуассонівської моделі потоку відомі всі аналітичні формули розрахунку основних характеристик *QoS* будь-яких СРІ [1–6].

2.2 Система з втратами *M*/*M*/*m*, *M*/*G*/*m*

Аналітичні розрахунки характеристик якості обслуговування (*QoS*) основані на математичному описі реакції системи на зовнішні впливи. Реакцією системи вважається її стан, а зовнішніми впливами – потоки вимог.

Повнодоступна схема СРІ з *m* серверами обслуговує пуассонівський потік вимог за дисципліною із втратами. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні розподіли. *Інтенсивність надходження вимог* λ, *інтенсивність обслуговування вимог* μ. Необхідно визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0 \ldots m$) та характеристики *QoS*.

Стан системи – це поточна кількість вимог системи, що обслуговуються у серверах. Закон розподілу станів системи, на відміну від середньої кількості вимог в ній (завантаженості), більш повно характеризує функціонування СРІ під впливом випадкових факторів.

Дискретні стани системи S_k змінюються за кожної події надходження вимоги або закінченні її обслуговування. Для пуассонівського потоку інтенсивність надходження вимог λ є величина незмінна. Якщо система знаходиться у стані S_k і надходить вимога, то система переходить з однаковою інтенсивністю λ у стан S_{k+1} за будь-якого *k*. В разі закінчення обслуговування вимоги інтенсивність, з якою система переходить зі стану S_k у стан S_{k-1} залежить від кількості зайнятих серверів або поточного стану системи.

Стаціонарний розподіл імовірностей станів системи P_k для дискретного ланцюга Маркова визначається системою лінійних алгебраїчних рівнянь:

$$P_k = \sum_i P_i p_{i,k}\,, \tag{2.3}$$

де $p_{i,k}$ – імовірність переходу системи зі стану *i* у стан *k* [5]. При цьому розподіл (2.3) має задовольняти умові нормування, за якої $\sum_{k=0}^{m} P_k = 1$.

Через ординарність потоку рівняння (2.3) спроститься і математичною моделлю процесу обслуговування вимог стане рівняння:

$$(\lambda + \mu k)P_k = \lambda P_{k-1} + \mu(k+1)P_{k+1}\,. \tag{2.4}$$

Система рівнянь (2.4) є рівнянням рівноваги або балансу. Розв'язання рівняння балансу дає співвідношення між імовірностями суміжних станів системи $P_k = \dfrac{\lambda}{\mu k} P_{k-1}$ та будь-якого стану відносно нульового стану:

$$P_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 \qquad (k = 1, 2, \ldots m). \tag{2.5}$$

З урахуванням умови нормування, внесеної до стаціонарного розподілу ймовірностей станів системи (2.3), імовірність P_0 визначиться так:

$$P_0 = \left[\sum_{i=0}^{m} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}\right]^{-1}. \tag{2.6}$$

Розподіл (2.5) є стаціонарним розподілом імовірностей станів повнодоступної системи з втратами, яка обслуговує пуассонівський потік вимог. Дана задача розв'язана А.К. Ерлангом у припущенні про експонентний розподіл тривалості обслуговування, але згодом доведено, що (2.5) вірно за будь-якого довільного закону, тобто й для моделі $M/G/m$.

Щоб отримані імовірності станів системи не залежали від того, у якому стані система була у початковий момент часу, даний процес має бути *ергодичним*. Достатнім критерієм ергодичності марковського процесу є умова:

$$\Lambda = \frac{\lambda}{\mu} < m, \tag{2.7}$$

тобто у середньому має надходити менше вимог, ніж їх можна обслужити.

З (2.7) випливає, що в (2.5) параметром розподілу ймовірностей станів системи є інтенсивність навантаження Λ, а тому з (2.5) і (2.6) маємо:

$$P_k = \frac{\dfrac{\Lambda^k}{k!}}{\sum\limits_{i=0}^{m} \dfrac{\Lambda^i}{i!}}, \quad (k = 1, 2, \ldots m). \tag{2.8}$$

Розподіл (2.8) називається *першим розподілом Ерланга*, за яким можна розрахувати імовірність зайнятості k серверів повнодоступної m-серверної системи з втратами при обслуговуванні пуассонівського потоку вимог.

Для стану $k = m$ з (2.8) отримуємо *B-формулу Ерланга*, що позначається $E_m(\Lambda)$, і за якою можна розрахувати основну характеристику *QoS* – імовірність втрати вимоги P_B, яка дорівнює розрахунку за формулою (1.10):

$$P_m = P_B = E_m(\Lambda) = \frac{\dfrac{\Lambda^m}{m!}}{\sum\limits_{i=0}^{m} \dfrac{\Lambda^i}{i!}}, \qquad (k = 1, 2, \ldots m). \tag{2.9}$$

Розраховані значення $E_m(\Lambda)$ часто надаються в таблицях Ерланга.

З (2.9) маємо значення імовірності зайняття всіх m серверів системи. Чергова вимога, що надходить до системи, буде втрачена (заблокована) в тому випадку, коли всі сервери зайняті. Але тільки за умов пуассонівського потоку значення імовірності зайняття всіх серверів системи P_m збігається з імовірністю втрати вимоги P_B незалежно від того, надходить в цей момент вимога чи ні.

Для односерверної системи $M/G/1$ із (2.8) за умови $m = 1$ отримуємо:

– імовірність того, що єдиний сервер вільний: $P_0 = \dfrac{1}{1+\Lambda}$;

– імовірність того, що сервер зайнятий P_1:

$$P_1 = 1 - P_0 = P_B = E_m(\Lambda) = Y = N = \frac{\Lambda}{1+\Lambda}. \tag{2.10}$$

Імовірність P_1 дорівнює імовірності втрати вимоги P_B і збігається з інтенсивністю обслуженого навантаження Y, а також із середньою кількістю вимог у системі N.

2.3 Система з необмеженою чергою *M*/*M*/*m*/∞

Повнодоступна схема СРІ з m серверами та необмеженою кількістю місць очікування у черзі обслуговує пуассонівський потік вимог з дисципліною обслуговування черги *FIFO*. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні закони розподілу. Інтенсивність потоку вимог λ, інтенсивність обслуговування вимог μ. Визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0 \ldots \infty$) та характеристики *QoS*.

Стан системи – це поточна кількість вимог системи, що обслуговуються у серверах або очікують у черзі. Дискретні стани системи S_k змінюються аналогічно до моделі системи з втратами *M*/*M*/*m*, але визначення стаціонарних імовірностей станів системи P_k складається з двох частин для станів системи:

$k = 0 \ldots m$ – кількість зайнятих серверів або вимог, що обслуговуються;

$k = m \ldots \infty$ – кількість зайнятих місць черги, або вимог, що очікують.

Інтенсивність потоку звільнення для станів системи $k = 0 \ldots m$ залежить від k і визначається аналогічно до її визначення у системі з втратами *M*/*M*/*m* – вона є змінною та дорівнює $k\mu$. Для станів системи $k = m \ldots \infty$ ця інтенсивність незмінна і дорівнює $m\mu$, оскільки при цьому вже є черга і зміна її стану залежить від звільнення будь-якого з усіх m серверів.

Розподіл кількості зайнятих серверів для станів $k = 0 \ldots m$ аналогічний розподілу (2.5) системи *M*/*M*/*m*:

$$P_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 \qquad (k = 1, 2, \ldots m). \tag{2.11}$$

Для станів $k = m \ldots \infty$ рівняння балансу: $(\lambda + m\mu)P_k = \lambda P_{k-1} + m\mu P_{k+1}$. Від нього підставлянням дістаємо співвідношення $P_k = \frac{1}{m}\frac{\lambda}{\mu} P_{k-1}$, що дає

$$P_k = \left(\frac{\lambda}{m\mu}\right)^{k-m} P_m . \tag{2.12}$$

Оскільки з (2.11) для P_m маємо:

$$P_m = \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} P_0 , \tag{2.13}$$

то у підсумку з (2.12) і (2.13) для станів $k = m \ldots \infty$ маємо, що

$$P_k = \left(\frac{\lambda}{m\mu}\right)^{k-m} \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} P_0 , \quad (k = m \ldots \infty). \tag{2.14}$$

Рівняння (2.11) і (2.14) визначають стаціонарний розподіл імовірностей станів системи P_k для всіх k. Імовірність P_0 для цього випадку визначається з умови $\sum_{k=0}^{\infty} P_k = 1$, звідки підставлянням значень P_k формул (2.11) та (2.14) маємо:

$$P_0 = \left[\sum_{k=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} \frac{m\mu}{m\mu - \lambda} \right]^{-1} . \tag{2.15}$$

Формула (2.15) існує за умови збіжності ряду $(\lambda / m\mu) < 1$ та (2.7), що є умовою ергодичності процесу. Це, у свою чергу, не дозволяє нескінченно зростати черзі, що є запорукою нормального функціонування всієї системи.

Оскільки $\Lambda = \lambda / \mu$, то за умови $\Lambda < m$ рівняння (2.11), (2.14) і (2.15) остаточно визначають стаціонарний розподіл імовірностей станів P_k, тому:

$$\begin{cases} P_k = \dfrac{\Lambda^k}{k!} P_0 & \text{при } k = 1 \dots m \qquad (2.16) \\ P_k = \left(\dfrac{\Lambda}{m}\right)^{k-m} \dfrac{\Lambda^m}{m!} P_0 & \text{при } k = m \dots \infty \qquad (2.17) \\ P_0 = \left[\sum\limits_{k=0}^{m-1} \dfrac{\Lambda^k}{k!} + \dfrac{\Lambda^m}{m!} \dfrac{m}{m-\Lambda} \right]^{-1} & \qquad (2.18) \end{cases}$$

Це є *другий розподіл Ерланга*, за яким можна розрахувати імовірність станів (зайнятості серверів та місць очікування у черзі) повнодоступної системи з необмеженою чергою при обслуговуванні пуассонівського потоку вимог.

Імовірність очікування P_w визначається з функції розподілу станів системи P_k. За пуассонівського потоку P_w дорівнює імовірності зайнятості всіх m серверів системи $P_{зн}$ з урахуванням усіх можливих станів черги:

$$P_w = P_{зн} = \sum_{k=m}^{\infty} P_k \,, \qquad (2.19)$$

де k – стан системи ($0 < k \le m$ – місця у серверах, $m < k \le \infty$ – місця у черзі).

Підставленням (2.17) в (2.19) і з урахуванням (2.18), отримаємо:

$$P_w = \frac{\dfrac{\Lambda^m}{m!} \dfrac{m}{m-\Lambda}}{\sum\limits_{k=0}^{m-1} \dfrac{\Lambda^k}{k!} + \dfrac{\Lambda^m}{m!} \dfrac{m}{m-\Lambda}}. \qquad (2.20)$$

Формула (2.20) є *С-формулою Ерланга* і позначається $C_m(\Lambda)$. Для системи $M/M/m/\infty$ за нею можна розрахувати імовірність очікування $P_w = C_m(\Lambda)$, яка дорівнює розрахунку за формулою (1.13).

Дану формулу можна переписати так:

$$C_m(\Lambda) = \frac{E_m(\Lambda) m}{m - \Lambda\left[1 - E_m(\Lambda)\right]}.$$

Ця формула встановлює взаємозв’язок між *В*- та *С*-формулами Ерланга, а це дозволяє розраховувати значення імовірності $C_m(\Lambda)$ за значеннями $E_m(\Lambda)$ формули (2.8), які надано у відповідних таблицях. Очевидно, що за однакових умов завжди імовірність $C_m(\Lambda) > E_m(\Lambda)$.

Середня довжина черги Q (середня кількість вимог у черзі) може бути знайдена з розподілу ймовірностей станів системи P_k при $k = m+1, m+2, \dots, \infty$ (в черзі 1, 2 і т.д. вимог) за правилом обчислення математичного сподівання:

$$Q = \sum_{k=m+1}^{\infty} (k-m)P_k \, .$$

З (2.17), (2.18) і (2.19) випливає, що $P_m = \left(1 - \frac{\Lambda}{m}\right) P_w$. Використовуючи це рівняння та формулу (2.13), підставляючи та скорочуючи маємо:

$$Q = \frac{\Lambda}{m - \Lambda} P_w \, . \tag{2.21}$$

За формулою Літтла за інтенсивності навантаження Λ за W одиниць часу очікування до черги надійде у середньому ΛW вимог:

$$Q = \Lambda W \, . \tag{2.22}$$

Тому з (2.21) та (2.22) середній час очікування для будь-якої вимоги

$$W = \frac{1}{m - \Lambda} P_w \, . \tag{2.23}$$

Середня кількість вимог у черзі Q – це інтенсивність навантаження, яке створюється вимогами, що очікують у черзі. Величина ΛP_w – це інтенсивність потоку вимог з черги, де кожна вимога в середньому очікує час t_q. Отже, $Q = \Lambda P_w t_q$, звідки з урахуванням (2.21) середній час очікування у черзі

$$t_q = \frac{1}{m - \Lambda} \, . \tag{2.24}$$

Величини W та t_q дано в одиницях середньої тривалості обслуговування, що показує, в скільки разів ці значення зростають або убувають порівняно із середнім часом обслуговування $\bar{x}$. З формул (2.23) та (2.24) видно, що $W \le t_q$.

Імовірність P_w показує середню частку затриманих вимог. З тих самих формул (2.23) та (2.24) видно, що

$$W = t_q P_w \, . \tag{2.25}$$

Середня кількість вимог у системі $N = \Lambda + Q$ (обслуговуються та очікують обслуговування), а середній час перебування вимоги у системі $T = \bar{x} + W$ і складається з середнього часу обслуговування та очікування.

Інтенсивність обслуженого навантаження Y менша за Λ і дорівнює:

$$Y = \Lambda(1 - P_w) \, . \tag{2.26}$$

Функція розподілу часу очікування у системі $P_{w>t}$ – це імовірність того, що вимога, яка надходить до системи в будь-якому з її станів очікує час понад t:

$$P_{w>t} = P_w e^{-(m\mu - \Lambda)t} \, .$$

Середня тривалість очікування у системі W:

$$W = \int_0^{\infty} P_{w>t} dt = \frac{P_w}{m\mu - \Lambda} \, ,$$

що за умови $\mu = 1$ (тобто W вимірюється в одиницях середньої тривалості обслуговування $\bar{x} = 1/\mu$) дає аналогічний до (2.23) результат.

Хоча дисципліна вибору з черги не впливає на середній час очікування W, але випадковий вибір з черги збільшує дисперсію часу очікування порівняно з вибором у порядку надходження.

2.4 Система з обмеженою чергою *M*/*M*/*m*/*r*

Повнодоступна схема СРІ з m серверами та обмеженою кількістю місць очікування у черзі r обслуговує пуассонівський потік вимог з дисципліною обслуговування черги *FIFO*. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні закони розподілу. Інтенсивність вхідного навантаження Λ. Необхідно визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0 \ldots m + r$) та характеристики *QoS*.

За обмеженої черги, де кількість місць очікування $r < \infty$, має місце комбінована система $M/M/m/r$ – з чергами та втратами. Вимозі відмовляється в обслуговуванні у тому разі, коли у системі знаходяться $l = m + r$ вимог. З цих l вимог m обслуговуються, а r очікують своєї черги.

Комбінована система $M/M/m/r$ частково аналогічна до системи $M/M/m/\infty$ тому рівняння (2.16) і (2.17) визначають стаціонарний розподіл імовірностей станів системи P_k для $k = 1 \ldots m$ та $k = m \ldots l$ відповідно. Імовірність P_0 визначається з умови нормування $\sum_{k=0}^{l} P_k = 1$, звідки підставленням усіх значень P_k формул (2.16) та (2.17) дістаємо:

$$P_0 \left[\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \sum_{k=m}^{l} \left(\frac{\Lambda}{m} \right)^{k-m} \right] = 1. \qquad (2.27)$$

Оскільки сума від $k = m$ до l може бути розрахованою за формулою суми $r = l - m$ членів геометричної прогресії зі знаменником Λ / m, то

$$P_0 = \left[\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \left(\frac{\Lambda^m}{m!} \right) \frac{1 - \left(\frac{\Lambda}{m} \right)^{r+1}}{1 - \frac{\Lambda}{m}} \right]^{-1}. \qquad (2.28)$$

Імовірність повної зайнятості системи $P_{зн}$ (очікування або втрати вимоги) визначається аналогічно (2.19), але сума береться не до ∞, а тільки до l. Замінивши цю суму подібно, як у формулах (2.27) і (2.28), маємо:

$$P_{зн} = \sum_{k=m}^{l} P_k = \frac{\Lambda^m}{m!} P_0 \frac{1 - \left(\frac{\Lambda}{m} \right)^{r+1}}{1 - \frac{\Lambda}{m}}. \qquad (2.29)$$

Підставлення P_0 з (2.28) у (2.29) дає

$$P_{зн} = \frac{\frac{\Lambda^m}{m!} \frac{m}{m - \Lambda} \left[1 - \left(\frac{\Lambda}{m} \right)^{r+1} \right]}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \frac{m}{m - \Lambda} \left[1 - \left(\frac{\Lambda}{m} \right)^{r+1} \right]}. \qquad (2.30)$$

Очевидно, якщо $r = \infty$, то формула (2.28) збігається з (2.18), а формула (2.30) – з *C*-формулою Ерланга (2.20). Якщо $r = 0$, то (2.28) збігається з (2.6), а формула (2.30) – з *B*-формулою Ерланга (2.9). З цього випливає, що при зміні довжини черги від ∞ до 0 величина імовірності $P_{зн}$ моделі *M*/*M*/*m*/*r* знаходиться в діапазоні значень, отриманих за формулами (2.20) та (2.8) для моделей *M*/*M*/*m*/∞ та *M*/*M*/*m*/0 відповідно. Отже, (2.30) є узагальнюючим розв'язком для повнодоступних систем з втратами, з необмеженою та обмеженою чергами за умови обслуговування пуассонівського потоку вимог та експонентної тривалості їх обслуговування.

Формулу (2.30) спростили з урахуванням того, що $\rho = \Lambda / m$:

$$P_{зн} = \frac{1-\rho^{r+1}}{\frac{1-\rho}{E_m(\Lambda)} + \rho - \rho^{r+1}}. \qquad (2.31)$$

У системі з обмеженою до *r* чергою подія втрати вимоги відбудеться у тому випадку, коли при надходженні вимоги у черзі вже є *r* вимог, що надійшли раніше. Отже, імовірність втрати вимоги P_B дорівнює імовірності того, що у системі вже є $l = m + r$ вимог (отже $r = l - m$) або система у стані P_l. Для пуассонівського потоку імовірність $P_B = P_l$ за будь-якого закону тривалості обслуговування (хоча саме значення P_B за різних законів буде різним).

Тому відповідно до (2.17) та (2.28), отримаємо:

$$P_B = P_l = \left(\frac{\Lambda}{m}\right)^{l-m} \frac{\Lambda^m}{m!} P_0 = \frac{\left(\frac{\Lambda}{m}\right)^r \frac{\Lambda^m}{m!}}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \left(\frac{\Lambda^m}{m!}\right) \frac{1-\left(\frac{\Lambda}{m}\right)^{r+1}}{1-\frac{\Lambda}{m}}}. \qquad (2.32)$$

Імовірність очікування за обмеженої до *r* черги $P_w(r)$ або частку вимог, які очікують: $P_w(r) = P_{зн} - P_B$. Підставивши сюди (2.30) і (2.32) відповідно, маємо

$$P_{w(r)} = \frac{\frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}\left[1-\left(\frac{\Lambda}{m}\right)^r\right]}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}\left[1-\left(\frac{\Lambda}{m}\right)^{r+1}\right]}. \qquad (2.33)$$

Формулу (2.32) з урахуванням (2.29) можна записати так:

$$P_B = \left(\frac{\Lambda}{m}\right)^{l-m} \frac{\Lambda^m}{m!} P_0 = P_{зн} \frac{\left(\frac{\Lambda}{m}\right)^r \left(1-\frac{\Lambda}{m}\right)}{1-\left(\frac{\Lambda}{m}\right)^{r+1}}.$$

Враховуючи те, що $\rho = \Lambda / m$, дану формулу можна спростити:

$$P_B = P_{\text{зн}} \frac{\rho^r (1-\rho)}{1-\rho^{r+1}}. \tag{2.34}$$

Підставивши у (2.34) формулу (2.31) маємо

$$P_B = \frac{\rho^r}{\dfrac{1}{E_m(\Lambda)} + \dfrac{\rho - \rho^{r+1}}{1-\rho}}. \tag{2.35}$$

Формулу (2.33) можна переписати не як різницю (2.30) і (2.32), а як різницю (2.31) та (2.34), що дає

$$P_w(r) = \frac{1-\rho^r}{\dfrac{1-\rho}{E_m(\Lambda)} + \rho - \rho^{r+1}}. \tag{2.36}$$

Із (2.34) і врахуванням (2.12), де $k = l$ та $r = l - m$, випливає, що

$$P_{\text{зн}} = P_l \frac{1-\rho^{r+1}}{\rho^r (1-\rho)} = P_m \frac{1-\rho^{r+1}}{1-\rho}. \tag{2.37}$$

Інтенсивність обслуженого навантаження Y або пропускна здатність менша вхідного навантаження на величину $\Lambda P_{\text{зн}}$, тобто:

$$Y = \Lambda(1 - P_{\text{зн}}). \tag{2.38}$$

Для системи $M/M/1/r$ із (2.28), (2.30), (2.33) та (2.34) можна розрахувати:

– імовірність P_0 (сервер і місця у черзі вільні) із (2.28)

$$P_0 = \left[1 + \rho \frac{1-\rho^{r+1}}{1-\rho}\right]^{-1} = \frac{1-\rho}{1-\rho^{r+2}}; \tag{2.39}$$

– імовірність $P_{\text{зн}}$ із (2.30) або $P_{\text{зн}} = 1 - P_0$, що дає

$$P_{\text{зн}} = \frac{\rho(1-\rho^{r+1})}{1-\rho^{r+2}}; \tag{2.40}$$

– імовірність $P_w(r)$ з (2.33) та підставленням $\rho = \Lambda / m$ дає

$$P_{w(r)} = \frac{\rho(1-\rho^r)}{1-\rho^{r+2}}; \tag{2.41}$$

– імовірність P_B із (2.34) та підставленням вище знайденого $P_{\text{зн}}$ дає

$$P_B = \frac{\rho^{r+1}(1-\rho)}{1-\rho^{r+2}}. \tag{2.42}$$

Із (2.14) випливає, що у системі $M/M/1/r$ стани системи мають розподіл

$$P_k = \rho^k P_0 = \frac{\rho^k (1-\rho)}{1-\rho^{r+2}}. \tag{2.43}$$

З цієї формули маємо геометричний розподіл станів системи

$$P_k = \rho^k (1-\rho), \tag{2.44}$$

який властивий для системи $M/M/m/r$.

2.5 Система з необмеженою чергою *M*/*G*/1/∞

Односерверна система з необмеженою кількістю місць очікування (рис. 2.1) обслуговує пуассонівський потік вимог. Тривалість обслуговування вимог має довільний (будь-який) закон розподілу $B(x)$ з математичним сподіванням $\bar{x}$. Інтенсивність потоку вимог λ. Визначити характеристики *QoS*.

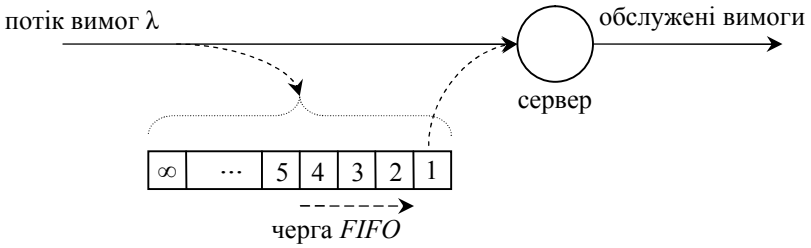


Рисунок 2.1 – Модель односерверної системи з чергою

Дана задача розв'язується методом вкладених ланцюгів Маркова, де для випадкового процесу, що описує поведінку моделі, конструюється зручний марковський ланцюг, досліджуваний аналітичними методами, звичайними для ланцюгів Маркова, і отримані результати (наприклад, стаціонарні імовірності станів) перераховуються для величин вихідної системи.

У моделі *M*/*G*/l/∞ імовірності того, що у стаціонарному стані, надходячи до системи, вимога застає в ній *j* інших вимог, дорівнюють стаціонарним ймовірностям P_j наявності у системі *j* вимог безпосередньо після закінчення обслуговування деякої вимоги. За пуассонівського потоку вимог ці імовірності дорівнюють стаціонарним ймовірностям P_j у будь-який момент часу. Математичне сподівання (середнє значення) кількості вимог *N*, що є у системі:

$$N = \rho + \frac{\lambda^2 \mu_x^2}{2(1-\rho)}, \tag{2.45}$$

де

$$\mu_x^2 = \int_0^\infty x^2 dB(x) = \bar{x}^2 + \sigma_x^2. \tag{2.46}$$

Параметри $\bar{x}^2$ та σ_x^2 характеризують математичне сподівання та дисперсію функції розподілу випадкової тривалості обслуговування вимог *x*.

Формулу (2.45) називають формулою Поллачека-Хінчина.

Кількість вимог, що перебувають у системі, коли її залишає деяка вимога, дорівнює кількості вимог, що надходять до системи за час перебування у системі вимоги, що йде з неї. Тому між функцією розподілу часу перебування $V(t)$ *i* розподілом P_j існує такий зв'язок:

$$P_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dV(t). \tag{2.47}$$

Середній час очікування у системі *W* для стаціонарного стану:

$$W = \frac{\lambda \mu_x^2}{2(1-\rho)} = \frac{\lambda(\bar{x}^2 + \sigma_x^2)}{2(1-\rho)}. \qquad (2.48)$$

Для характеристики другого центрального моменту розподілу тривалості обслуговування σ_x^2 часто використовується *коефіцієнт варіації* ν_x:

$$\nu_x = \frac{\sigma_x}{\bar{x}}. \qquad (2.49)$$

Суму $\bar{x}^2 + \sigma_x^2$, з урахуванням (2.49), можна представити як $\bar{x}^2(1+\nu_x^2)$. Підставивши це до (2.45) і (2.48), дістанемо формули розрахунку N та W. З урахуванням (1.15) та (1.16) отримуємо всі інші характеристики *QoS*, формули яких занесені до табл. 2.1. Значення W, t_q та T нормуються за $\bar{x}$.

Таблиця 2.1 – Характеристики якості обслуговування моделі $M/G/1/\infty$

Характеристика *QoS*	Модель		
	$M/G/1/\infty$	$M/D/1/\infty$	$M/M/1/\infty$
P_w	ρ	ρ	ρ
Q	$\frac{\rho^2(1+\nu_x^2)}{2(1-\rho)}$	$\frac{\rho^2}{2(1-\rho)}$	$\frac{\rho^2}{1-\rho}$
W	$\frac{\rho(1+\nu_x^2)}{2(1-\rho)}$	$\frac{\rho}{2(1-\rho)}$	$\frac{\rho}{1-\rho}$
t_q	$\frac{1+\nu_x^2}{2(1-\rho)}$	$\frac{1}{2(1-\rho)}$	$\frac{1}{1-\rho}$
N	$\rho + \frac{\rho^2(1+\nu_x^2)}{2(1-\rho)}$	$\rho + \frac{\rho^2}{2(1-\rho)}$	$\rho + \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho}$
T	$1 + \frac{\rho(1+\nu_x^2)}{2(1-\rho)}$	$1 + \frac{\rho}{2(1-\rho)}$	$1 + \frac{\rho}{1-\rho} = \frac{1}{1-\rho}$

До табл. 2.1 записано й формули всіх характеристик *QoS* моделей $M/D/1/\infty$ та $M/M/1/\infty$, для яких коефіцієнт варіації ν тривалості обслуговування регулярного та експонентного розподілів дорівнює 0 та 1 відповідно. З таблиці випливає, що за постійної тривалості обслуговування значення W, t_q та Q в два рази менші, ніж за експонентної. Надані в табл. 2.1 значення характеристик *QoS* для моделі $M/M/1/\infty$ відповідно збігаються з (2.20), (2.21), (2.23) і (2.24).

Для m-серверної системи $M/G/m/\infty$ простого розв'язання не існує, однак є наближена оцінка середнього часу очікування W_m [6]

$$W_m \le \frac{\bar{x}^2 + \sigma_x^2}{2m - 2\lambda\bar{x}}. \qquad (2.50)$$

2.6 Система з пріоритетами *M*/*G*/1/∞

Введення пріоритетів для вимог (надходять або очікують) – ефективний спосіб керування розмірами черги і часом перебування в ній. При надходженні до системи вимоги з високим пріоритетом обслуговування вимоги з більш низьким пріоритетом або переривається (абсолютний пріоритет), або вимога з високим пріоритетом стає в початок черги вимог (відносний пріоритет) [1, 2].

2.6.1 Відносний пріоритет

Вимоги з r пріоритетами (менше номер – вище пріоритет) надходять до односерверної системи і утворюють r пуассонівських потоків з інтенсивністю λ_k, де $k = 1 \ldots r$. Сукупний потік є пуассонівським з інтенсивністю $\lambda_C = \lambda_1 + \lambda_2 + + \ldots, \lambda_r$. Вимоги кожного пріоритету обслуговуються в порядку надходження без переривань обслуговування. Після закінчення обслуговування будь-якої вимоги з черги вибирається вимога з вищим пріоритетом.

Інтенсивність питомого навантаження k-го пріоритету:

$$\rho_k = \frac{\lambda_k}{\mu_k}, \qquad \text{де} \qquad k = 1, \ldots, r, \tag{2.51}$$

а сумарне навантаження всіх потоків з 1 по k-й пріоритет:

$$R_k = \sum_{i=1}^{k} \rho_i, \quad \text{де } R_0 = 0, \quad k = 1, \ldots, r. \tag{2.52}$$

Сумарне навантаження потоків усіх пріоритетів $R_r < 1$ (ергодичність), тому з використанням (2.46) середній час очікування вимоги з k-м пріоритетом

$$W_k = \frac{\lambda_C(\bar{x}^2 + \sigma_x^2)}{2(1 - R_{k-1})(1 - R_k)}. \tag{2.53}$$

На рис. 2.2 показано графіки залежності середнього часу очікування у системі вимог k-го пріоритету від сукупної інтенсивності навантаження ρ потоків усіх, наприклад, п’яти пріоритетів.

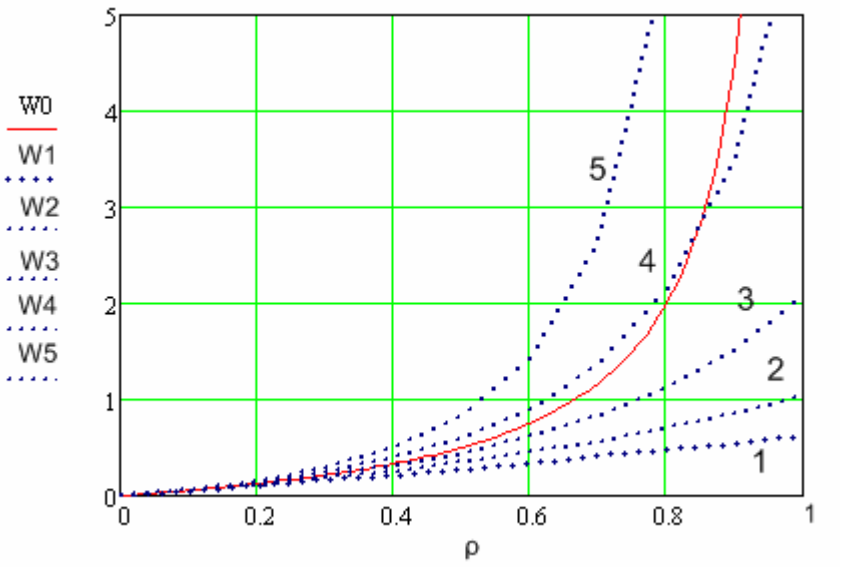


Рисунок 2.2 – Вплив відносних пріоритетів на середній час очікування

Пунктирними лініями зображені графіки середнього часу очікування W_k, вираженого в одиницях середньої тривалості обслуговування $\bar{x}$ при відносних пріоритетах (1, …, 5) і постійному часі обслуговування, однаковому для всіх п'яти однакових за інтенсивністю λ_k пуассонівських потоків вимог. Видно, що при введенні пріоритетів середній час очікування для окремих потоків є менше, ніж у випадку без пріоритетного обслуговування (безперервна лінія).

Якщо $r = 2$, то вимоги першого пріоритету є пріоритетними, а другого – непріоритетними. Середній час очікування вимог визначиться з (2.53):

$$W_1 = \frac{\lambda_C(\bar{x}^2 + \sigma_x^2)}{2(1-\rho_1)}; \qquad W_2 = \frac{\lambda_C(\bar{x}^2 + \sigma_x^2)}{2(1-\rho_1)(1-\rho_1-\rho_2)}.$$

Середній час перебування вимоги k-го пріоритету у системі T_k визначається як сума W_k та середнього часу обслуговування $1 / \mu_k$.

2.6.2 Абсолютний пріоритет

У цьому разі, якщо до системи надходить вимога більш високого пріоритету, то переривається обслуговування вимоги низького пріоритету.

З (2.53) середній час очікування вимог першого пріоритету

$$W_1 = \frac{\lambda_1(\bar{x}_1^2 + \sigma_1^2)}{2(1-\rho_1)}, \tag{2.54}$$

а згідно з (2.45) середня кількість вимог першого пріоритету у системі

$$N_1 = \rho_1 + \frac{\lambda_1^2(\bar{x}_1^2 + \sigma_1^2)}{2(1-\rho_1)}. \tag{2.55}$$

Загальний вираз для середнього часу очікування вимог інших пріоритетів $1 \le k \le r$ маємо аналогічно (2.53):

$$W_k = \frac{\sum_{i=1}^{r} \lambda_i(\bar{x}_i^2 + \sigma_i^2)}{2(1-R_{k-1})(1-R_k)}. \tag{2.56}$$

Наприклад, при $r = 2$ є пріоритетні і непріоритетні вимоги (пріоритети 1 і 2). Для непріоритетних вимог оцінимо середні значення: часу очікування у системі W_2, часу перебування у системі T_2, кількості вимог у системі N_2:

$$W_2 = \frac{\lambda_1(\bar{x}_1^2 + \sigma_1^2) + \lambda_2(\bar{x}_2^2 + \sigma_2^2)}{2(1-\rho_1)(1-\rho_1-\rho_2)}; \qquad T_2 = W_2 + \frac{1}{\mu_2(1-\rho_1)};$$

$$N_2 = \frac{1}{1-\rho_1}\left[\rho_2 + \frac{\lambda_1\lambda_2(\bar{x}_1^2 + \sigma_1^2) + \lambda_2(\bar{x}_2^2 + \sigma_2^2)}{2(1-\rho_1-\rho_2)}\right],$$

де λ_2 – інтенсивність потоку непріоритетних вимог. Тривалість обслуговування визначається розподілом $B_2(x)$ із середнім значенням $\bar{x}_2 = 1/\mu_2$ та $\rho_2 = \lambda_2/\mu_2$.

3 МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА ДЛЯ МУЛЬТИСЕРВІСНИХ МЕРЕЖ ЗВ'ЯЗКУ

3.1 Модель трафіка мультисервісних мереж зв'язку

Інтегральний характер мультисервісної мережі зв'язку з розширеним спектром надаваних послуг зумовлює різнорідність трафіка, яка сильно змінює його параметри та математичну модель. Параметрами мультисервісного трафіка є інтенсивність навантаження Λ (середня кількість вимог, що надійшла до системи за середню тривалість обслуговування x) та дисперсія інтенсивності навантаження σ^2 (див. табл. 1.1). Математичною моделлю трафіка є імовірнісна функція розподілу випадкової величини z (інтервал часу між вимогами) або випадкової величини i (кількість вимог за час обслуговування вимоги $\bar{x}$).

Потоки вимог у мультисервісних мережах формуються множиною джерел з різною питомою інтенсивністю навантаження λ_i (різнорідні потоки). В процесі створення потоку вимог беруть участь джерела, що належать до i-ї групи споживачів k сервісів з близькими інтенсивностями навантаження. Значення інтенсивності результуючого потоку вимог в кожну мить залежить від того, до якої i-ї групи за інтенсивністю навантаження належить джерело і яке співвідношення кількості цих джерел з іншими. Отже, адекватною моделлю такого потоку, тобто розподілом інтервалів часу між вимогами буде не експонентний розподіл (M), а їх суміш – гіперекспонентний розподіл (H):

$$P(z) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i z} \qquad \text{при} \sum_{i=1}^{k} p_i = 1. \tag{3.1}$$

Цей розподіл описує більший розкид величини інтервалу часу між вимогами z і забезпечує значення коефіцієнтів варіації $v_z \geq 1$, а це дозволяє описувати мультисервісний трафік з дисперсією інтенсивності навантаження σ^2, що перевищує її математичне сподівання Λ від одиниць до десятків разів.

Гіперекспонентний інтервал часу між вимогами призводить до такого розподілу вхідної кількості вимог i за середню тривалість їх обслуговування $\bar{x}$, а це є інтенсивність навантаження Λ, який апроксимується *нормальним* (Гауса) законом. Отже, для трафіка мультисервісних мереж адекватною є математична модель з гіперекспонентним розподілом інтервалу часу між вимогами (іноді достатньо другого порядку) або з нормальним розподілом інтенсивності навантаження з параметрами Λ та σ^2:

$$P_i = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{\frac{-(i-\Lambda)^2}{2 \cdot \sigma^2}}. \tag{3.2}$$

Гістограми статистики вимірів трафіка з апроксимацією пуассонівським та нормальним законами розподілу показані на рис. 3.1. Перевірка за критерієм Пірсона показала, що для пуассонівської апроксимації $\chi^2 = 4{,}51$ при рівні значимості $p = 0{,}34$, а для апроксимації нормальним законом $\chi^2 = 1{,}29$ при $p = 0{,}87$. Менше значення χ^2 та більше значення рівня значимості свідчать на користь гіпотези про нормальний закон розподілу статистичних даних.

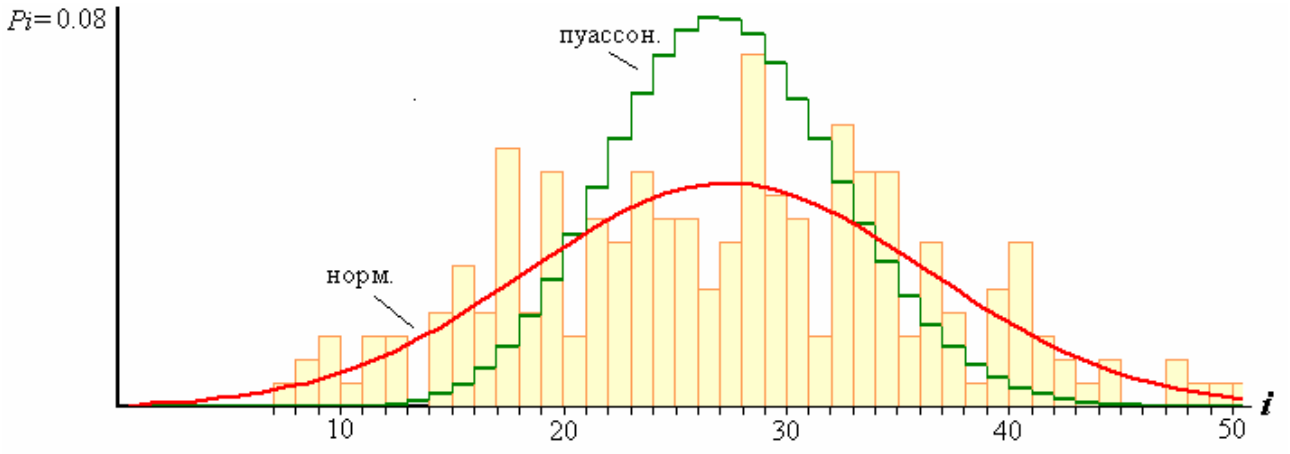


Рисунок 3.1 – Гістограми статистичних вимірів трафіка та апроксимуючі криві

Для цього прикладу інтенсивність вхідного навантаження $\Lambda = 27$ (це є математичне сподівання кількості вимог), а середньоквадратичне відхилення $\sigma = 9$ (дається в тих самих одиницях виміру, що і математичне сподівання). При цьому коефіцієнт скупченості навантаження $S = \sigma^2 / \Lambda = 3$.

Для мультисервісних мереж зв'язку властива підвищена нерівномірність трафіка, за якої дисперсія інтенсивності трафіка σ^2 перевищує її математичне сподівання Λ від 2 до 15 разів. Іноді дане перевищення є більшим, але це відбувається або за межами ГНН, або на невеликих пучках каналів [3]. Отже, у мультисервісного трафіка дисперсія інтенсивності навантаження $\sigma^2 > \Lambda$.

Нерівномірність надходження вимог характеризується коефіцієнтом варіації інтервалу часу між вимогами $v_z = \sigma_z / \bar{z}$ і для мультисервісного трафіка $v_z > 1$. Зі збільшенням v_z співвідношення перших двох центральних моментів інтенсивності трафіка теж збільшується. У [13] за результатами статистичного моделювання встановлено, що у випадку гіперекспонентного розподілу інтервалу часу між вимогами коефіцієнт варіації інтервалу часу v_z і пікфактор інтенсивності трафіка S знаходяться в такій залежності:

$$v_z^2 = \left(\frac{\sigma_z}{\bar{z}} \right)^2 \approx S, \qquad (3.3)$$

де σ_z – стандартне відхилення та $\bar{z}$ – математичне сподівання випадкового інтервалу z.

Для підвищення точності розрахунку характеристик якості обслуговування можна враховувати не тільки математичне сподівання і дисперсію інтенсивності трафіка, але і моменти розподілу більш вищого порядку, таких як асиметрія та ексцес.

3.2 Апроксимація Хейворда

Повнодоступна схема СРІ з m серверами функціонує за дисципліною обслуговування з втратами. Середня тривалість обслуговування дорівнює $\bar{x}$. До системи надходить неординарний потік вимог з інтенсивністю Λ, в якому вимоги надходять однаковими групами через інтервали часу з експонентним розподілом, причому кожна надійшовши одночасно група вимог кількістю $i > 1$

вимог потребує для свого обслуговування одночасно $k > 1$ серверів відповідно (одна вимога групи при з'єднанні займає один сервер групи). У разі завершення обслуговування (з'єднання) всі сервери даної групи одночасно звільняються. Якщо в момент надходження групи вимог у системі відсутня необхідна кількість k вільних серверів, то вся група вимог кількістю i втрачається.

При визначенні ймовірності втрат вимог π для аналізованої системи, яку названо системою U, пряме застосування B-формули Ерланга неможливе через неординарність потоку вимог і займань серверів відповідно. Замість цього досліджується модифікована система U', яка складається з v комплектів серверів (комплект k серверів), де

$$v = m / k \,. \tag{3.4}$$

Кожній групі вхідних вимог для обслуговування потрібний один комплект і потік зайняття комплектів є ординарним.

Навантаження у системі U' визначається кількістю зайнятих комплектів (а не серверів), тобто з точки зору груп вимог потік ординарний з інтенсивністю груп $\Lambda_g = \Lambda / k$**.** Цей потік експонентний і тому вхідне навантаження за групами є пуассонівським, де математичне сподівання Λ_g і дисперсія D_g рівні – $\Lambda_g = D_g$. Тому для розрахунку ймовірності втрат групи вимог π' у системі U' можна вжити B-формулу Ерланга (2.9) з параметрами Λ_g та v, тобто знайти $E_v(\Lambda_g)$:

$$\pi' = E_v(\Lambda_g) = \frac{\dfrac{\Lambda_g}{v!}}{\displaystyle\sum_{i=0}^{v} \frac{\Lambda_g^i}{i!}}. \tag{3.5}$$

У системі U кількість зайнятих серверів в k разів більше кількості груп вимог, що обслуговуються (з'єднань). Звідси інтенсивність вхідного навантаження на сервери і його дисперсія:

$$\Lambda = k\Lambda_g \,, \qquad D = k^2 D_g \,. \tag{3.6}$$

Тут застосовується наступне правило, відоме з теорії ймовірностей: якщо випадкова величина збільшується на постійний коефіцієнт k, то на цей самий коефіцієнт потрібно помножити її математичне сподівання, а дисперсія повинна збільшуватися на коефіцієнт k у квадраті.

З даних формул видно, що $D > \Lambda$ і це навантаження є скупченим через неординарність потоку займань серверів. Тут коефіцієнт скупченості S (1.9) дорівнює кількості серверів, які потрібні для обслуговування одного з'єднання:

$$S = \frac{D}{\Lambda} = \frac{k^2 D_g}{k\Lambda_g} = k \,. \tag{3.7}$$

Втрата групи вимог з імовірністю π' у системі U' означає й втрату кожної окремої вимоги з групи у системі U. Тому імовірність втрати вимог π у системі U з параметрами Λ та m можна розрахувати як імовірність втрати вимог π' у системі U' з відповідними параметрами інтенсивності навантаження Λ / S, що випливає з (3.6) та (3.7), та ємністю системи m / S, що випливає з (3.4) та (3.7). Звідси з урахуванням вказаних співвідношень маємо:

$$\pi = E_{(m/S)}\left(\frac{\Lambda}{S}\right). \qquad (3.8)$$

Тобто, щоб імовірності втрати вимоги π та групи вимог π' відповідали одна одній як в єдиній системі ($\pi = \pi'$), необхідно імовірність π розраховувати не у системі U з непуассонівським потоком, а у системі U' з пуассонівським потоком, але зі зменшеними у S разів параметрами B-формули Ерланга.

Якщо вимоги надходять не однаковими групами з кількістю $i \geq 1$ вимог, то відповідно для обслуговування вхідних вимог необхідна й не однакова кількість $k \geq 1$ вільних серверів. Так може бути у мультисервісних мережах, де одними й тими ж трактами зв'язку передаються мовні потоки, потоки даних, відео й ін. При цьому передавання окремих видів інформації вимагає різного ресурсу або кількості серверів (наприклад, якщо під сервером розуміється порт певної швидкості і треба зайняти одночасно кілька портів для досягнення необхідної для послуги швидкості передавання інформації). Тому, в залежності від категорії з'єднання на надання певної послуги для кожного з потоків інформації необхідно зайняти певний неоднаковий ресурс із спільного ресурсу пропускної здатності. При цьому коефіцієнт скупченості мультисервісного навантаження вже дорівнює середньозваженій кількості серверів k (3.7), які потрібні для обслуговування навантажень з'єднань окремих категорій, з вагами $k_i\Lambda_i$ та $k_i^2 D_i$, що дорівнюють навантаженню та дисперсії з'єднань i-ї категорії. Отже, коли мультисервісне навантаження створюється декількома категоріями джерел з різною кратністю з'єднання k_i, суперпозицію вхідних потоків можна замінити одним потоком з інтенсивністю навантаження Λ і коефіцієнтом скупченості S об'єднаного навантаження.

Для різних категорій джерел навантаження вимоги втрачаються за різних станів системи і, як наслідок, імовірнісні характеристики якості обслуговування вимог будуть відрізнятися, що дає наближену оцінку середніх (або загальних) втрат. Порівняльний аналіз такого підходу й імітаційного моделювання свідчить про те, що запропонована апроксимація Хейворда (3.8) є наближеною і лише за постійної тривалості обслуговування має точність, цілком достатню для розв'язання інженерних задач.

Хоча реально непуассонівські потоки мають досить складні статистичні властивості, і для повного опису таких потоків потрібне використання більшої кількості характеристик, на практиці звичайно припускають, що імовірність втрати вимоги слабко залежить від моментів навантаження більш високого порядку і їх можна не враховувати.

3.3 Функція розподілу станів системи з втратами *H*/*D*/*m*

Повнодоступна схема СРІ з m серверами функціонує за дисципліною обслуговування з втратами. Тривалість обслуговування постійна й дорівнює $\bar{x}$. До системи надходить потік вимог з інтенсивністю Λ, в якому інтервал між вимогами має гіперекспонентний розподіл, а кількість вимог Λ за одиницю

часу $\bar{x}$ розподілена за нормальним законом. Визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0, \ldots, m$).

Стан системи визначається кількістю зайнятих серверів. Випадковий процес надходження вимог модифікує стан системи зі швидкістю, обумовленою інтенсивністю навантаження Λ. Інтенсивність навантаження Λ – це сумарна кількість вимог j, які надійшли за час, що дорівнює середній тривалості обслуговування $\bar{x}$. Таким чином, інтенсивність переходу системи з одного стану в інший залежить від властивостей потоку вимог, що описується відповідним розподілом імовірностей Q_j, де j – кількість вимог за час $\bar{x}$, яка може бути в діапазоні від 0 до ∞. Усі події надходження вимог належать простору станів Q. Події зайняття серверів утворюють новий дискретний підпростір P, який описано розподілом імовірностей P_i, де i – кількість зайнятих серверів ($i = 0, \ldots, m$). Простір P, безсумнівно, менше простору Q, оскільки стани $i > m$ для системи неможливі, а j може бути як завгодно велике.

Оскільки система обслуговування з втратами, то незалежно від її поточного стану для кожного з випадків надходження $j > m$ вимог відбувається подія «відмова в обслуговуванні». Це очевидно, оскільки за час $\bar{x}$ (постійна тривалість зайняття серверів) надійде $j > m$ вимог і жоден зі знову зайнятих серверів за цей самий час не звільниться. Якщо система перебуває в початковому стані $i = 0$ (усі сервери вільні), то в цьому випадку для кожного з варіантів надходження за той самий відрізок часу $j \le m$ вимог подія «відмова в обслуговуванні» не відбувається. Події, які полягають в надходженні за час $\bar{x}$ точно j вимог, утворюють повну групу неспільних гіпотез $H_0, H_1, \ldots, H_j$ з апріорними ймовірностями $Q(H_0), Q(H_1), \ldots, Q(H_j)$ і тому $\sum_{j=0}^{\infty} Q(H_j) = 1$. Подія A, яка полягає у „відсутності відмови в обслуговуванні” може відбуватися тільки разом з однією із гіпотез групи $H_0, H_1, \ldots, H_m$. Умовні імовірності гіпотез надходження $j \le m$ вимог (апостеріорні), за умови здійснення події A (відсутність втрат вимог), обчислюються за формулою Байеса:

$$Q(H_j \mid A) = \frac{Q(H_j)Q(A \mid H_j)}{\sum_{k=0}^{m} Q(H_k)Q(A \mid H_k)}.$$

Через те що подія A (відсутність відмови в обслуговуванні), яка з’являється з кожною із гіпотез групи $H_0, H_1, \ldots, H_m$ є достовірною, то умовні імовірності $Q(A|H_j) = 1$. Тому, якщо для здійснення події можлива тільки частина гіпотез, а інші неможливі, то для отримання апостеріорних ймовірностей необхідно кожну з апріорних ймовірностей цієї частини можливих гіпотез розділити на їх суму. Отже:

$$Q(H_j \mid A) = \frac{Q(H_j)}{\sum_{k=0}^{m} Q(H_k)}.$$

Для системи, що перебуває в початковому стані (усі сервери вільні), умовні імовірності $Q(H_j|A)$ відповідають ймовірностям зайняття $i = j$ серверів системи, тобто ймовірностям P_i. Вони також утворять повну групу неспільних подій підпростору P і тому $\sum_{i=0}^{m} P_i = 1$. Таким чином, імовірності зайняття серверів P_i представлено через імовірності надходження вимог $Q(H_j)$, тобто

$$P_i = \frac{Q(H_i)}{\sum_{k=0}^{m} Q(H_k)}. \qquad (3.9)$$

Умовою стаціонарності отриманого розподілу є ергодичність процесу обслуговування вимог. Це означає, що отримані імовірності станів системи (кількості зайнятих серверів) не повинні залежати від того, у якому стані система була у початковий момент (вважалось, що в початковий момент усі сервери вільні). Відомо, що властивість ергодичності мають *марковські* процеси, для яких після досить тривалого часу функціонування системи обов'язково настане стаціонарний режим, де імовірність того, що система буде в *i*-му стані, не залежить від того, у якому стані вона знаходилася в початковий момент часу. Відповідно до теореми Маркова умовою ергодичності процесу є (2.7) і тоді система працює у стані статистичної рівноваги.

Підставивши в (3.9) замість імовірностей $Q(H_i)$ імовірності з розподілу Пуассона (2.2), дістаємо перший розподіл Ерланга (2.8), справедливий для моделі *M*/*G*/*m*, і за яким можна розрахувати імовірності всіх станів системи. Однак у цьому випадку, неодмінно, математичне сподівання випадкового процесу надходження вимог з інтенсивністю Λ дорівнює її дисперсії σ^2. Але якщо при розрахунках (3.9) замість імовірностей $Q(H_i)$ і $Q(H_k)$ підставити імовірності, що розраховано за нормальним законом, то дістаємо формулу розрахунку станів системи саме в моделі *H*/*D*/*m*:

$$P_i = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-(i-\Lambda)^2/2\sigma^2}}{\sum_{k=0}^{m} \frac{1}{\sqrt{2\pi}\sigma} e^{-(k-\Lambda)^2/2\sigma^2}} \approx \frac{1}{\sum_{k=0}^{m} \exp\left[\frac{-(k-2\Lambda+i)(k-i)}{2\sigma^2}\right]}. \qquad (3.10)$$

При $\sigma^2 = \Lambda$ нормальний випадковий процес є *марковським* і при цьому відповідно до (2.7) при $\Lambda < m$ розрахункові значення, отримувані за формулами розподілу Ерланга (2.8) і (3.10) досить близькі – розбіжність не більше 1 %. При $\sigma^2 > \Lambda$ розподіл інтервалів часу між вимогами вже не експонентний і потік вимог втрачає властивість відсутності післядії. Процес обслуговування стає складнішим і виникає залежність ймовірностей станів системи від її початкового стану. Крім того, виникає залежність і від виду розподілу тривалості обслуговування вимог, на відміну від розподілу Ерланга, точному за будь-якого закону розподілу тривалості обслуговування. Через це імовірність зайняття всіх серверів системи вже не збігається з імовірністю відмови в обслуговуванні, як це спостерігається за пуассонівського потоку.

Дану задачу для моделі *H*/*D*/*m* можна вирішити і в інший спосіб.

В умовах необмеженої кількості серверів (модель $H/D/\infty$) вимоги обслуговуються без втрат. За постійної тривалості обслуговування x, коли нема втрат, властивості потоку звільнень збігаються із властивостями потоку надходження вимог, тому що відбувається тільки зсув у часі на величину x між моментом надходження вимоги та моментом її виходу із системи (завершення обслуговування). При цьому стани системи повністю визначаються властивостями потоку вимог, а імовірнісні функції розподілу кількості вимог у системі i і вхідної кількості вимог j за час x, повністю збігаються. За нормального розподілу кількості вимог, що надходять у систему за середню тривалість їх обслуговування, нормальна функція розподілу потоку вимог P_j визначить функцію розподілу станів системи P_i:

$$P_j = P_i = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(i-\Lambda)^2}{2\sigma^2}}. \qquad (3.11)$$

За обмеженої до m кількості серверів простір станів системи буде також обмеженим від 0 до m. У цьому випадку моделі $H/D/m$ імовірнісна функція розподілу станів системи може бути апроксимована усіченим нормальним законом, що визначає імовірності станів системи P_i в межах $0 \le j \le m$:

$$P_i = \frac{A}{\sqrt{2\pi}\sigma} e^{-\frac{(i-\Lambda)^2}{2\sigma^2}}, \qquad (3.12)$$

де

$$A = \frac{1}{\frac{1}{\sqrt{2\pi}}\left[\int_0^{\frac{m-\Lambda}{\sigma}} e^{-\frac{u^2}{2}} du - \int_0^{\frac{0-\Lambda}{\sigma}} e^{-\frac{u^2}{2}} du\right]}.$$

Якщо $m = \infty$, то $A = 1$ і тому (3.11) та (3.12) збігаються. За всіх інших m значення формул (3.10) та (3.12) збігаються, що підтверджує вірність обох способів розв'язання задачі. Для (3.10), (3.11) та (3.12) виконується $\sum_{i=0}^{m} P_i = 1$.

На рис. 3.2 показані графіки апроксимації функції розподілу станів системи, яка обслуговує потік вимог з параметрами: інтенсивність $\Lambda = 100$ Ерл та дисперсія $\sigma^2 = 400$ (пікфактор $\sigma^2 / \Lambda = 4$). Пунктирною лінією показана система з необмеженою кількістю серверів, де $m = \infty$; безперервною – система з $m = 125$; штриховою – з $m = 115$ та штрих-пунктирною – з $m = 110$ серверів.

В широкому діапазоні зміни параметрів трафіка Λ і σ^2 та кількості серверів m відносна похибка запропонованої апроксимації усіченим нормальним законом не перевищує 1% для всіх значень P_i, окрім $P_{j=m}$. Процес обслуговування вимог у системі є ергодичним або не залежить від початкового стану системи за умови $m > \Lambda$ [2]. Якщо кількість серверів m суттєво перевищує значення інтенсивності Λ, то початковий стан системи не дуже впливає на функцію розподілу станів системи. При зменшенні m початковий стан системи

виявляє свій найбільший вплив саме на імовірність $P_{i=m}$. Це відбувається через «скупченість» потоку вимог реального трафіка, яка випливає з його властивості $\sigma^2 > \Lambda$. Саме у разі «перевантаження» системи, коли зайняті всі сервери, при звільненні будь-якого з серверів він тут же займається черговою вимогою через скупченість вимог на даному інтервалі часу, а це підтримує систему більш тривалий час у стані «насичення», тобто у стані $P_{i=m}$.

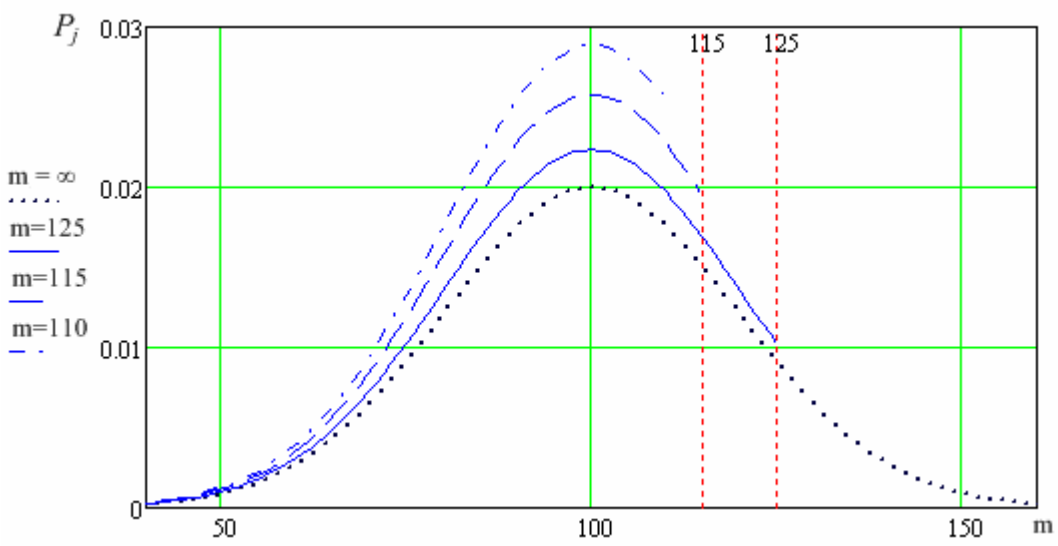


Рисунок 3.2 – Апроксимація станів усіченим нормальним законом

При $\sigma^2 = \Lambda$ закони Гауса та Пуассона майже збігаються вже при $\Lambda > 20$. Тому тут даний усічений нормальний розподіл дає значення ймовірностей станів системи, які дуже близькі до значень, що отримуються за розподілом Ерланга (2.8). З (2.8) визначаються стани системи типу $M/G/m$ з експонентним розподілом інтервалу часу між вимогами потоку, для якого $\sigma^2 \equiv \Lambda$. Гіперекспонентий розподіл за $k = 1$ обертається в експонентний, і тому він є його узагальненням. Але на відміну від розподілу Ерланга, який дійсний за будь-якого (G) закону розподілу тривалості обслуговування вимог, усічений нормальний розподіл станів системи при $\sigma^2 > \Lambda$ справедливий тільки для моделі $H/D/m$, тобто за постійної тривалості обслуговування.

3.4 Імовірність втрат системи *H*/*D*/*m*

Визначення основної характеристики *QoS* – імовірності відмови в обслуговуванні через зайнятість усіх серверів системи – базується на визначенні імовірнісної функції розподілу станів системи P_i.

У моделі $H/D/m$ імовірність відмови в обслуговуванні залежить від пікфактора інтенсивності навантаження S і за умови $m > \Lambda$ дорівнює імовірності зайняття всіх m серверів системи (3.10), помноженій на S:

$$P_B = \frac{1}{\sum_{k=0}^{m} \exp\left[\frac{-(k-2\Lambda+m)(k-m)}{2\sigma^2}\right]} \frac{\sigma^2}{\Lambda}. \qquad (3.13)$$

Графіки залежності ймовірності втрат вимог від кількості каналів (серверів) та коефіцієнта S показано на рис. 3.3. Всі графіки відповідають інтенсивності навантаження $\Lambda = 100$ Ерл. Пунктирна крива E_m дає залежність ймовірності втрат від m, розрахованих за формулою Ерланга [8].

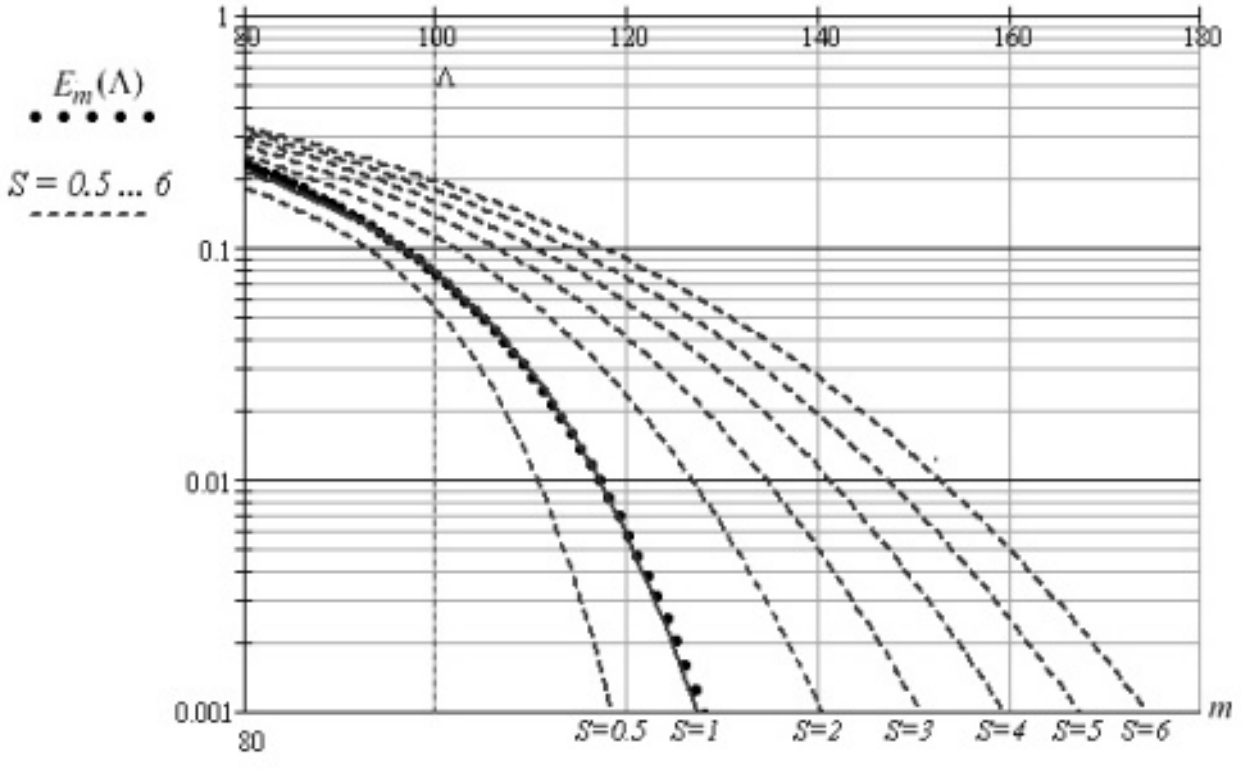


Рисунок 3.3 – Графіки залежності ймовірності втрат вимог від кількості каналів m та коефіцієнта скупченості S

Як видно з рисунку, графіки ймовірностей, розрахованих за формулою Ерланга та запропонованою формулою (3.13) для $S = 1$, збігаються. При $S < 1$ (більш «вирівняний» потік) імовірність втрат вимог зменшується. При збільшенні коефіцієнта S імовірність втрат вимог за однакової ємності пучка суттєво зростає, на що вказують графіки для $S = 2, ..., 6$ (штрихові лінії).

У [8] показано, що регулярний закон розподілу тривалості обслуговування з точки зору якості обслуговування є найгіршим і тому формула (3.13) є не що інше, як верхня оцінка можливих значень імовірності відмови в обслуговуванні. Для інших законів розподілу тривалості обслуговування вимог до формули (3.13) застосовується спеціальна апроксимаційна функція, яка отримана за результатами імітаційного моделювання [7], і в якій коефіцієнтом h задається вид закону розподілу тривалості обслуговування:

$$P_B = \frac{S}{\sum_{k=0}^{m} \exp\left[\frac{-(k-2\Lambda+m)(k-m)}{2\Lambda S}\right]}\left[1 - \frac{(S^2-1)(\sigma-\Lambda+m)}{\sigma(S^2 h - h + 5)}\right], \qquad (3.14)$$

де h дорівнює 4.25, 3.55 і 2.85 для рівномірного, експонентного і логарифмічно нормального законів розподілу тривалості обслуговування відповідно. Видно, що при $S = 1$ (тобто $\sigma^2 = \Lambda$) дана формула перетворюється у формулу (3.10) для випадку $i = m$ – стан системи m (зайняті всі сервери).

На рис. 3.4 показано залежності імовірності втрат P_B від кількості серверів у системі $H/G^*/m$ та виду закону розподілу тривалості обслуговування при $\Lambda = 100$ Ерл та $S = 8$ (G^* – регулярний, рівномірний, експонентний та логарифмічно нормальний закони розподілу тривалості обслуговування).

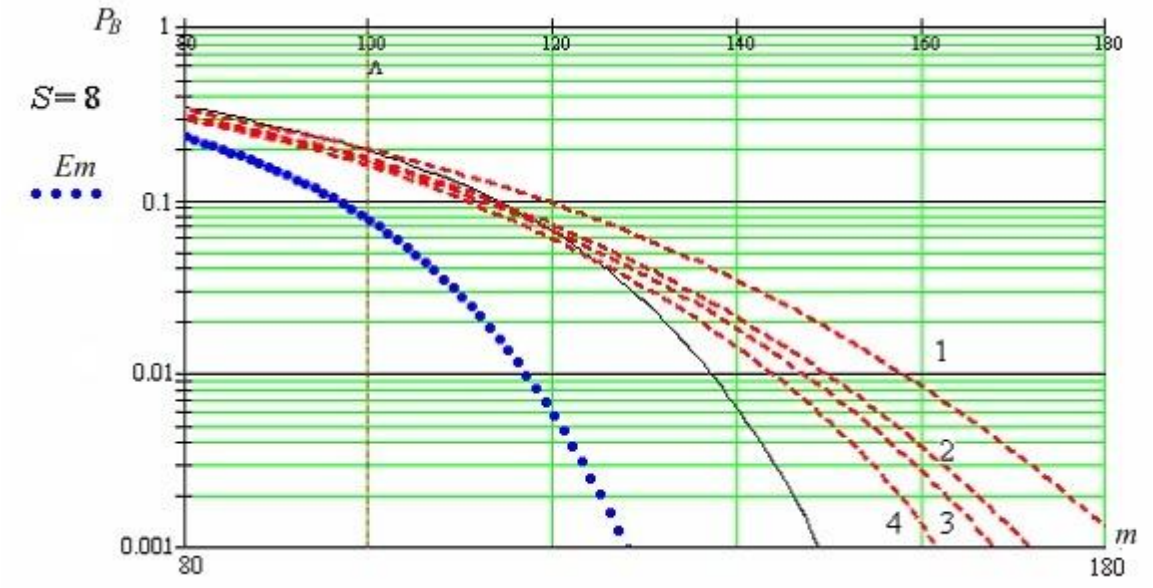


Рисунок 3.4 – Залежність P_B від m та закону розподілу тривалості обслуговування

Пунктирна крива $E_m(\Lambda)$ репрезентує залежність втрат від m, розраховану за B-формулою Ерланга. Штриховими лініями 1, 2, 3 та 4 показано залежність імовірності втрат вимог для регулярного, рівномірного, експонентного та логарифмічно нормального законів розподілу тривалості обслуговування.

Дані графіки демонструють, що, наприклад, за ємності системи $m = 125$ серверів імовірність втрати вимоги, що розрахована за B-формулою Ерланга складе $P_B = 0{,}002$. У той самий час, якщо до системи надходить не пуассонівський потік, а більш нерівномірний потік з пікфактором $S = 8$, то за постійної тривалості обслуговування вимог реальні втрати складуть $P_B = 0{,}07$, що у 35 разів більше. Для того, щоб за цих умов утримати якість обслуговування на заданому рівні $P_B = 0{,}002$, то необхідно у системі мати реально $m = 172$ сервери, а не 125. Саме такими великими є помилки в розрахунках CPI, коли використовуються неадекватні методи, які не враховують характеру вхідного потоку вимог.

З рисунка видно, що за умов значної ($S = 8$) дисперсії інтенсивності навантаження краща якість обслуговування буде там, де в законі розподілу більша частка (імовірність) коротших за тривалістю обслуговування вимог. Коротші за тривалістю обслуговування вимоги скоріше йдуть із системи, звільняючи сервери на використання наступними вимогами.

3.5 Система з необмеженою чергою *H*/*D*/*m*/∞

Повнодоступна схема СРІ з m серверами та необмеженою кількістю місць очікування у черзі обслуговує гіперекспонентний потік вимог з інтенсивністю Λ, пікфактором $S > 1$ та нормальним розподілом кількості вимог за одиницю часу x, де x – постійна тривалість обслуговування вимоги. Вибір вимог із черги в порядку надходження. Необхідно визначити основні характеристики *QoS*:

– імовірність очікування P_w;
– середню кількість вимог у черзі (довжину черги) Q;
– середню тривалість очікування вимог у системі W;
– середню тривалість очікування вимог у черзі t_q.

Розподіл імовірностей P_i випадкової кількості вимог i, що надходять до системи за одиницю часу x, описується нормальним законом розподілу (3.2).

Незалежно від виду потоку вимог Q та W розраховуються за формулами (2.22) та (2.25) на підставі формули Літтла та визначення відповідно. Звідси видно, що дані характеристики *QoS* залежать від P_w та t_q, які визначаються з функцій розподілу станів системи P_j та розподілу часу очікування вимог у черзі $P(t_q)$. Однак для непуассонівського потоку не існує загального методу отримання таких функцій, і формули для них не є простими.

Для розрахунку t_q використаємо такі відомі результати:

– за формулою (2.24) у системі *M*/*M*/*m*/∞ середня тривалість очікування вимог у черзі $t_{q(M)} = 1 / (m - \Lambda)$;
– за формулою Поллачека-Хінчина (табл. 2.1) у системі *M*/*D*/1/∞ середня тривалість очікування вимог у черзі $t_{q(D)} = t_{q(M)} / 2$.

Очевидно, що в шуканому виразі для розрахунку t_q системи *H*/*D*/*m*/∞ мають враховуватися дані наслідки. Перший – тому, що пуассонівський потік (*M*) є окремим випадком гіперекспонентного (*H*). Другий – тому що односерверна система ($m = 1$) є окремим випадком *m*-серверної.

В [9] для системи *H*/*D*/*m*/∞ показано, що $t_{q(D)} > t_{q(M)}$ у $S / (k = 2)$ разів за кількості серверів $m = \Lambda$. Через пікфактор S враховано відмінність гіперекспонентного потоку від пуассонівського та відмінність у два рази середньої тривалості очікування за постійної та експонентної тривалості обслуговування, але в характерній точці, де кількість серверів системи $m = \Lambda$.

Далі при збільшенні m коефіцієнт $k = 2$ спадає зі швидкістю $k(m) \approx (m + \Lambda) / m$. За результатами імітаційного моделювання системи *H*/*D*/*m*/∞ встановлено, що точність розрахунку t_q підвищується при заміні даної залежності на $k(m) \approx (m + \Lambda + 1 + \Lambda / m) / m$. Остаточна формула для розрахунку t_q системи *H*/*D*/*m*/∞ має вид:

$$t_q \approx \frac{1}{m - \Lambda} \cdot S \cdot \frac{m}{m + \Lambda + 1 + \Lambda / m} = \frac{S}{(m+1) \cdot [1 - (\Lambda / m)^2]}. \qquad (3.15)$$

Імовірність P_w – це імовірність зайнятості всіх m серверів і з (2.19) маємо:

$$P_w = 1 - \sum_{k=0}^{m-1} P_k \,. \qquad (3.16)$$

При кінцевому числі m і необмеженій кількості місць очікування вимоги обслуговуються без втрат. На сервери системи надходять вимоги з первинного потоку з інтенсивністю Λ та із черги з інтенсивністю $\Lambda \cdot P_w \cdot t_w = Q$. Тому загальна інтенсивність навантаження на систему збільшується до величини $\Lambda_2 = \Lambda + Q$.

Однак функція розподілу кількості вимог у системі (обслуговуються та у черзі) або станів системи P_j не така, як нормальна функція розподілу P_i кількості вимог, що надходять. На рис. 3.5 показано розподіли станів системи за гіперекспонентного потоку вимог з параметрами $\Lambda = 100$ Ерл та $S = 4$ для ємності системи m = 105, 110 та 120 серверів ($m > \Lambda$).

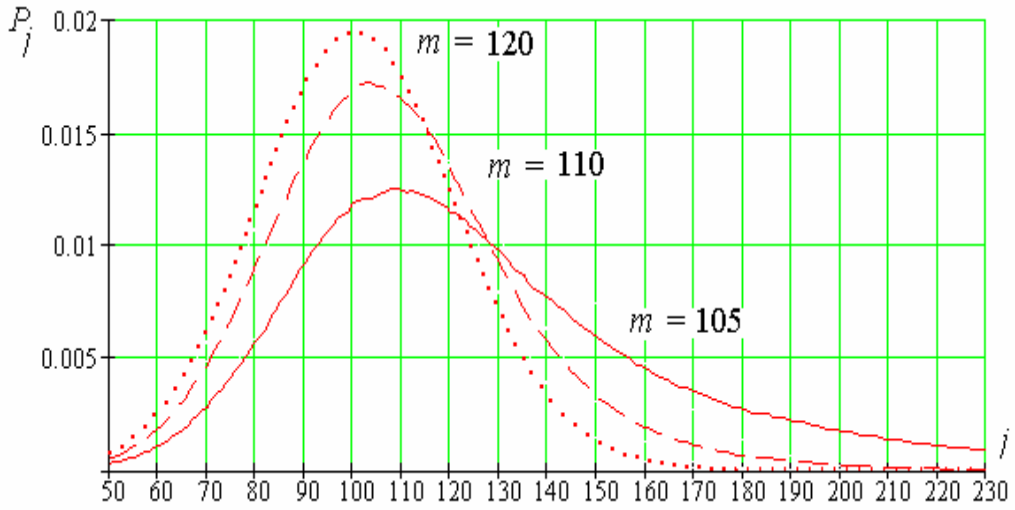


Рисунок 3.5 – Розподіл станів системи *H/D/m/∞*

Порівнянно з m =120 при меншому m =105 розкид окремих значень функції розподілу станів системи від середнього значення збільшується, оскільки додатковий потік вимог із черги збільшує сукупну інтенсивність навантаження Λ_2 та її дисперсію σ_2^2. Але при m = 120 (пунктирна лінія) функція розподілу станів системи доволі симетрична, що дозволяє її апроксимувати нормальним законом розподілу. Апроксимація цих функцій нормальним законом (3.2) з параметрами

$$\Lambda_2 = \Lambda + Q \,; \qquad (3.17)$$

$$\sigma_2 \approx \sigma + \frac{Q}{2} \qquad (3.18)$$

дає достатньо точні результати на лівому відрізку функції розподілу станів системи, обумовленому межами підсумовування у (3.16), тобто від 0 до m – 1.

На рис. 3.6 для даного прикладу виконана апроксимація станів системи нормальним законом розподілу з параметрами $\Lambda_2 = \Lambda + Q$ і $\sigma_2 \approx \sigma + Q/2$ на ділянці, обумовленій межами підсумовування у формулі (3.16).

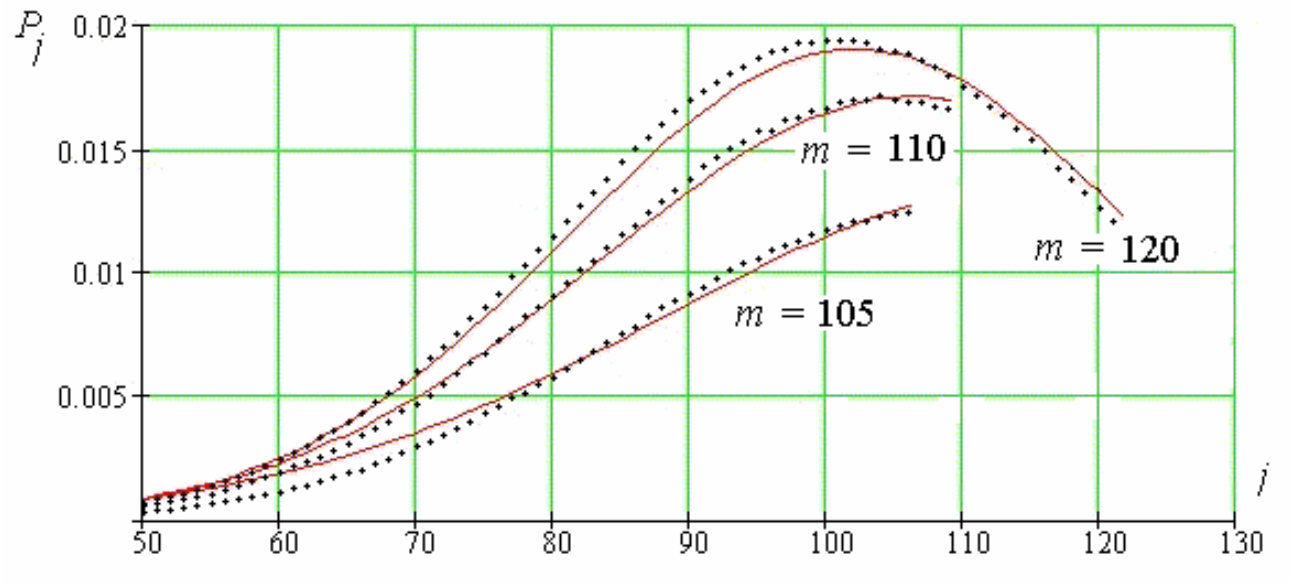


Рисунок 3.6 – Апроксимація функцій розподілу станів системи $H/D/m/\infty$

З цього випливає простий ітераційний алгоритм розрахунку основних характеристик якості обслуговування системи $H/D/m/\infty$:

– із (3.15) для заданих Λ, S та m розраховується t_q;

– із (3.16) та (3.2) для заданих Λ і σ^2 визначається первинна імовірність P_w (нібито для випадку, коли вимоги із черги не йдуть у систему і не збільшують навантаження на неї);

– для розрахованих t_q і P_w відповідно до (2.25) і (2.22) визначаються первинні значення W і Q;

– для розрахованих із (3.17) і (3.18) значень Λ_2 та σ_2 відповідно до (3.16) і (3.2) визначається уточнена імовірність P_w, тобто з урахуванням впливу додаткового навантаження на сервери системи із черги. (При цьому довжина черги більш реальна, оскільки вимоги, що не йдуть із системи негайно, сприяють зростанню черги);

– за уточненим P_w із (2.22) і (2.25) уточнюються значення Q та W.

Імітаційним моделюванням встановлено, що цей алгоритм розрахунку у великому діапазоні варіювання параметрів Λ, S і m дає завжди занижену оцінку імовірності P_w, однак при цьому відносна похибка ніколи не перевищує -10 %. Тому, оскільки на останньому кроці знову уточнюються значення W і Q, то можна ще раз перерахувати P_w з більш точними значеннями Λ_3 і σ_3. Перевірка цього кроку показала, що результати розрахунків після третьої ітерації завжди дають верхню оцінку імовірності P_w, яка не перевищує +10 % [7].

На підтвердження сказаного для прикладу на рис. 2.2 при $m = 105$ відповідно до запропонованого алгоритму зроблено розрахунки, результати яких зведені до табл. 3.1. Тут W і t_q надані в одиницях середньої тривалості обслуговування.

Таблиця 3.1 – Результати імітаційного моделювання

Параметр *QoS*	Моделю-вання	1-я ітерація		2-я ітерація		3-я ітерація	
		розрахунок	помилка	розрахунок	помилка	розрахунок	помилка
P_w	0,71332	0,41050	-42,6%	0,66514	-6,8%	0,74821	4,9%
Q	28, 89636	16,67339	-42,3%	27,01616	-6,5%	30,39005	5,2%
W	0, 28908	0,16681	-42,3%	0,27029	-6,5%	0,30404	5,2%
t_q	0,40526	0,40636	0,3%	0,40636	0,3%	0,40636	0,3%

Як випливає з табл. 3.1 відносна помилка розрахунку Q і W визначається сумарною точністю розрахунку t_q і P_w, і разом з тим помилка розрахунку всіх характеристик *QoS* залишається в межах ±10% (друга і третя ітерації). Розрахунки, що виконані у цьому випадку методом Кроммеліна, дають заниження результату від 50 до 95% (зі зростанням m збільшується похибка).

3.6 Система з необмеженою чергою *GI*/*M*/1/∞

У мультисервісних пакетних мережах зв'язку вхідні інформаційні потоки можуть мати постійну (*CBR*), змінну (*VBR*) і змішану бітову швидкість, від чого математична модель потоку може бути від найпростішої пуассонівської до складної моделі фрактальних процесів (самоподібний трафік). Закон розподілу інтервалу часу між вимогами в цих потоках може бути довільний і тому в узагальненій моделі резонно досліджувати узагальнений (*G*) вид розподілу випадкової величини цього інтервалу. Довжина пакетів кожної зі служб загальної для них мультисервісної мережі (отже, і тривалість обслуговування) може бути різною – для одних служб постійною, а для інших – змінною. У такому випадку також бажано досліджувати загальний вид розподілу випадкової величини тривалості обслуговування.

Системи з чергою типу *GI*/*G*/*m*/∞ є найважливішими моделями, що розглядаються у теорії телетрафіка (*GI* – узагальнений рекурентний потік). При їх дослідженні застосовувалися різні методи й отримані численні наближені результати. Окремий випадок – система з чергою й одним сервером ($m = 1$) розглядається, наприклад в [6], де отримано тільки приблизні результати. Однак дотепер у загальному випадку не існує простих і точних, безпосередньо застосовуваних на практиці, формул для розрахунку характеристик *QoS* у стаціонарному режимі.

У багатьох випадках односерверної системи хорошим наближенням є експонентна функція розподілу тривалості обслуговування вимог. Тому визначимо характеристик *QoS* для моделі *GI*/*M*/1/∞ [12].

Коефіцієнт використання серверів системи ρ (*utilization factor*) визначається як відношення інтенсивності вхідного потоку вимог λ до

інтенсивності обслуговування μ. В *m*-серверній системі всі сервери забезпечать інтенсивність обслуговування $m\mu = m\frac{1}{\bar{x}}$. Отже, в *m*-серверній системі $\rho = \frac{\lambda\bar{x}}{m}$. В односерверній системі ρ в *m* разів більше і збігається з інтенсивністю навантаження $\Lambda = \lambda\bar{x}$ або з інтенсивністю вхідного потоку вимог λ, якщо $\bar{x}$ є однією умовною одиницею часу обслуговування. За умови $0 \le \rho < 1$ процес у системі ергодичний і стаціонарний розподіл імовірностей станів системи існує.

Для будь-якої односерверної системи $\rho = 1 - p_0$, де p_0 – імовірність вільності системи (стан системи p_0 – зайнято 0 серверів). Отже, ρ – чисельно збігається з імовірністю зайнятості системи $P_{зн}$ (стан системи p_1 – зайнятий єдиний сервер, відповідає частці часу зайнятості серверу). З урахуванням вимог, що знаходяться у черзі, у стаціонарному режимі існує стаціонарний розподіл кількості вимог у системі p_k, де *k* – кількість вимог. Цей розподіл не залежить від моменту прибуття вимоги до системи.

За пуассонівського потоку імовірність очікування P_w збігається з імовірністю зайнятості системи $P_{зн}$ (див. (3.16)). Для односерверної моделі *M*/*G*/1/∞ з довільним розподілом тривалості обслуговування дані імовірності однакові і $P_w = \rho$. Однак для моделі *GI*/*M*/1/∞ такої рівності нема, тобто за цим параметром моделі не інваріантні. В [5, с. 272] показано, що система *GI*/*M*/1∞ призводить до геометричного розподілу кількості вимог у системі в моменти надходження нових вимог r_k, де *k* – кількість вимог. Розподіл p_k відрізняється від розподілу r_k тим, що $p_0 = 1 - P_{зн}$ (або $p_0 = 1 - \rho$), у той час як $r_0 = 1 - P_w$. Для системи *M*/*G*/1/∞ виконується рівняння $p_k = r_k$.

Вимога повинна очікувати обслуговування з імовірністю $P_w = 1 - r_0$. Тому за експонентного розподілу тривалості обслуговування безумовний розподіл тривалості очікування визначиться так:

$$W(t) = 1 - P_w e^{-\mu(1-P_w)t} \quad \text{при} \quad t \ge 0. \qquad (3.19)$$

З цього можна розрахувати середній час очікування у системі *W* і всі інші параметри якості обслуговування:

$$P_{зн} = \rho; \qquad P_w = 1 - \frac{\rho}{N}; \qquad W = \frac{P_w}{1 - P_w}; \qquad Q = \frac{\rho \cdot P_w}{1 - P_w};$$

$$t_q = T = \frac{1}{1 - P_w}; \qquad N = \frac{\rho}{1 - P_w}.$$

Оскільки $\mu = 1$, тобто середня тривалість обслуговування $\bar{x} = 1/\mu$ приймається за умовну одиницю часу, то *W*, t_q і *T* оцінюються в одиницях середньої тривалості обслуговування. В табл. 3.2 наведено співвідношення між основними параметрами *QoS* для моделі *GI*/*M*/1/∞.

Таблиця 3.2 – Співвідношення між параметрами *QoS* у системі *GI*/*M*/1/∞

Хар-ка *QoS*	Характеристика *QoS*				
	Q	W	t_q	N	
$P_{зн}$	ρ	ρ	ρ	ρ	
P_w	$\frac{Q}{\rho+Q}$	$\frac{W}{1+W}$	$1-\frac{1}{t_q}$	$1-\frac{\rho}{N}$	$\frac{Q}{N}$
Q	–	$\rho\cdot W$	$\rho\cdot(t_q-1)$	$N-\rho$	$N\cdot P_w$
W	$\frac{Q}{\rho}$	–	t_q-1	$\frac{N}{\rho}-1$	
t_q	$1+\frac{Q}{\rho}$	$1+W$	–	$\frac{N}{\rho}$	
N	$\rho+Q$	$\rho\cdot(1+W)$	$t_q\rho$	–	$\frac{Q}{P_w}$
T	$1+\frac{Q}{\rho}$	$1+W$	t_q	$\frac{N}{\rho}$	

З таблиці випливає, що за наявності лише одного відомого параметра (наприклад, відомий параметр Q, W, t_q або N) всі інші параметри розраховуються через навдані в таблиці співвідношення.

Для моделі *GI*/*M*/1/∞ виявлено й перевірено за допомогою імітаційного моделювання важливі властивості односерверної системи, які виконуються тільки за експонентної тривалості обслуговування (виділені фоном в табл. 3.2).

По-перше, середній час очікування у черзі t_q чисельно збігається із середнім часом перебування вимоги у системі T. Це означає, що середній час очікування у системі W менше середнього часу очікування у черзі t_q на величину середньої тривалості обслуговування (одиниця часу $\bar{x}$):

$$W = t_q - 1. \tag{3.20}$$

По-друге, імовірність очікування можна визначити як

$$P_w = \frac{Q}{N}. \tag{3.21}$$

За визначенням (1.12) імовірність очікування P_w – це відношення кількості затриманих вимог до загальної кількості вимог, але у даному випадку частку затриманих вимог можна визначити і як відношення середньої кількості вимог у черзі Q до середньої кількості вимог у системі N. Звідси з урахуванням формули Літтла випливають кілька важливих співвідношень між параметрами *QoS*, справедливих для моделей *GI*/*G*/1/∞ та *GI*/*G*/*m*/∞ (праворуч у табл. 3.2).

Для системи *GI*/*G*/1/∞ Маршаллом запропонована наближена верхня (*High*) оцінка середнього стаціонарного часу очікування W_H [1]:

$$W_H \le \frac{\sigma_z^2 + \sigma_x^2}{2(\bar{z} - \bar{x})}.$$

Нижня (*Low*) оцінка середнього стаціонарного часу очікування W_L запропонована Штойяном в [6]:

$$W_L \ge \frac{\sigma_x^2}{2(\bar{z} - \bar{x})} - \frac{\bar{x}}{2}.$$

Аналіз системи типу $GI/G/1/\infty$, в якій враховується вид функції розподілу інтервалу часу між вимогами вхідного потоку та тривалості обслуговування вимог, за допомогою наданих наближених верхньої та нижньої оцінок іноді дає більш точний результат порівняно з максимальними оцінками, що отримуються при використанні марковських моделей.

У пакетних мережах зв'язку вимогою на обслуговування можна вважати кожний пакет інформації. Довжина пакетів може бути незмінною за постійної тривалості їх обслуговування, наприклад, в технології *ATM*. У такому разі в дослідженнях можна обмежитися тільки детермінованим законом розподілу тривалості обслуговування (D). Однак і за змінної довжини пакетів при аналізі пакетних комутаторів слід враховувати наявність у кожного пакета заголовку фіксованої довжини, що вимагає врахування в тривалості обслуговування деякого постійного доданку, навіть якщо розподіл довжин пакетів є, наприклад, експонентним. Отже, розробка методу розрахунку основних характеристик *QoS* у системах**,** представлених моделлю $GI/D/1/\infty$ також є актуальною.

У п. 3.6 показано, що для моделі $GI/M/1/\infty$ з геометричним розподілом r_k (розподіл кількості вимог у системі в моменти надходження нових вимог) за експонентного закону розподілу тривалості обслуговування середній час очікування у системі W визначається як

$$W = \frac{P_w}{1 - P_w}. \qquad (3.22)$$

З урахуванням цього результату та формули Літтла в табл. 3.3 надано формули розрахунку характеристик *QoS* для моделі $GI/M/1/\infty$ і відомий результат для моделі $M/D/1/\infty$ за формулою Поллачека-Хінчина (2.45).

У формули розрахунку характеристик *QoS* моделі $GI/G/1/\infty$ можна ввести величину $t_q - W$ (табл. 3.3). Правильність цього кроку легко перевірити підстановленням в кожну з них справедливого за визначенням для будь-якої моделі рівняння:

$$P_w = \frac{W}{t_q}.$$

Для розрахунку характеристик *QoS* моделі $GI/D/1/\infty$ можна використати виявлену вище властивість моделі $GI/M/1/\infty$, а саме: $t_q = T$. Оскільки з визначення середнього часу перебування вимоги в будь-якій системі (в тому числі й в моделі $GI/G/1/\infty$) випливає, що

$$T = W + 1, \qquad (3.23)$$

то з урахуванням наведеної властивості моделі *GI*/*M*/1/∞

$$t_q - W = 1. \qquad (3.24)$$

Видно, що при виконанні умови (3.24), кожна з формул розрахунку характеристик *QoS* моделі *GI*/*G*/1/∞ збігається з аналогічною формулою моделі *GI*/*M*/1/∞. Також відомо, що для моделі *M*/*D*/1/∞:

$$t_q - W = 0{,}5. \qquad (3.25)$$

Тому при виконанні умови (3.25) разом з умовою $P_w = \rho$, властивої за пуассонівського потоку, кожна з формул моделі *GI*/*G*/1/∞ перетвориться у відповідну формулу моделі *M*/*D*/1/∞ (стовпчики 4 та 2 табл. 3.3).

Таблиця 3.3 – Основні параметри якості обслуговування

Хар-ка *QoS*	Модель			
	M/*D*/1/∞	*GI*/*M*/1/∞	*GI*/*G*/1/∞	$GI^*/D/1/\infty$
			точно	*приблизно*
$P_{зн}$	ρ	ρ	$\rho,\ \dfrac{Q}{t_q P_w}$	
P_w	ρ	$1-\dfrac{\rho}{N}$	$\dfrac{W}{t_q},\ \dfrac{Q}{t_q \rho}$	
Q	$\dfrac{\rho^2}{2(1-\rho)}$	$\dfrac{\rho P_w}{1-P_w}$	$\dfrac{\rho P_w}{1-P_w}(t_q - W)$	$\dfrac{\rho P_w}{2(1-P_w)}$
W	$\dfrac{\rho}{2(1-\rho)}$	$\dfrac{P_w}{1-P_w}$	$\dfrac{P_w}{1-P_w}(t_q - W)$	$\dfrac{P_w}{2(1-P_w)}$
t_q	$\dfrac{1}{2(1-\rho)}$	$\dfrac{1}{1-P_w}$	$\dfrac{1}{1-P_w}(t_q - W)$	$\dfrac{1}{2(1-P_w)}$
N	$\rho + \dfrac{\rho^2}{2(1-\rho)}$	$\dfrac{\rho}{1-P_w}$	$\rho + \dfrac{\rho P_w}{1-P_w}(t_q - W)$	$\dfrac{\rho(2-P_w)}{2(1-P_w)}$
T	$1 + \dfrac{\rho}{2(1-\rho)}$	$\dfrac{1}{1-P_w}$	$1 + \dfrac{P_w}{1-P_w}(t_q - W)$	$\dfrac{2-P_w}{2(1-P_w)}$

Якщо припустити, що умова (3.25) здійсненна і при $P_w \neq \rho$, то з тих же формул моделі *GI*/*G*/1/∞ можна отримати відповідні формули розрахунку характеристик *QoS* моделі *GI*/*D*/1/∞ (стовпчик 5, табл. 3.3). Але оскільки, як показує імітаційне моделювання, отримані у такий спосіб формули дають не зовсім точні і різні результати для різних вхідних потоків, то краще назвати таку модель з *умовно загальним* розподілом випадкової величини інтервалу часу між вимогами і позначити $GI^*/D/1/\infty$.

Для моделі $GI^*/D/1/\infty$ різниця (3.25) не завжди дорівнює 0,5. За результатами імітаційного моделювання [7], встановлено, що для потоків, у яких коефіцієнт варіації тривалості інтервалу часу між вимогами $v_z < 1$ (більш

вирівняний потік порівняно з пуассонівським) $t_q - W < 0{,}5$. Наприклад, за рівномірного розподілу інтервалу часу між вимогами (модель $U/D/1/\infty$) $v_z = 0{,}58$. Для пуассонівського потоку з експонентним розподілом зазначеного інтервалу при $v_z \equiv 1$ або з розподілом Вейбулла при підібраному коефіцієнті форми розподілу так, щоб $v_z = 1$, умова (3.24) виконується. Для гіперекспонентного, логарифмічно нормального або розподілів Парето та Вейбулла при параметрах розподілів, що забезпечують значення v_z від одиниць до декількох десятків (дуже нерівномірні потоки), різниця t_q і W завжди трохи більше 0,5.

У випадку $t_q - W < 0{,}5$ результати розрахунків трохи вище результатів моделювання, при $t_q - W = 0{,}5$ – вони цілком збігаються, а при $t_q - W > 0{,}5$ – результати розрахунків виявляються трохи заниженими. При цьому найбільша похибка обчислень за формулами моделі $GI^*/D/1/\infty$ порівняно з результатами моделювання спостерігається у випадку самоподібного потоку вимог, згенерованому за розподілом Парето. Наприклад, при завданні параметра форми розподілу $a = 1{,}3$ (відповідає коефіцієнтові самоподібності трафіка $H = 0{,}85$ [11] і $v_z \geq 30$) похибка обчислень не перевищує 20 % для всіх параметрів *QoS*. При збільшенні a і пов'язаним з цим зменшенням H, наприклад, до значень 1,9 і 0,55 відповідно (ще самоподібний процес), дана похибка не перевищує вже 10 %.

Слід зазначити, що точність розрахунків декілька підвищується при зменшенні інтенсивності навантаження ρ і в багатьох випадках при $\rho < 0{,}5$ похибка розрахунків не перевищує 5 %.

Важливість даного методу мотивується суттєвим ускладненням моделей трафіка у сучасних мережах зв'язку і відсутністю адекватних цьому ускладненню методів розрахунку параметрів якості обслуговування. Корисним цей метод є у випадку обслуговування самоподібного трафіка, оскільки його точність незрівнянно вище точності існуючих методів розрахунку.

4 МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА ДЛЯ ПАКЕТНИХ МЕРЕЖ ЗВ’ЯЗКУ

4.1 Модель трафіка пакетних мереж зв’язку

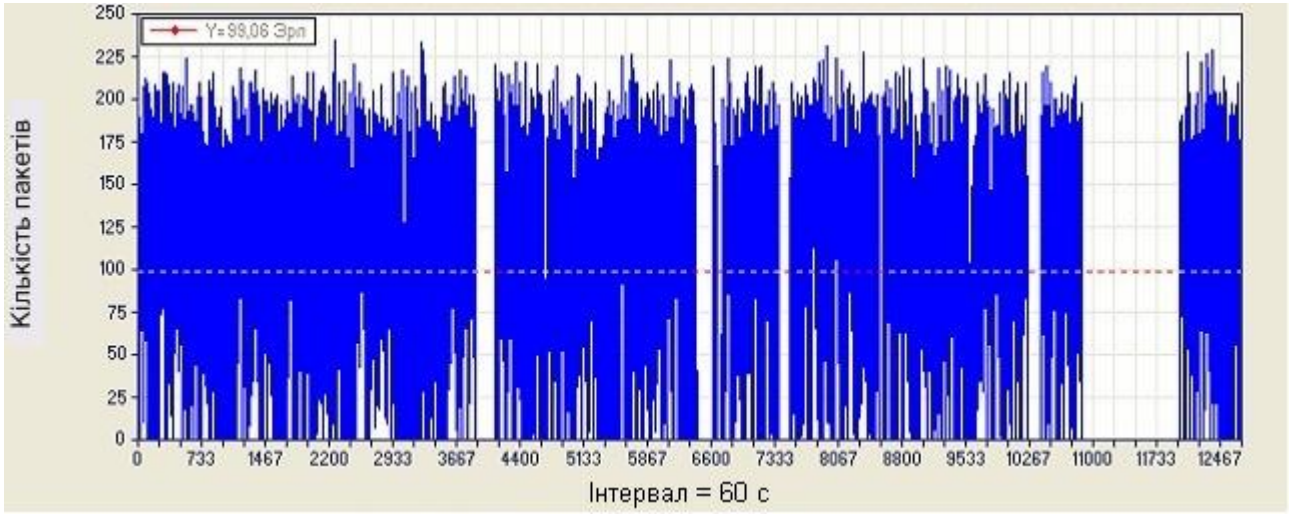
У пакетних мережах зв'язку потоки пакетів (трафік) істотно відрізняються від моделі пуассонівського потоку. Пакетний трафік формується множиною джерел запитів на надавані мережею послуги та мережними додатками, що забезпечують послуги передавання відео, даних, мови та ін. Джерела вимог, беручи участь у процесі створення потоку пакетів, істотно відрізняються між собою значеннями питомої інтенсивності навантаження. Інтенсивність навантаження результуючого потоку пакетів у кожний момент часу залежить від того, якими додатками обслуговуються джерела вимог і яке співвідношення їхньої кількості для різних додатків.

На структуру трафіка впливають і технологічні особливості алгоритмів обслуговування. Наприклад, якщо послуга забезпечується декількома додатками або якщо у використовуваних протоколах застосовується повторна передача невірно прийнятих пакетів, то моменти виникнення вимог на встановлення сеансів зв'язку сильно корелюються. Через це в процесі обслуговування вхідні потоки значно змінюються й у сумарному трафіку з'являються довгострокові залежності в інтенсивності надходження пакетів. Отже, пакетний трафік уже не є простою сумою множини незалежних стаціонарних і ординарних потоків, що властиво пуассонівським потокам.

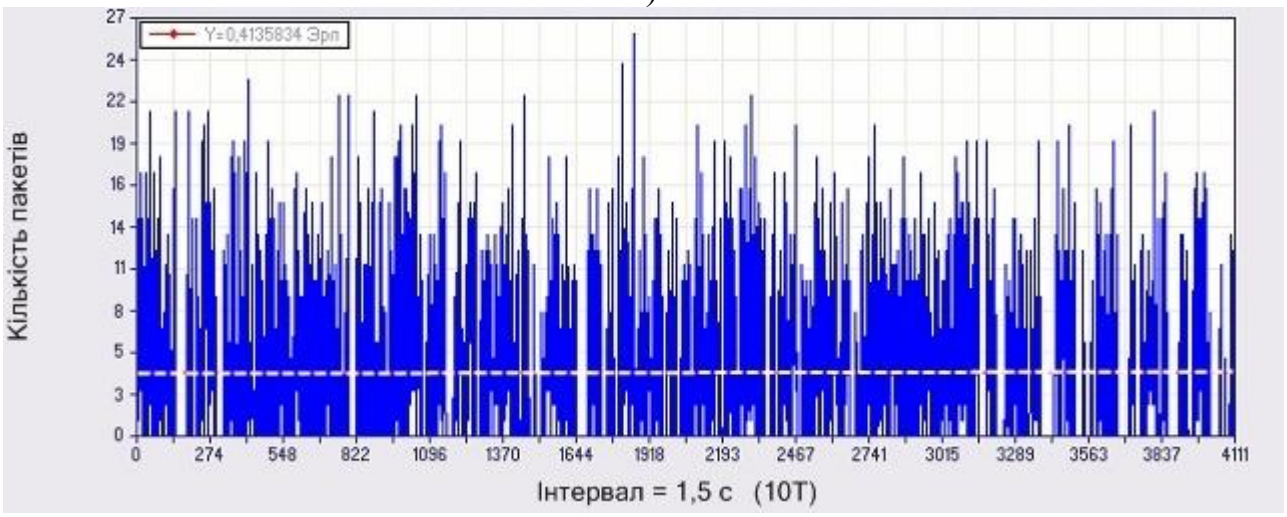
У мережах з комутацією пакетів трафік є різнорідним, а потоки різних додатків вимагають забезпечення певного рівня якості обслуговування. У цих умовах передачу потоків усіх додатків забезпечує єдина пакетна мережа з загальними протоколами і законами керування, незважаючи на те, що джерела кожного додатка мають різні швидкості передавання інформації або змінюють її в процесі сеансу зв'язку (максимальна і середня швидкості). Через це об'єднаному потоку пакетів властиве так зване «пачкування» трафіка (*burstness*) з випадковою періодичністю та тривалістю піків навантаження, вимірювана коефіцієнтом пачкування [3]. Наприклад, пачкування для мовленнєвих служб можливе через паузи в розмові. Це пачкування обумовлює ще більшу нерівномірність трафіка, за якої дисперсія інтенсивності трафіка перевищує її математичне сподівання від 15 до 60 разів і більше.

На підставі статистичних даних про кількість і розміри переданих пакетів, характеристик інтервалів часу між пакетами протягом встановленого з'єднання (сеансу зв'язку), а також даних про тривалість встановлюваних з'єднань і т.ін. можна скласти математичну модель пакетного трафіка.

На рис. 4.1 показані результати вимірів параметрів трафіка пакетної мережі зв’язку. Тут показана кількість пакетів у сеансах зв'язку по фіксованих інтервалах часу 60 і 1,5 с (понад 12 000 і 4 000 інтервалів на рис. 4.1,а та 4.1,б відповідно).



а)



б)

Рисунок 4.1 – Статистичні результати вимірів пакетного трафіка

З рис. 4.1 видно, що пакетному трафіку властива сильна нерівномірність інтенсивності пакетів. Пакети не плавно розосереджені по різних інтервалах часу, а групуються в «пачки» в одних, і цілком відсутні або їх дуже мало в інших. Тому в пачковому трафіку за порівняно невеликого середнього значення інтенсивності пакетів присутні значні викиди. Наприклад, для трафіка рис. 4.1,б середнє значення складає 4,1 пакети за 1,5 с (показано штриховою лінією), а окремі викиди досягають значень до 20 і більше пакетів за інтервал. Згідно з «правилом трьох сигма» для пуассонівського потоку з імовірністю $P = 0{,}99$ для цього випадку викиди могли б досягати тільки 10 пакетів. На рис. 4.1,а за середнього значення 99 пакетів за 60 с є викиди до 200 і більше, а для пуассонівського потоку вони могли б досягати лише 130.

Ефективність обслуговування такого трафіка дуже низька, оскільки в процесі його оброблення для забезпечення заданого рівня втрат (якості обслуговування) необхідно збільшувати пропускну здатність каналів. При цьому в періоди спаду піків навантаження ресурси системи дуже сильно недовикористовуються, а, як правило, проектування пропускної здатності каналів відбувається в розрахунку на середнє значення інтенсивності трафіка.

Результати досліджень [13] показали, що функція розподілу інтервалів часу між моментами надходження пакетів має вид, показаний на рис. 4.2.

Рисунок 4.2 – Апроксимація інтервалу часу між пакетами

Графіки на рис. 4.2 підтверджують, що результати статистичної обробки даних вимірів краще узгоджуються не з функцією експонентного розподілу (на графіку пряма лінія через логарифмічну шкалу), а з функціями, що мають так званий «довгий хвіст», більш вагомий, ніж за експонентного розподілу. При цьому гістограми реальних вимірів добре апроксимуються розподілами Парето та Вейбулла, іноді функцією логарифмічно нормального закону розподілу. У деяких випадках для цієї апроксимації можна використовувати гамма-розподіл або гіперекспонентний (іноді достатньо другого порядку):

$$P(z) = p_1\lambda_1 e^{-\lambda_1 z} + p_2\lambda_2 e^{-\lambda_2 z}, \qquad (4.1)$$

де з імовірністю p_1 інтервал часу між пакетами має експонентний розподіл з параметром λ_1, а з імовірністю p_2 – з параметром λ_2 (природно, що $p_1 + p_2 = 1$).

«Довгий хвіст» означає, що, наприклад, за експонентного розподілу ймовірність існування інтервалу часу між вимогами тривалістю 2500 мс дорівнює $P_{2500} = 1{,}4\cdot10^{-13}$, а за логарифмічно нормального – $P_{2500} = 8{,}3\cdot10^{-7}$, тобто значно більше. Отже, дуже довгі інтервали пауз в процесі надходження вимог існують й їхня частка велика. Розподіли з «довгим хвостом» дозволяють обґрунтувати пачкування трафіка – якщо є тривалі інтервали часу, на яких немає жодної вимоги, то щоб «витримати» загальну середню кількість вимог (інтенсивність трафіка), на інших інтервалах часу ці вимоги «компенсуються» надто значною їх кількістю.

Як показали дослідження, для потоків трафіка з розподілом інтервалів часу між пакетами з логарифмічно нормальним законом і законам Парето або Вейбулла, функція розподілу кількості пакетів за одиницю часу (інтенсивність) має явно виражену позитивну асиметрію, що і показано на рис. 4.3.

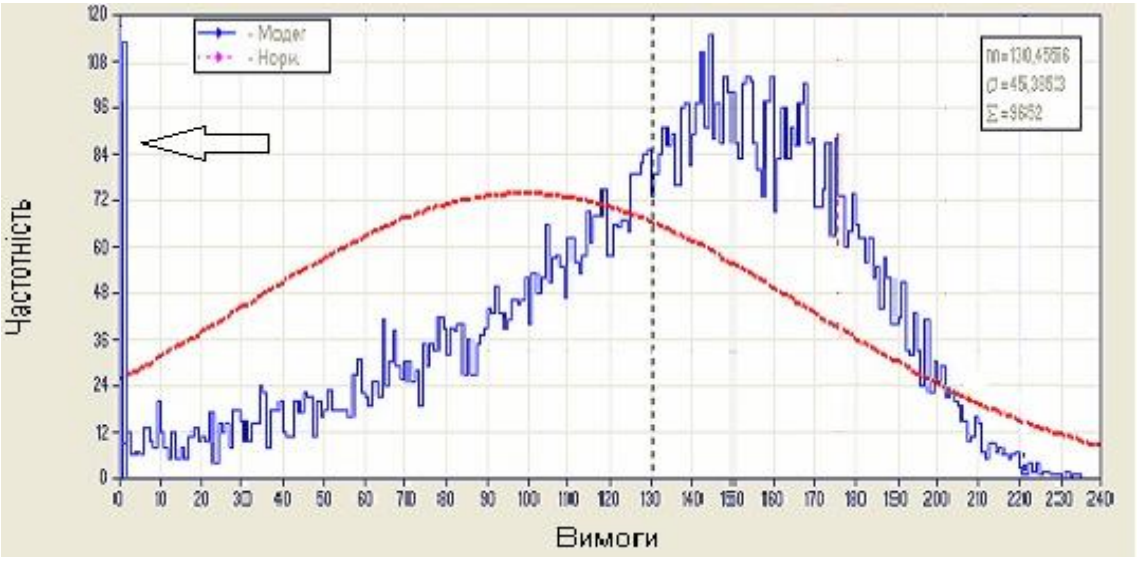


Рисунок 4.3 – Розподіл кількості пакетів за інтервал часу

З рис. 4.3 за інтенсивності трафіка $\Lambda = 130$ Ерл випливає, що апроксимація даної статистики нормальним законом розподілу (штрихова лінія) з математичним сподіванням $\Lambda = 130$ і $\sigma^2 = 2025$ не узгоджується з реальним розподілом трафіка. Тут пікфактор інтенсивності трафіка $S = \sigma^2 / \Lambda = 15{,}6$. При цьому реальний розподіл має істотну асиметрію порівняно з нормальним законом і великою часткою (частотністю) інтервалів, на які випало «нуль вимог» – на них немає жодної вимоги (пакета). Саме ця ймовірність «нульового стану» або нульової кількості вимог (позначено стрілкою) і приводить до настільки значної асиметрії розподілу в цілому.

Трафік, якому властиве «пачкування», прийнято вважати трафіком з ефектом самоподібності. Для пакетних мереж зв'язку математична модель самоподібного трафіка є найбільш популярною, але через відсутність строгої теоретичної бази, здатної доповнити класичну теорію масового обслуговування при проектуванні пакетної мережі з самоподібним трафіком, не існує достовірної та визнаної методики розрахунку параметрів і характеристик якості систем розподілу інформації в умовах обслуговування такого трафіка. Зі зростанням ступеня самоподібності пакетного трафіка характеристики *QoS* у системі значно погіршуються порівняно з обслуговуванням трафіка аналогічної інтенсивності, але без ефекту самоподібності.

Метод формування самоподібного потоку [1] заснований на суперпозиції декількох незалежних з однаковим розподілом ON/OFF джерел, де інтервали між ON та OFF періодами мають ефект Ноа. Ефект Ноа в розподілі тривалості ON/OFF періодів є базовим при моделюванні самоподібного трафіка та є синонімом синдрому нескінченної дисперсії. Для досягнення ефекту Ноа використовують розподіли Парето або Вейбулла, часто називані «розподілами з довгим хвостом». Наявність у розподілі «довгого хвоста» забезпечує властивість пачкового трафіка, тому що в розподілі зростають ймовірності довгих інтервалів між подіями, а також для їх компенсації зростають й імовірності надто коротких інтервалів.

Густина розподілу Парето задається функцією:

$$f(x) = \frac{a}{b}\left(\frac{b}{x}\right)^{a+1}, \qquad (4.2)$$

де a – параметр форми (*shape*); b – мода розподілу (мінімальне значення випадкової величини x). Причому, при $a \le 2$ дисперсія нескінченна, що і необхідно як однієї з умов самоподібності.

При практичному моделюванні самоподібного трафіка розподіл Парето отримаємо при переході від рівномірного розподілу методом зворотної функції:

$$Z_i = \frac{b}{\sqrt[a]{U_i}}, \qquad (4.3)$$

де Z_i – i-й інтервал між заявками потоку; U_i – випадкове число, рівномірно розподілене на інтервалі 0 … 1. Для забезпечення самоподібних властивостей модельованого трафіка необхідно задавати значення параметра форми a в межах від 2 до 1, що має забезпечувати значення коефіцієнта самодібності Херста в діапазоні H = 0,5…1 відповідно. Крім того, необхідно кожне отримане значення інтервалу часу до наступної вимоги Z_i зменшити на величину b для забезпечення реалістичності трафіка (розподіл Парето не дає величини, менші за b, але реально такі значення можуть бути).

Густина розподілу Вейбулла задається функцією:

$$\lambda_0 a x^{a-1} e^{-\lambda_0 x^a}, \qquad (4.4)$$

де a – параметр форми; λ_0 – коефіцієнт масштабу.

При практичному моделюванні самоподібного трафіка розподіл Вейбулла отримаємо при переході від рівномірного розподілу методом зворотної функції:

$$Z_i = \left(-\frac{1}{\lambda_0}\ln U_i\right)^{1/a}, \qquad (4.5)$$

де Z_i – i-й інтервал між заявками потоку; U_i – випадкове число, рівномірно розподілене на інтервалі 0 … 1. Для забезпечення самоподібних властивостей модельованого трафіка необхідно задавати значення параметра форми a в межах від 1 до 0, що має забезпечувати значення коефіцієнта самодобності Херста в діапазоні H = 0,5 ... 1 відповідно.

Для реальних потоків трафіка, що не мають властивості самоподібності, H < 0,5, а для самоподібних потоків з довгостроковою залежністю цей параметр змінюється в межах 0,65 … 0,9 (процес має тривалу пам'ять).

Для моделей самодібного трафіка при використанні вищезгаданих розподілів коефіцієнт варіації інтервалу часу між вимогами v_z істотно зростає. Також збільшується й коефіцієнт скупчення навантаження або пікфактор трафіка S > 15. При цьому значення обох параметрів пропорційні показнику самоподібності Херста. Хоча для випадку мультисервісного трафіка з пік-фактором S > 1 в розд. 3 запропоновано методи розрахунку характеристик якості обслуговування, проте, як показали результати дослідження, у випадку пакетного трафіка ці рішення не підходять.

4.2 Система з необмеженою чергою *fBM*/*D*/1/∞

Трафік мереж зв'язку з комутацією пакетів характеризується наявністю довгострокових залежностей в інтенсивності навантаження й істотною відмінністю статистичних властивостей потоків пакетів від пуассонівського потоку. Адекватною моделлю потоків в таких мережах вважаються самоподібні (*self-similarity*) процеси, де вхідний потік описується фрактальним броунівським рухом (*fBM – fractal Brownian Motion*). Протое дослідження характеристик якості обслуговування СРІ в цих умовах є дуже складною математичною задачею. Причиною цьому є слабка формалізованість моделі самоподібних потоків, внаслідок чого й неможливо отримати аналітично обґрунтовані результати для оцінки параметрів *QoS* у СРІ. Символи *fBM* використовуються для позначення узагальненої моделі трафіка за будь-якого закону розподілу інтервалу *Z*, що призводить до саподібного процесу.

Для односерверної системи з нескінченною чергою та постійним часом обслуговування (модель *fBM*/*D*/1/∞) наближене рішення наведено в [1] у вигляді так званої формули Норроса, де показано, що у стані статистичної рівноваги при $\rho = \lambda / \mu < 1$ середня кількість вимог у системі

$$N = \frac{\rho^{\frac{1}{2(1-H)}}}{(1-\rho)^{\frac{H}{1-H}}}. \qquad (4.6)$$

Оскільки з формули Літтла випливає, що $T = N / \rho$, то середній час перебування вимоги у системі при $\mu = 1$, тобто в одиницях часу тривалості обслуговування, визначається формулою:

$$T = \frac{\rho^{\frac{H-0.5}{1-H}}}{(1-\rho)^{\frac{H}{1-H}}} \qquad (4.7)$$

Виходячи з того, що для будь-якої односерверної системи середня довжина черги $Q = N - \rho$, то з урахуванням (4.6) отримаємо:

$$Q = \frac{\rho^{\frac{1}{2(1-H)}}}{(1-\rho)^{\frac{H}{1-H}}} - \rho. \qquad (4.8)$$

Формули (4.6), (4.7) і (4.8) вважаються аналітичним розв'язанням для системи *fBM*/*D*/1/∞. Проте їх аналіз показує, що за показника Херста $H = 0{,}5$ (не самоподібний процес) маємо відомий результат для середньої кількості вимог, середньої тривалості перебування і середньої довжини черги у системі *M*/*M*/1/∞. Це нелогічно, оскільки спочатку розглядалася система *fBM*/*D*/1/∞. При зміні коефіцієнта Херста від значення $H = 1$ до мінімального $H = 0{,}5$ потік вимог і його відповідна імовірнісна функція розподілу інтервалу часу між вимогами змінюється, але не змінюється функція розподілу тривалості обслуговування. При $H = 0{,}5$ потік втрачає властивості самоподібності, але

формули (4.6 … 4.8) мають корелюватися з результатами для моделі з постійною тривалістю обслуговування, а не експонентною [10].

З показаного на рис. 4.4 графіка залежності середнього часу перебування вимог у системі T випливає, що при навантаженні $\rho > 0{,}3$ система $fBM/D/1/\infty$ із самоподібним вхідним потоком вимог буде витрачати на оброблення більше часу, ніж за відсутності самоподібності, тобто ніж системи $M/M/1/\infty$ і $M/D/1/\infty$.

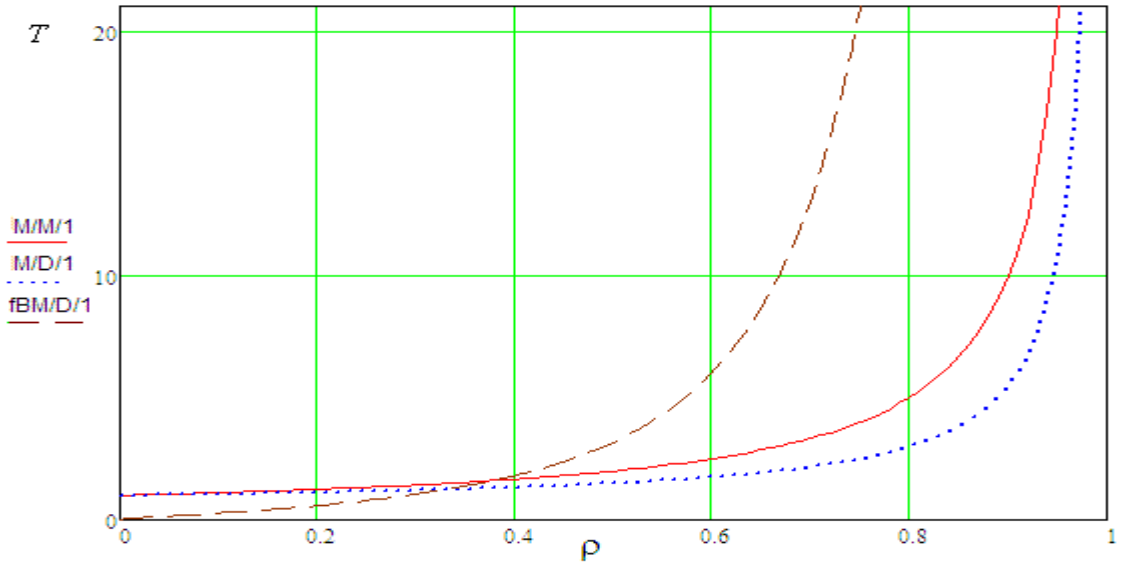


Рисунок 4.4 – Залежність середнього часу перебування вимог у системі T для моделей $M/M/1/\infty$, $M/D/1/\infty$ та $fBM/D/1/\infty$ при $H = 0{,}7$

З приведеного на рис. 4.5 графіка залежності N від ρ для систем $M/M/1/\infty$, $M/D/1/\infty$ і $fBM/D/1/\infty$ маємо аналогічний висновок, що при навантаженні $\rho > 0{,}3$ у системі із вхідним потоком вимог, що має характер самоподібного процесу, буде більше вимог, ніж за відсутності самоподібності.

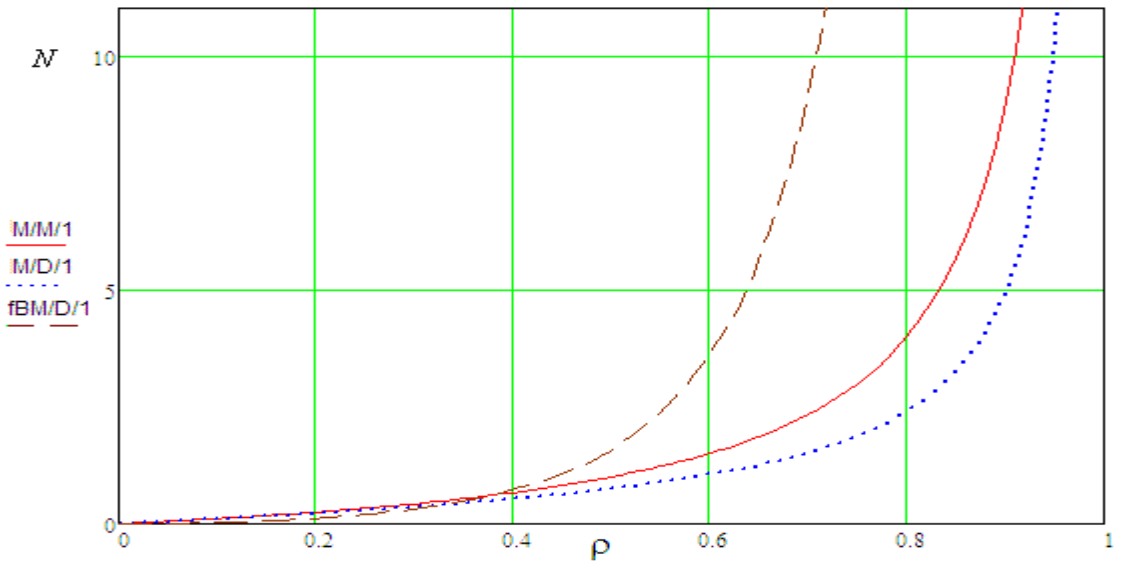


Рисунок 4.5 – Залежність середньої кількості вимог у системі N для моделей $M/M/1/\infty$, $M/D/1/\infty$ та $fBM/D/1/\infty$ при $H = 0{,}7$

До подібного висновку можна дійти й після аналізу графіків залежності середньої довжини черги Q від ρ у тих самих системах, показаних на рис. 4.6.

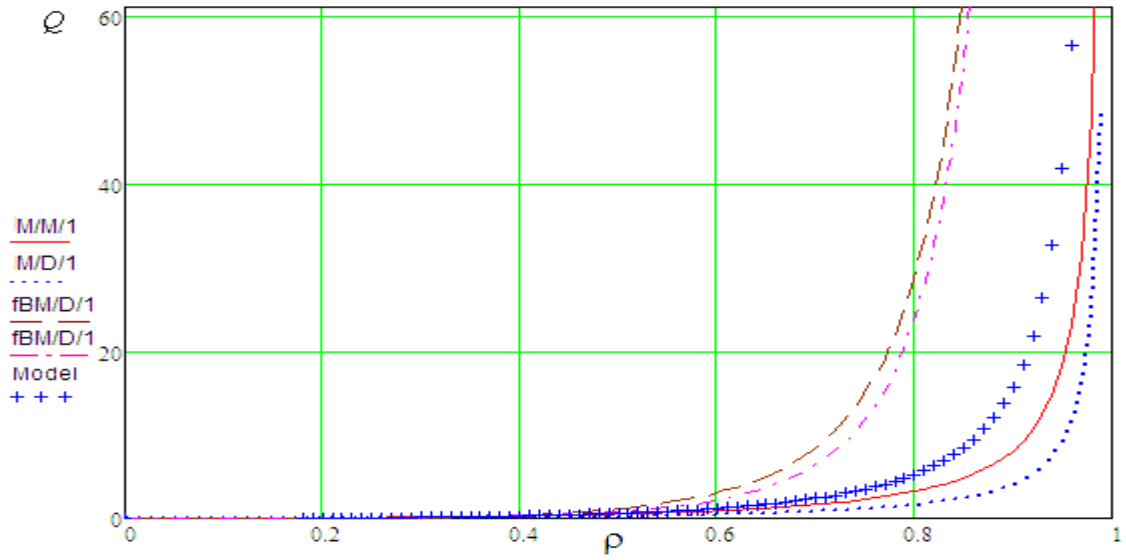


Рисунок 4.6 – Залежність середньої довжини черги Q для моделей M/M/1/∞, M/D/1/∞ та fBM/D/1/∞ при H = 0,7

З усіх графіків на рис. 4.4 … 4.6 випливає, що за пуассонівського потоку зі зростанням інтенсивності навантаження ρ погіршуються характеристики якості обслуговування T, N і Q, але ще більш істотно вони погіршуються за наявності властивостей самоподібності у вхідному потоці вимог. Формули (4.6 – 4.8) показують ступінь цього впливу залежно від показника H.

Проте, формули Норроса (4.6 … 4.8), що отримані в припущенні постійної тривалості обслуговування, при H = 0,5 дають оцінки характеристик QoS, які властиві СРІ за пуассонівського потоку з експонентним законом розподілу тривалості обслуговування, а не регулярного (модель M/M/1/∞).

При імітаційному моделюванні достатньо оцінити тільки один з параметрів, наприклад N, оскільки Q і T пов'язані з N відомими співвідношеннями. Результати імітаційного моделювання СРІ типу fBM/D/1/∞ при H = 0,7 показані на рис. 4.6 лінією, у вигляді знаків „+”.

Результати моделювання підтверджують висновок про те, що за наявності властивості самоподібності у вхідному потоці вимог зі зростанням інтенсивності навантаження ρ погіршуються характеристики якості обслуговування, але не настільки, як це передбачається за формулою Норроса. Розбіжність результатів моделювання й оцінок, отриманих за формулами (4.6 … 4.8), може становити сотні відсотків [10]. Очевидно, що оцінка Норроса значно завищена, а це вимагає знаходження більше точного розв’язання.

У зв'язку із цим, найбільш перспективним є метод оцінки параметрів якості обслуговування самоподібного трафіка, у якому запропоновано використання методів розрахунку відомих класичних розподілів, ентропія яких найбільш близька до ентропії станів системи за умов обслуговування самоподібного трафіка [11]. При цьому стає можливим розрахунок

характеристик *QoS* у моделях з самоподібним трафіком за будь-якого закону розподілу тривалості обслуговування за формулою Поллачека-Хінчина.

Випадковий процес (ВП) надходження пакетів до СРІ на обслуговування утворює потік пакетів (трафік), який характеризується певним законом розподілу, що встановлює зв'язок між значенням випадкової величини (кількістю пакетів) й імовірністю появи цього значення. Часто для розрахунку параметрів *QoS* досить знати про закон розподілу тільки деякі його числові характеристики – моменти розподілу різних порядків. Для розрахунку в умовах пуассонівського розподілу достатньо математичного сподівання Λ, а для нормального розподілу – необхідні математичне сподівання Λ і дисперсія D.

Основні характеристики випадкового процесу Λ і D, хоча й досить важливі, у той самий час не є вичерпними, а іноді й марними для прогнозування значення випадкової величини. Іноді ВП характеризуються однаковими значеннями Λ і D, але внутрішня структура цих процесів різна. Одні можуть мати плавно мінливі реалізації, а інші – яскраво виражену коливальну структуру за стрибкоподібної зміни окремих значень випадкової величини (наприклад, різке зростання кількості пакетів у мережі, що призводить до «пачкування» трафіка). Для «плавних» процесів характерна велика передбачуваність реалізацій, а для «пачкових» – дуже мала імовірнісна залежність між двома випадковими величинами ВП. У таких випадках закон розподілу, що характеризує ВП, несе у собі деяку невизначеність і дозволяє з більшою або меншою надійністю прогнозувати значення випадкової величини.

Таким чином, використовувані ймовірнісні закони розподілу, що описують трафік у пакетних мережах, не дають такої кількісної оцінки невизначеності стану системи розподілу інформації, як ентропія розподілу:

$$\mathbf{H} = -\sum_{j=1}^{m} P_j \log P_j . \qquad (4.9)$$

Ентропія не залежить від значень, яких набуває випадкова величина, а тільки від їхніх імовірностей.

Оцінка параметрів якості обслуговування самоподібного трафіка можлива ентропійним методом, який полягає у використанні методів розрахунку відомих моделей СРІ, ентропія станів яких співпадає або найбільш близька до ентропії станів системи при обслуговуванні самоподібного трафіка [11].

Результатами моделювання встановлено, що в тих точках графіків залежності ентропії станів системи від завантаженості ρ, де збігається ентропія розподілу станів різних систем, збігаються й досліджувані параметри якості обслуговування, такі як середня довжина черги Q та середня тривалість очікування вимог W. Наприклад, з рис. 4.7 випливає, що для моделей $M/M/1/\infty$ і $fBM/D/1/\infty$ при коефіцієнті Херста $H = 0{,}8$ та $\rho = 0{,}6$ ентропії розподілів станів системи досить близькі і дорівнюють 1,683 та 1,719 відповідно. При цьому для моделі $fBM/D/1/\infty$ середня довжина черги $Q = 0{,}982$ й середня тривалість очікування всіх вимог $W = 1{,}611$, що перевищує відповідні значення для моделі $M/M/1/\infty$ усього на 3 % (настільки ж відмінність і значень ентропії). Такий самий збіг основних параметрів якості обслуговування СРІ з чергою

спостерігається в усіх інших точках, для яких однакові значення ентропії розподілу станів системи, незалежно від закону розподілу тривалості обслуговування. Тому для цих розрахунків можна застосовувати формули Поллачека-Хінчина з табл. 2.1 моделі $M/G/1/\infty$.

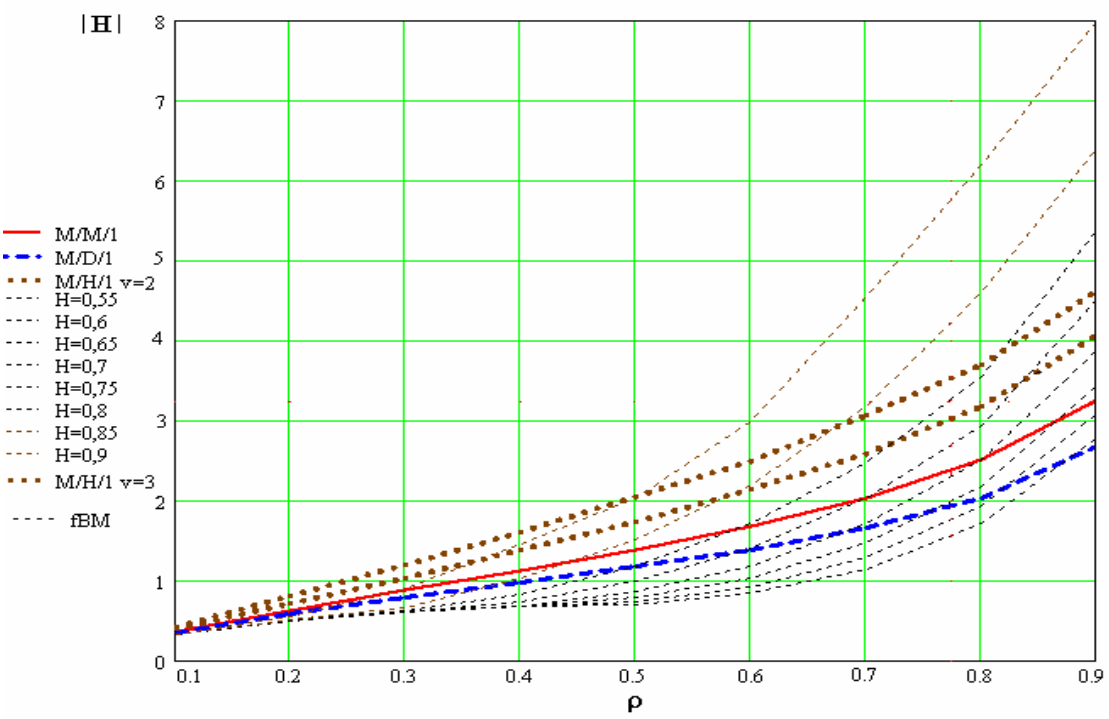


Рисунок 4.7 – Залежність ентропії розподілу станів системи **H** від завантаженості системи ρ

Алгоритм застосування ентропійного методу розрахунку характеристик *QoS* такий:

1. Для певного значення ρ за встановленим законом розподілу станів системи з самоподібним трафіком $fBM/G/1/\infty$ за відомими формулами, що подаються в довідниках, визначається ентропія цього розподілу $\mathbf{H}_{fBM}$.

2. Зміною коефіцієнта варіації v_x для моделі, наприклад $M/H/1/\infty$, досягається збіг значень ентропії $\mathbf{H}_{M/H/1} = \mathbf{H}_{fBM}$.

3. За допомогою знайденого коефіцієнта варіації v_x визначається середня кількість пакетів у системі N за формулою Поллачека-Хінчина (табл. 2.1).

4. Через співвідношення, які надані в табл. 3.2, через N визначаються й інші характеристики *QoS*.

Таким чином, розрахунок параметрів *QoS* в моделі з самоподібним трафіком за будь-якого (G) закону розподілу тривалості обслуговування здійснимий. Необхідною умовою такого розрахунку є визначення ентропії розподілу станів системи із самоподібним трафіком.

4.3 Показник Херста трафіка з розподілами Парето та Вейбулла

Модель самоподібного трафіка широко застосовується для пакетних мереж зв'язку, однак достовірність оцінок якості обслуговування є сумнівною. Результати моделювання (п. 4.2) показують, що оцінки Норроса сильно завищені. Однією з причин цього може бути невідповідність визначеного показника Херста, на якому й основана формула, реальному коефіцієнту самоподібності генерованого при моделюванні трафіка. Пачковий характер генерованого трафіка сприяє його адекватності реальному характеру трафіка в мультисервісних пакетних мережах зв'язку.

Підвищити точність розрахунку характеристик якості обслуговування пакетного трафіка можна шляхом отримання уточненої формули розрахунку коефіцієнта самоподібності трафіка в залежності від параметрів конкретно обраного розподілу самоподібного трафіка.

Для опису інтервалу часу між вимогами або пакетами самоподібного пачкового трафіка часто обираються розподіли Парето або Вейбулла, називаними «розподілами з довгим хвостом» (п. 4.1).

Розподіл Парето. Параметр форми *a* (*shape parameter*) розподілу Парето і показник Херста *H* прийнято вважати [1], що перебувають у такій залежності:

$$H = \frac{3-a}{2}. \tag{4.10}$$

Проте, результати моделювання на рис. 4.8 показують, що для розподілу Парето нема лінійної залежності (4.10) показника Херста *H* від параметра форми *a* розподілу.

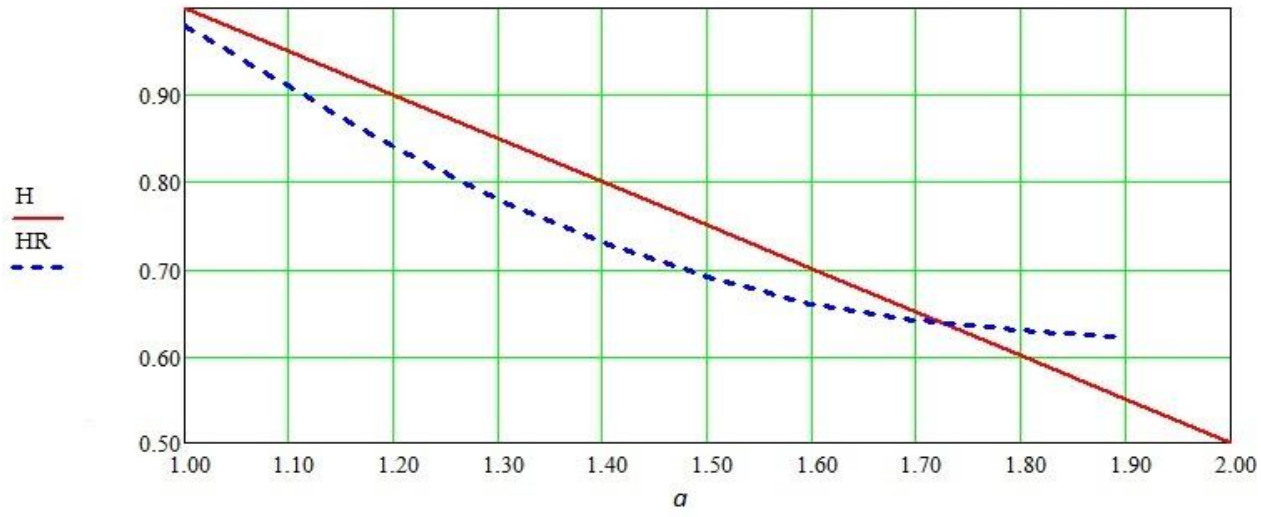


Рисунок 4.8 – Моделювання коефіцієнта самоподібності *H*

Із рис. 4.8 випливає, що реальний показник Херста HR (пунктирна крива) залежить від параметра форми *a* розподілу Парето не лінійно (суцільна лінія), а за законом, близьким до експонентного.

Таким чином, якщо реальна статистика трафіка (інтервал часу між пакетами) апроксимується розподілом Парето, то для розрахунку характеристик якості обслуговування за формулою Норроса (4.6) необхідно розраховувати коефіцієнт самоподібності Херста не за формулою (4.10), а за формулою апроксимуючої кривої HR, що показана на рис. 4.8 штриховою лінією. У цьому випадку точність розрахунку зростає на порядок і вже похибка не перевищує 10…20 % [14].

За результатами імітаційного моделювання для розрахунку показника Херста самоподібного трафіка, описуваного розподілом Парето (позначено *Pa*), запропонована формула:

$$H = a^{-0,8}, \tag{4.11}$$

де a – параметр форми розподілу Парето.

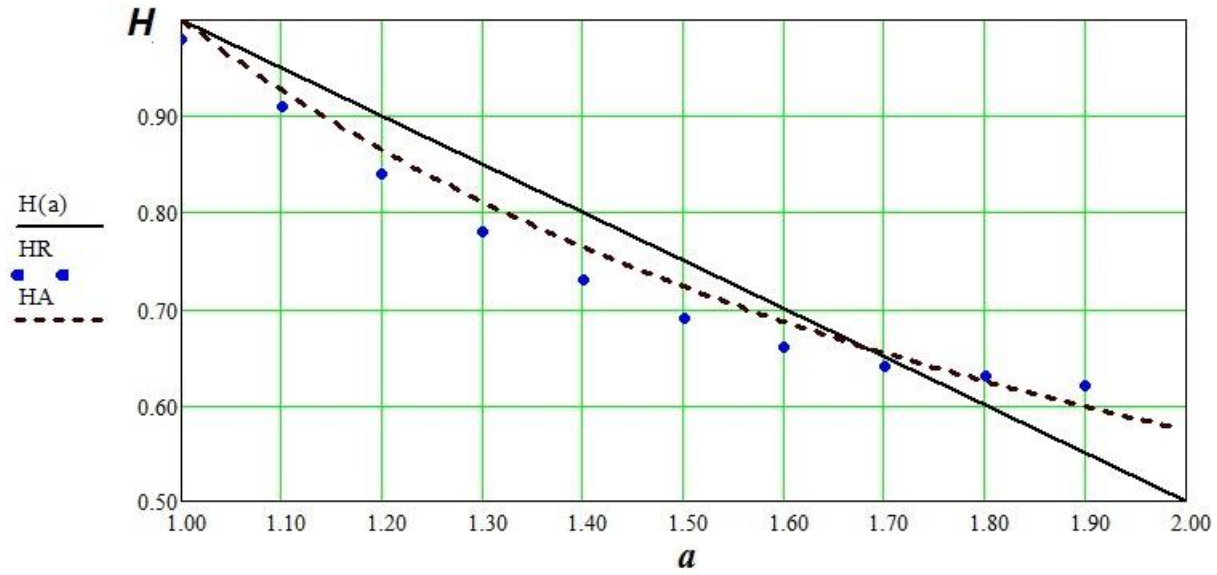


Рисунок 4.9 – Апроксимація показника Херста HA

Апроксимація (4.11) показника Херста *HA* (штрихова лінія на рис. 4.9) не повністю відповідає кривій реальної зміни показника Херста в залежності від параметра форми a розподілу Парето (позначено точками, але і в такому вигляді забезпечує точність розрахунку характеристик якості обслуговування в середньому на порядок вищу, ніж при розрахунку з використанням формули (4.10). При цьому похибка розрахунку у середньому не перевищує 10…20 %.

З урахуванням апроксимації (4.11) у системі *Pa*/*D*/1/∞ середня кількість вимог розраховується через параметр форми розподілу Парето a так:

$$N = \frac{(1-\rho)^{\frac{a^{-0,8}}{a^{-0,8}-1}}}{\rho^{\frac{0,5}{a^{-0,8}-1}}}. \tag{4.12}$$

Розподіл Вейбулла. Параметр форми a розподілу Вейбулла і показник Херста H прийнято вважати [1], що знаходяться в такій залежності

$$H = \frac{2-a}{2}. \qquad (4.13)$$

Проте, результати моделювання на рис. 4.10 показують, що для розподілу Вейбулла теж немає лінійної залежності (4.13) показника Херста H від параметра форми a цього розподілу.

З рис. 4.10 випливає, що реально коефіцієнт Херста HR (пунктирна крива) залежить від параметра a форми розподілу Вейбулла не лінійно (суцільна лінія), а згідно із законом, близьким до експоненти. Якщо реальна статистика трафіка (інтервал часу між вимогами або пакетами) апроксимується розподілом Вейбулла, то для формули Норроса (4.6) треба розраховувати коефіцієнт самоподібності Херста не за формулою (4.13), а за формулою апроксимуючої кривої HR, показаною на рис. 4.10 пунктирною лінією. При цьому точність розрахунку (4.13) істотно зростає. Наприклад, з (4.13) маємо, що при a = 0,3 показник H = 0,85, а фактично H = 0,6, що на 40 % менше.

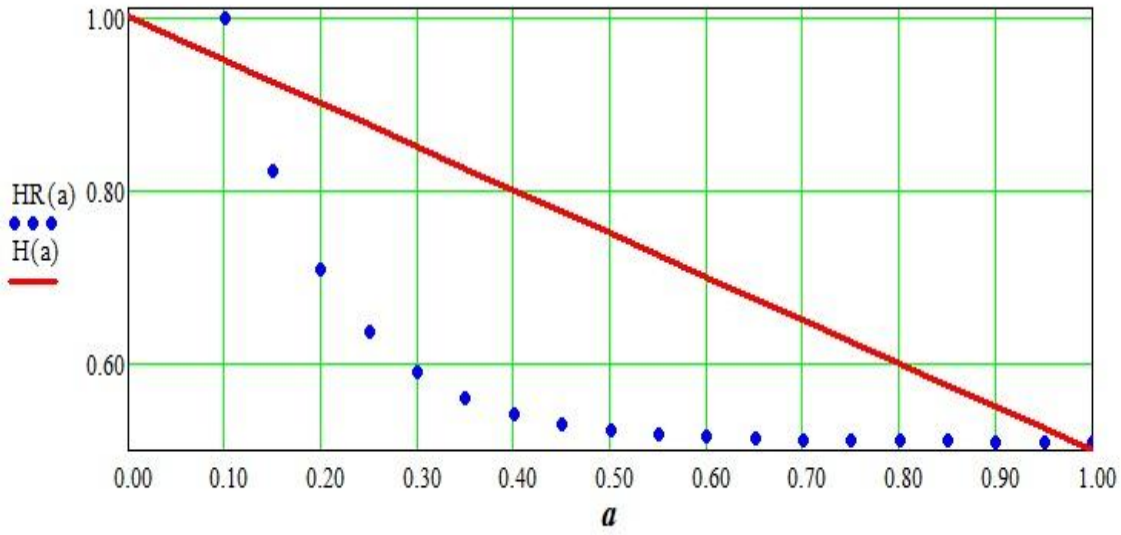


Рисунок 4.10 – Моделювання коефіцієнта самоподібності *HR* в моделі *fBM*/*D*/1/∞ (*Wb*/*D*/1/∞)

За результатами імітаційного моделювання, здійсненого за допомогою алгоритму [7], для розрахунку показника Херста трафіку, описуваного розподілом Вейбулла (позначено *Wb*), запропоновано наступну формулу:

$$Hw = 1{,}2e^{-9a} + 0{,}51. \qquad (4.14)$$

де a – параметр форми розподілу Вейбулла.

На рис. 4.11 показана апроксимація показника Херста HR (штрихова лінія) за формулою (4.14).

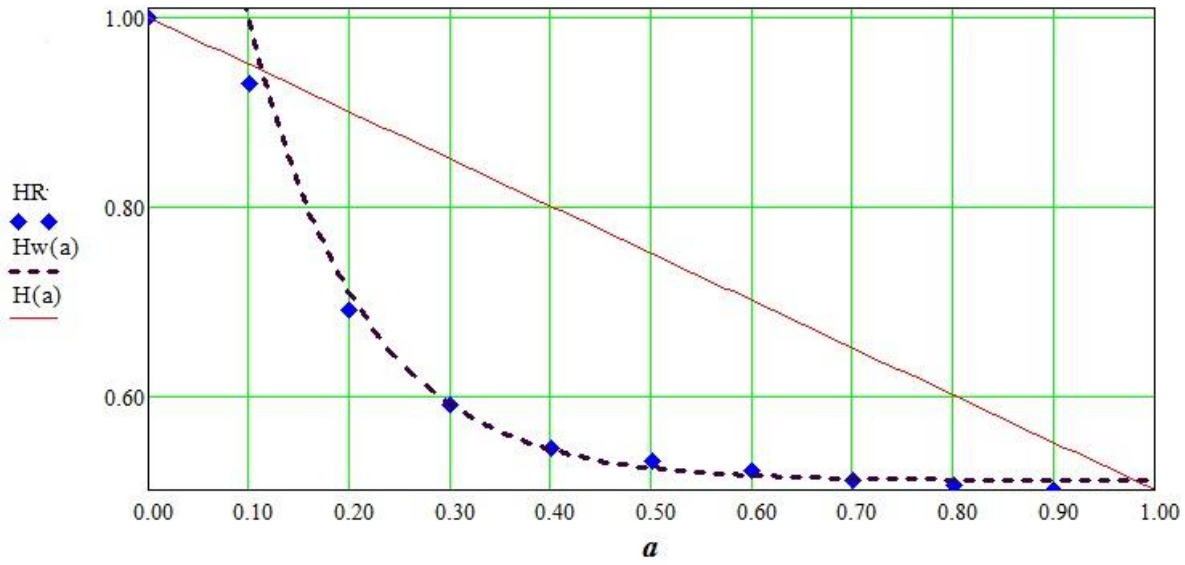


Рисунок 4.11 – Апроксимація показника Херста *HR* в моделі *Wb*/*D*/1/∞

Апроксимація (4.14) показника Херста HR (штрихова лінія), показана на рис. 4.11, хоча і не повністю відповідає кривій реальної зміни показника Херста в залежності від параметра *a* форми розподілу Вейбулла, але забезпечує точність розрахунку характеристик якості обслуговування на порядок вище, ніж при розрахунках з використанням формули (4.13). При цьому похибка розрахунку у середньому не перевищує 10 ... 20 % [15].

На рис. 4.12 для діапазону значень показника H = 0,6 ... 0,9 надана апроксимація виду

$$Hw1 = 4{,}1e^{-19a} + 0{,}57, \qquad (4.15)$$

яка забезпечує ще більш точний розрахунок показника Херста H в залежності від параметра форми a розподілу Вейбулла. Саме в цьому діапазоні, в основному, й знаходяться значення коефіцієнтів H реального самоподібного трафіка пакетних мереж зв'язку.

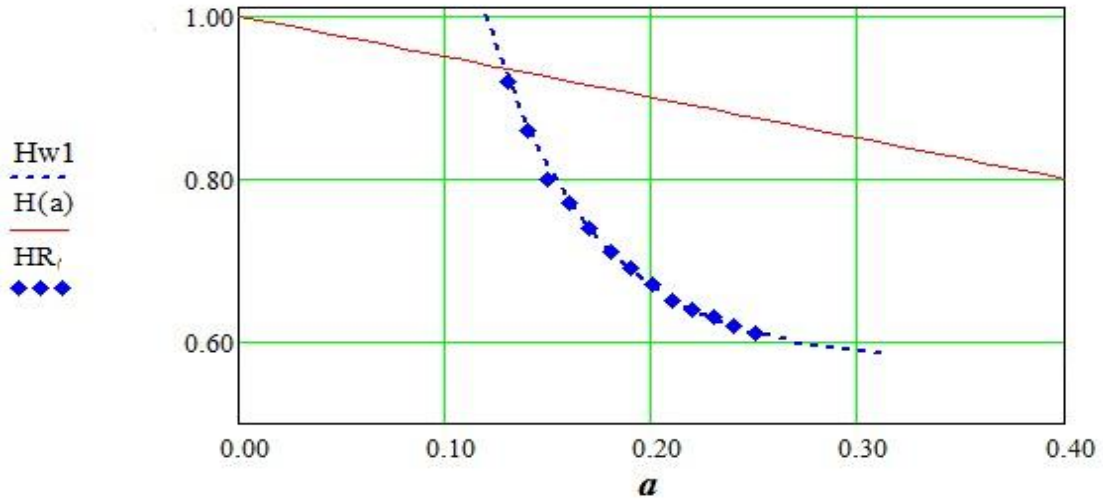


Рисунок 4.12 – Апроксимація показника Херста *HR* при H = 0,6…0,9

Всі надані на рис. 4.9 … 4.12 графіки демонструють істотну відмінність реальної (4.11) або (4.14) і використовуваної зараз лінійної залежності (4.10) або (4.13) показника Херста (коефіцієнта самоподібності) трафіка від параметра форми *a* розподілів Парето або Вейбулла у системах *Pa*/*D*/1/∞ або *Wb*/*D*/1/∞ відповідно.

Для оцінки характеристик якості обслуговування в односерверній системі *fBM*/*D*/1/∞ з дискретним часом обслуговування вимог та самоподібним трафіком, який являє собою фрактальний броунівський процес, необхідно знати його ймовірнісно-часові параметри або функцію розподілу інтервалу часу між вимогами. Якщо нею є функція Парето або Вейбулла, то потрібно розрахувати через параметр форми *a* обраного розподілу показник Херста (коефіцієнт самоподібності *H*). проте цей розрахунок слід виконувати не за лінійними формулами (4.10) або (4.13), а за уточненими формулами апроксимації (4.11) або (4.14) відповідно до моделей *Pa*/*D*/1/∞ та *Wb*/*D*/1/∞ (узагальнена назва – *fBM*/*D*/1/∞). При цьому нема потреби розраховувати для даного трафіка показник Херста або коефіцієнт самоподібності достатньо складним способом, наприклад, методом абсолютних моментів (найменших квадратів) або методом *R*/*S*-статистики.

Застосовуючи реальні функціональні залежності показника Херста *H* від параметра *a* форми розподілу Парето або Вейбулла (4.11) або (4.14), можна підвищити точність розрахунку характеристик якості обслуговування на порядок порівняно з використанням лінійних залежностей (4.10) або (4.13).

Визначивши даним способом показник Херста *H*, потім достатньо розрахувати середню кількість вимог у системі *N* за формулою Норроса (4.6). Інші характеристики *QoS*, такі як середня кількість вимог у черзі *Q*, середній час перебування вимог у системі *T* і середній час затримки вимог у системі *W* розраховуються за наступними формулами:

$$Q = N - \rho, \qquad T = \frac{N}{\rho}, \qquad W = T - 1, \qquad (4.16)$$

де *T* и *W* дано в умовних одиницях середньої тривалості обслуговування.

4.4 Апроксимація станів системи *fBM*/*D*/1/∞

Розрахунок характеристик якості обслуговування (*QoS*) в одноканальній системі з нескінченною чергою при самоподібному трафіку (модель *fBM*/*D*/1/∞) часто зводиться до знаходження показника Херста або коефіцієнта самоподібності трафіка, після чого за відомою формулою Норроса (4.6) розраховується середня кількість пакетів у системі *N*. Інші характеристики, такі як середня кількість пакетів у черзі *Q*, середній час перебування пакетів у системі *T* і середній час затримки пакетів у системі *W* потім розраховуються виходячи із відомих їх функціональних співвідношень із розрахованим середнім значенням *N* (4.16). Проте такий алгоритм не дозволяє за встановленим значенням показника Херста *H* розрахувати ще й такі характеристики *QoS*, як імовірність очікування обслуговування пакета P_w та середній час затримки пакетів t_q в накопичувальному буфері.

Отже, необхідно визначити функцію розподілу станів одноканальної системи з нескінченною чергою та самоподобнім трафіком і на її основі отримати формулу розрахунку імовірності очікування обслуговування пакета та середнього часу затримки пакетів в накопичувальному буфері.

У математичних моделях теорії телетрафіка враховано вид вхідного потоку, схема системи розподілу інформації й дисципліна обслуговування. У даному випадку розглядається вхідний потік із самоподібними властивостями, який являє собою фрактальний броунівський процес *fBM*, що описується, наприклад, розподілами Парето або Вейбулла. Дисципліна обслуговування пакетів потоку – без втрат із можливістю очікування в нескінченній черзі, а дисципліна обслуговування пакетів із черги – за правилом *FIFO – firs input-firs output*. Схема СРІ – одноканальна (односерверна).

Оцінка характеристик якості обслуговування в СМО завжди виконується на основі математичного опису реакції системи на вхідний потік пакетів. Під реакцією системи розуміють її стани, які через випадкову природу потоку пакетів математично описуються імовірнісною функцією розподілу кількості зайнятих серверів та місць очікування P_i, де i – кількість пакетів у системі (у серверах і черзі). Ця функція збігається із функцією розподілу кількості пакетів у системі (обслуговуваних і тих, що чекають у черзі), оскільки кожний пакет займає один сервер при обслуговуванні або одне місце у черзі при очікуванні.

У випадку найпростішої пуассонівської моделі потоку у СРІ із втратами або очікуванням (чергою) стани системи описуються одним із відомих розподілів Ерланга (т. зв. *перший або другий розподіл Ерланга* відповідно) [2]. Знаходження функції розподілу станів системи при більш складних моделях потоків є надто важка задача, і тому для вищезгаданої моделі потоку аналогічні рішення відсутні.

Для пачкового трафіка характерна сильна нерівномірність інтенсивності надходження пакетів. Пакети не плавно розсерджені по різних інтервалах часу, а групуються в «пачки» на одних інтервалах, і повністю відсутні або їх дуже мало на інших інтервалах часу (п. 4.1). Тому, за такого трафіка у функції розподілу кількості пакетів в одноканальній системі істотно зростає ймовірність P_0 повної відсутності пакетів у ній, що і показано на рис. 4.13.

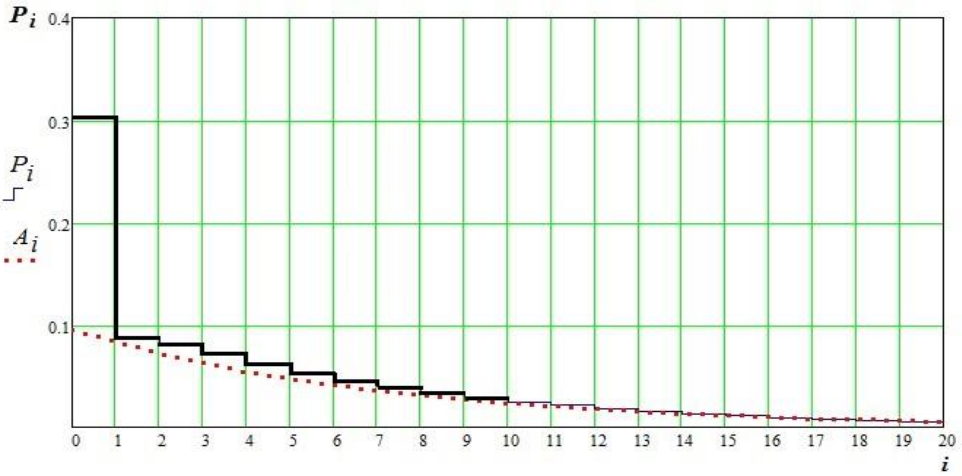


Рисунок 4.13 – Функція розподілу станів системи P_i та її апроксимація A_i

Ефективність обслуговування такого трафіка дуже низька, оскільки в процесі його оброблення у періоди спаду навантаження з імовірністю P_0 ресурси системи сильно недовикористовуються, а при піках навантаження необхідно збільшувати довжину накопичувального буфера для недопущення втрат пакетів. Проектування ж пропускної здатності системи, як правило, відбувається виходячи із середнього значення інтенсивності трафіка, що не забезпечує одночасно її ефективного використання та заданого рівня *QoS*.

Із рис. 4.13 випливає, що частина функції розподілу кількості вимог у системі P_i без імовірності P_0 достатньо якісно узгоджується із апроксимуючою функцією A_i, в якості якої запропоновано наступний вираз:

$$A_i = \rho \frac{\rho}{N} \exp\left(- \frac{\rho}{N} i \right), \tag{4.17}$$

де ρ – загрузка системи (0,3 < ρ < 1); N – середня кількість пакетів у системі.

Як видно із (4.17), апроксимуюча функція A_i являє собою результат помноження загрузки системи ρ на експонентну функцію з параметром розподілу (ρ / N) і тому $\sum_{0}^{\infty} A_i \neq 1$, тобто всі ймовірності A_i не відображають повну групу подій.

За непуассонівського потоку (наприклад, з розподілом *GI* або *fBM*) імовірність очікування в одноканальній системі згідно з [2, с. 89] визначиться:

$$P_w = \sum_{i=1}^{\infty} P_i^{'} = 1 - P_0^{'}, \tag{4.18}$$

де $P_i^{'}$ – імовірність наявності у системі i пакетів *у моменти надходження нових пакетів*. А у функції розподілу P_i, показаній на рис. 4.13, кожне значення P_i *не залежить від моменту надходження пакета у систему* (не залежить від того, надходить чи не надходить пакет у систему), тому імовірність P_0 для розрахунку імовірності очікування P_w не підходить.

З точки зору функції розподілу станів системи $P_i^{''}$, яка складається із імовірностей $P_i^{''}$ наявності у системі i пакетів тільки *під час ненадходження нових пакетів,* подія «очікування обслуговування» відбувається тільки тоді, коли у системі є два та більше пакетів, тобто імовірність очікування

$$P_w = 1 - P_0^{''} - P_1^{''}. \tag{4.19}$$

Функція A_i не є повною мірою функцією розподілу кількості пакетів у системі, а тільки лише її частина, починаючи з A_1, близька до частини функції P_i без імовірності P_0. Функція A_i без A_0 наближено описує новий простір подій наявності у системі від одного до нескінченної кількості пакетів. У цьому новому просторі подій можна розрахувати ймовірності $P_1^{'''}$, $P_2^{'''}$ і так далі, розглядуючи їх відповідно до класичного визначення ймовірності – «ймовірність події дорівнює відношенню кількості сприятливих цій події випадків до загальної кількості випадків». Таким чином, наприклад, імовірність $P_1^{'''}$ буде визначена так:

$$P_1''' = \frac{A_1}{\sum_{i=0}^{\infty} A_i}. \qquad (4.20)$$

Сума всіх імовірностей A_i в знаменнику (4.20) отримана із простору подій, в якому вилучено подію «повної відсутності пакетів у системі» з імовірністю P_0 із розподілу P_i, де кожне значення P_i не залежить від того, надходить чи не надходить пакет у систему. Тобто, нормування ймовірності A_1 виконується сумою імовірностей A_i або ймовірностей простору, у якому неможлива подія «відсутність пакетів у системі», від чого пакети у системі «завжди є» (в моменти їх надходження і ненадходження). Тому імовірність P_0'' у імовірності P_1''' вже врахована. Якщо пакети у системі «завжди є», то при цьому подія, яка полягає в наявності у системі одного пакета (мінімально можлива кількість пакетів за постійної їх наявності), може бути тільки під час ненадходження нових пакетів. Тому й імовірність P_1'' у імовірності P_1''' також уже врахована. Отже, імовірність P_1''' дорівнює сумі ймовірностей P_0'' та P_1'':

$$P_1''' = P_0'' + P_1''. \qquad (4.21)$$

Таким чином, відповідно до виразів (4.19), (4.20) і (4.22) імовірність очікування обслуговування пакета в одноканальній системі з нескінченною чергою типу *fBM*/*D*/1/∞ визначиться так:

$$P_w = 1 - \frac{A_1}{\sum_{i=0}^{\infty} A_i}. \qquad (4.22)$$

З урахуванням постійної частини $\rho\frac{\rho}{N}$ апроксимуючої функції (4.17), яка є присутня в чисельнику і знаменнику виразу (4.22), кінцевий вираз для розрахунку імовірності очікування буде такий:

$$P_w = 1 - \frac{\exp\left(-\frac{\rho}{N}\right)}{\sum_{i=0}^{\infty} \exp\left(-\frac{\rho}{N} i\right)}. \qquad (4.23)$$

Отже, якщо є можливість задати середню кількість пакетів у системі N (або після визначення за формулою Норроса (4.6) показника Херста і за ним розрахувати наближене значення N), то скориставшись апроксимацією (4.17) за формулою (4.22), або відразу за формулою (4.23) можна розрахувати імовірність очікування обслуговування пакета P_w.

Далі через співвідношення (4.16) розраховуються такі характеристики, як середній час перебування пакетів у системі T та з нього середній час затримки пакетів у системі W. І тільки після цього можна розрахувати середній час затримки пакетів у накопичувальному буфері за формулою

$$t_q = \frac{W}{P_w}. \qquad (4.24)$$

Слід зазначити, що імітаційне моделювання підтвердило коректність даного методу розрахунку характеристик якості обслуговування у системі *fBM*/*D*/1/∞ із самоподібним трафіком. При цьому розходження результатів моделювання і розрахунку не перевищує 5 % при зміні загрузки системи в діапазоні $0{,}3 < \rho < 1$ (при $\rho \geq 0{,}6$ похибка менше 2 %) і зміні значень коефіцієнта самоподібності Херста в діапазоні $0{,}5 < H < 0{,}9$.

З формули (4.18) випливає, що за наявності функції розподілу станів системи *у моменти надходження пакетів* розрахунок імовірності очікування обслуговування пакета стає значно простішим. Отже, необхідно визначити функцію розподілу станів одноканальної системи з нескінченною чергою та самоподобним трафіком *у моменти надходження пакетів*.

Загрузка системи ρ (*utilization factor*) визначається як відношення інтенсивності вхідного потоку вимог λ до інтенсивності обслуговування μ. Для одноканальної системи за будь-якого потоку пакетів (довільний розподіл *G* інтервалу часу між моментами надходження пакетів) $\rho = 1 - p_0$, де p_0 – імовірність вільності системи або стан системи p_0 (у системі 0 пакетів). Таким чином ρ збігається з імовірністю зайнятості системи або $P_{зн} = \rho$.

За пуассонівського потоку пакетів імовірність очікування P_w збігається з імовірністю зайнятості системи $P_{зн}$ (2.19) і тому для одноканальної моделі, наприклад, *M*/*G*/1/∞ (за будь-якого розподілу тривалості обслуговування) отримуємо $P_w = P_{зн} = \rho$.

З урахуванням пакетів, що знаходяться у черзі, у стаціонарному режимі існує стаціонарний розподіл станів системи або кількості пакетів у системі p_k, де k – кількість пакетів (стан p_0 – у системі 0 пакетів; стан p_1 – зайнятий єдиний канал; стан p_2 – зайнятий канал і одне місце у черзі і т.д.). Розподіл p_k *не залежить від моментів надходження пакетів* до системи (не залежить від того, *надходить чи не надходить пакет у систему*). За пуассонівського потоку цього розподілу достатньо для розрахунку імовірності очікування P_w, оскільки

$$P_w = \sum_{k=1}^{\infty} p_k = 1 - p_0. \qquad (4.25)$$

За довільного потоку пакетів, наприклад, система *GI*/*G*/1/∞, $P_w \neq P_{зн}$ і дану формулу можна використати тільки у випадку, якщо відомий розподіл кількості пакетів у системі *в моменти надходження нових пакетів* r_k, де k – кількість пакетів, а для цього необхідно в (4.25) замість ймовірностей p_k підставити ймовірності r_k, які аналогічні ймовірностям $P_i^{'}$ з формули (4.18). Розподіл p_k відрізняється від розподілу r_k тим, що $p_0 = 1 - P_{зн}$ (або $p_0 = 1 - \rho$), у той час як $r_0 = 1 - P_w$. З цього випливає, що пакет повинен очікувати обслуговування з імовірністю $P_w = 1 - r_0$. Для системи *M*/*G*/1/∞ виконується рівняння $p_k = r_k$, і тому замість розподілу r_k використовують розподіл p_k [5].

Отже, у випадку самоподібної моделі потоку пакетів з розподілом інтервалу часу між моментами надходження пакетів за законами Парето або

Вейбулла розрахунок імовірності очікування обслуговування можливий, якщо відомий розподіл станів системи або кількості пакетів у системі *в моменти надходження нових пакетів* r_k.

На рис. 4.14 для одноканальної системи з нескінченною чергою штриховою лінією показано функцію розподілу кількості пакетів у системі p_k, яка *не залежить від моментів надходження пакетів* до системи (імовірність p_0 найбільша), а суцільною ломаною лінією показано функцію розподілу кількості пакетів у системі *в моменти надходження нових пакетів* r_k. Дані функції отримано за допомогою програми моделювання самоподібного трафіка [7].

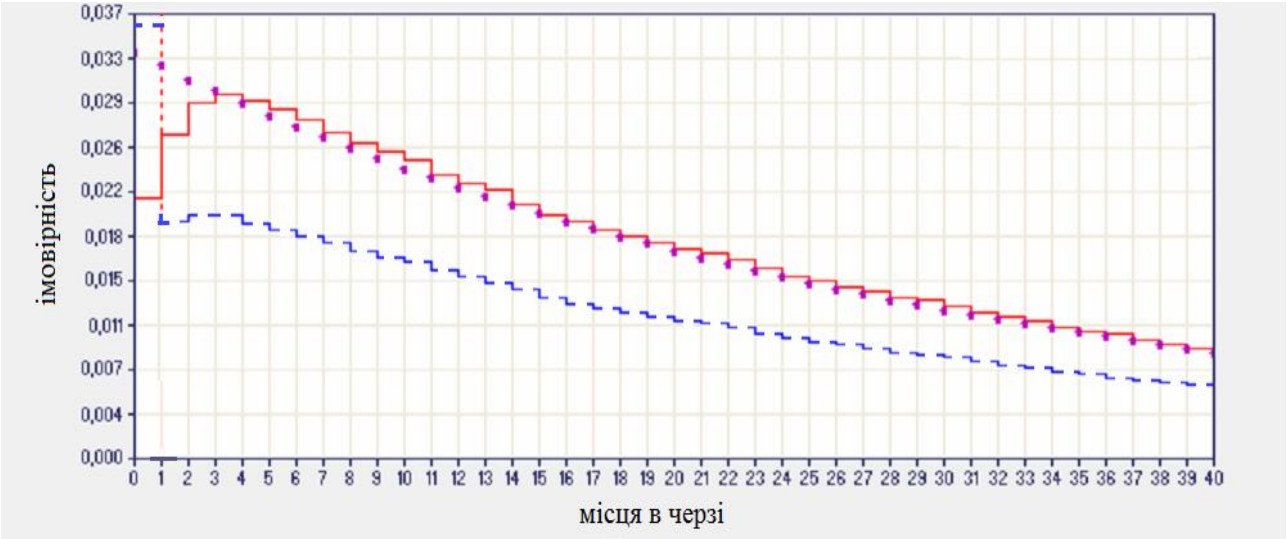


Рисунок 4.14 – Функція розподілу станів системи та її апроксимація

Із рис. 4.14 випливає, що основна частина функції розподілу кількості вимог у системі в моменти надходження нових пакетів r_k без імовірностей r_0, r_1 та r_2 достатньо якісно узгоджується із апроксимуючою функцією B_i (показано точками), в якості якої запропоновано наступний вираз:

$$B_i = \frac{1}{1+W}\exp\left(-\frac{1}{1+W}i\right), \qquad (4.26)$$

де W – середній час затримки пакетів у системі.

У формулі (4.26) апроксимуюча функція B_i являє собою експонентну функцію з параметром розподілу $1 / (1 + W)$.

За непуассонівського потоку з узагальненим розподілом *GI* інтервалу часу між моментами надходження пакетів (наприклад, потоку типу *fBM*) імовірність очікування обслуговування в одноканальній системі розраховується за формулою (4.25), але обов'язково із застосуванням функції розподілу кількості пакетів у системі *в моменти надходження нових пакетів* r_k, а не p_k. Але, як видно із рис. 4.14, безпосередній розрахунок імовірності r_0 (тобто B_0) з використанням апроксимуючої функції (4.26) буде з великою похибкою. Тому й похибка розрахунку імовірності очікування за формулою $P_w = 1 - B_0$ буде такою самою. Отже, відповідно до виразів (4.18) або (4.25) і (4.26) імовірність

очікування обслуговування пакета в одноканальній системі з нескінченною чергою типу *fBM*/*D*/1/∞ визначиться так:

$$P_w = \sum_{k=1}^{\infty} r_k \approx \sum_{k=1}^{\infty} B_k = \sum_{k=1}^{\infty} \frac{1}{1+W} \exp\left(-\frac{1}{1+W} k\right). \qquad (4.27)$$

Звідси, якщо задати середній час затримки пакетів у системі *W*, то скориставшись апроксимацією (4.26) за формулою (4.27) можна розрахувати імовірність очікування обслуговування пакета P_w. Оскільки у параметрі апроксимуючого розподілу (4.26) у знаменнику $1 + W = T$, де T – середній час перебування пакетів у системі, то у практичних розрахунках в розподілі (4.26) можна задавати не *W*, а відповідно *T*.

Далі через співвідношення (4.16) розраховуються такі характеристики, як середній час перебування пакетів у системі *T* та з нього середній час затримки пакетів у системі *W*. Після цього можна розрахувати середній час затримки пакетів у накопичувальному буфері t_q за формулою (4.24).

Імітаційним моделюванням підтверджено коректність даного методу розрахунку характеристик *QoS* у системі *fBM*/*D*/1/∞. При цьому розходження результатів моделювання і розрахунку не перевищує 5 % при зміні загрузки системи в діапазоні $0{,}3 < \rho < 1$ (при $\rho \geq 0{,}6$ похибка менше 2 %) і зміні значень коефіцієнта самоподібності в діапазоні $0{,}5 < H < 0{,}9$. Як видно з рис. 4.14, розрахована імовірність очікування обслуговування пакета P_w завжди буде дещо завищеною, оскільки апроксимуюча функція (4.26) дає також трохи завищені результати щодо реальних ймовірностей r_1 та r_2, які входять до суми формули розрахунку B_k (4.27). Наприклад, з рис. 4.14 видно, що імовірність $r_0 = 0{,}022$ і тому реальна імовірність очікування $P_w = 0{,}978$. Розрахунок цієї ж імовірності за формулою (4.27) дає значення $P_w = 0{,}983$, що тільки на 0,5 % перевищує реальне значення імовірності очікування обслуговування.

4.5 Розрахунок показника Херста методом *R*/*S*-аналізу

В пакетних мереж зв’язку прийнято вважати, що трафік описується самоподібним (*self-similarity*) випадковим процесом з параметром Херста близько 0,65 ... 0,9 [1, 4]. Трафік пакетних мереж зв’язку можна уявити як процес надходження пакетів, описуваний часовим рядом кількості пакетів по фіксованим відрізкам часу (миттєва інтенсивність надходження пакетів). Статистичний *R*/*S*-аналіз, створений Г. Херстом – це метод аналізу динаміки часових послідовностей. *R*/*S*-аналізом також вважають сукупність статистичних прийомів і методів аналізу часових рядів (переважно фінансових), що дозволяють визначити деякі важливі їх характеристики, такі як наявність неперіодичних циклів, пам’яті.

Метод Херста дозволяє виявити у статистичних даних пакетного трафіка такі його властивості, як кластерність, тенденцію слідувати у напрямку тренда (персистентність) і швидку змінюваність послідовних значень інтенсивності трафіка (сплески інтенсивності, що призводять до пачкування), післядію,

сильну пам’ять, фрактальність (самоподібність), наявність періодичних і неперіодичних циклів (через особливості застосованих протоколів передачі).

Проте, існуючі методи розрахунку показника Херста є досить трудомісткими, що ускладнює їх використання в умовах реального процесорного часу оброблення параметрів трафіка при виявленні його самоподібних властивостей.

Розрахунок коефіцієнта самоподібності Херста реального трафіка зазвичай виконується методом абсолютних моментів. Як значення випадкового процесу розглядається кількість пакетів, що надходять у СРІ за одиницю часу. Вихідна послідовність (ряд) кількості пакетів довжиною N ділиться на періоди довжиною m (окремі агреговані процеси розміром m). На непересічних часових інтервалах або на кордонах кожного періоду k-та послідовність має середнє значення:

$$X_k^{(m)} = \frac{1}{m}\sum_{j=1}^{m} X_{(k-1)m+1}, \qquad k = 1, 2, 3 \ldots, [N / m].$$

Після розрахунку середнього значення для всієї послідовності, потім для кожного періоду k розраховується дисперсія D_k:

$$D_k^{(m)} = \frac{1}{N/m}\sum_{j=1}^{m}\left(X_k^{(m)} - \overline{X}\right)^2 .$$

Для самоподібного процесу дисперсія агрегованих процесів повинна спадати повільніше, ніж величина, зворотна розміру вибірки m [1]. Для виявлення цієї властивості будується дисперсійно-часовий графік залежності дисперсій агрегованих процесів від ступеня агрегування m. Оскільки Херстом було показано, що:

$$\log\left(\frac{\max D - \min D}{D_k^{(m)}}\right) \approx H \log\left(\frac{N}{2}\right),$$

то графік цієї залежності будується теж в логарифмічному масштабі. Вираз в лівій частині цього рівняння називається R/S-статистикою або нормованим розмахом. З отриманого графіка визначається коефіцієнт β як тангенс кута нахилу апроксимуючої кривої до побудованої залежності. Дана апроксимація виконується методом мінімального середньоквадратичного відхилення від експериментальних даних. Коефіцієнт β ($0 < \beta < 1$), що задає асимптотичні властивості характеристик самоподібного випадкового процесу, пов'язаний з показником Херста наступним співвідношенням [1]:

$$H = 1 - \frac{\beta}{2}.$$

Для процесів, які не мають властивості самоподібності, $H < 0{,}5$, а для самоподібних процесів з довгостроковою залежністю цей параметр змінюється в межах 0,65 ... 0,9 (процес «має тривалу пам’ять»).

Підготовка ряду для розрахунку. З основного часового ряду довжиною N елементів (значень) створюються нові допоміжні ряди, для яких розраховуються певні числові характеристики. Кожен допоміжний ряд ділиться

на суміжні періоди $I_{k,n}$ довжиною m_k елементів, де k – номер допоміжного ряду, а n – номер періоду. Максимальна кількість періодів допоміжного ряду $n_{max} = N / m_{min}$, де m_{min} – мінімальна довжина періоду. Для кожного натурального $2 \le n \le n_{max}$ можна скласти допоміжний ряд, проте точність розрахунку істотно не погіршиться і при меншій кількості рядів. Наприклад, на рис. 4.15 показані основний 0-й ряд довжиною $N = 320$ елементів і допоміжні ряди 1, 2 і 3, для яких $m_1 = 32$, $m_2 = 80$ і $m_3 = 160$ елементів відповідно.

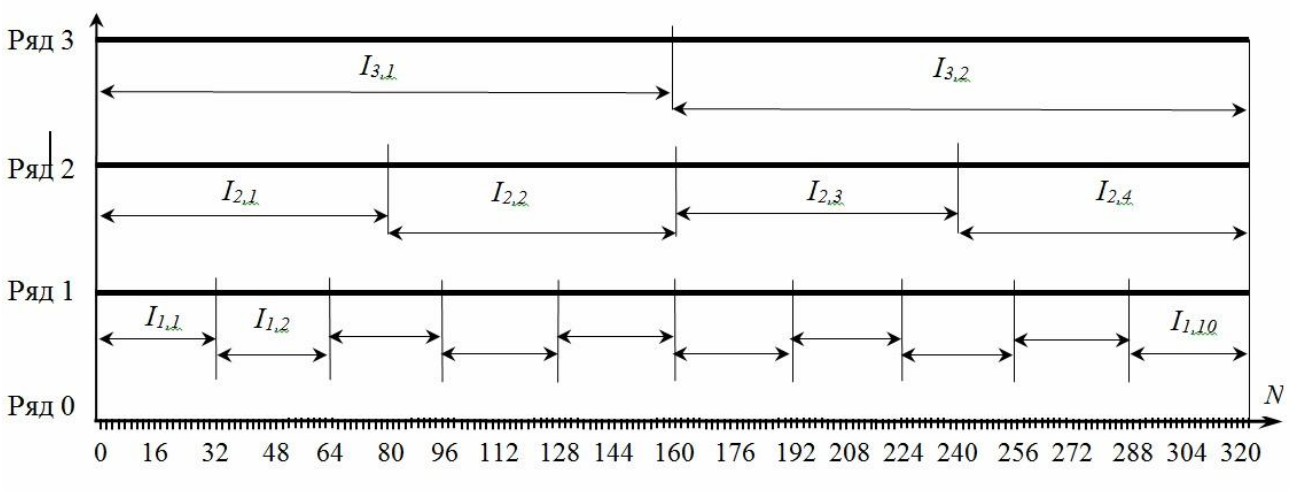


Рисунок 4.15 – Основний нульовий та допоміжні ряди 1…3

З рис. 4.15 випливає, що 1-й допоміжний ряд отриманий шляхом ділення основного ряду на десять, 2-й допоміжний ряд – на чотири і 3-й допоміжний ряд – на два періоди. Для отримання достовірного результату обов'язково $k_{max} \ge 5$ при $m_{min} \ge 10$. Тому, в даному випадку для отримання ще трьох допоміжних рядів можна, наприклад, розділити основний ряд на 5, 8 і 20 періодів довжиною $m_4 = 64$, $m_5 = 40$ і $m_6 = 16$ відповідно.

При кількості елементів в основному ряду $N = 2^i$ зручно визначити мінімальну довжину періоду $m_{min} = 2^z$, де, наприклад, $z = 3$, 4, 5 і більше. При цьому максимальна кількість періодів допоміжного ряду $n_{max} = N / m_{min} = 2^i/2^z = 2^{(i - z)}$, а кількість допоміжних рядів $k_{max} = i - z$ або $k_{max} = \log2\ (n_{max})$. Наприклад, при $N = 65536$ і $m_{min} = 16$ маємо $k_{max} = 12$ допоміжних рядів. У цьому випадку буде 12 значень $n = 2^{12}$, 2^{11}, 2^{10}, ..., 2^1 і відповідно для кожного з них $m_k = N / 212$, $N / 211$, $N / 210$, ..., $N / 2$.

Результати розрахунку кожного допоміжного ряду використовуються для розрахунку коефіцієнта Херста методом найменших квадратів. Після підготовки відповідно до наведеного алгоритму допоміжних рядів здійснюються такі розрахунки.

Для кожного n-го періоду $I_{k,n}$ кожного ряду k:

1. Розраховується середнє арифметичне значення $E_{k,n}$ елементів Nj

$$E_{k,n} = \frac{1}{m_k} \sum_{j=1}^{m_k} N_j \,, \qquad (4.28)$$

де j – порядковий номер елемента в періоді n.

2. Накопичується сума $X_{k,n}$ розрахованих відхилень кожного елемента N_j від середнього значення елементів

$$X_{k,n} = \sum_{j=1}^{m_k - 1} (N_j - E_{k,n}). \tag{4.29}$$

Кожне накопичуване в циклі (4.29) значення $X_{k,n}$ запам'ятовується, утворюючи для n-го періоду $I_{k,n}$ ряду k кумулятивний ряд X_j,
де: $X_1 = (N_1 - E_{k,n})$; $X_2 = X_1 + (N_2 - E_{k,n})$; $X_3 = X_2 + (N_3 - E_{k,n})$ і т.д.

З цього випливає, що значення кумулятивного ряду X_j розраховуються за нижче наведеною формулою, де j також змінюється від 1 до (m_k – 1):

$$X_j = \sum_{i=1}^{j} N_i - jE_{k,n}. \tag{4.30}$$

У виразі (4.29) при заміні межі підсумовування на m_k вся сума стає рівною нулю, що є ознакою правильного розрахунку.

3. З кумулятивного ряду відхилень знаходиться максимальне $X_{\max}(X_i)$ і мінімальне $X_{\min}(X_i)$ значення, а потім розраховується розмах (діапазон) відхилень $R_{k,n}$:

$$R_{k,n} = X_{\max}(X_j) - X_{\min}(X_j). \tag{4.31}$$

4. Розраховується сума квадратів відхилення кожного елемента N_j від середнього значення елементів

$$Q_{k,n} = \sum_{j=1}^{m_k} (N_j - E_{k,n})^2. \tag{4.32}$$

За цим значенням розраховуються стандартні відхилення $S_{k,n}$ (частіше позначається σ – *sigma*):

$$S_{k,n} = \sqrt{\frac{1}{m_k} \sum_{j=1}^{m_k} (N_j - E_{k,n})^2}. \tag{4.33}$$

5. Розраховується нормований розмах накопичених сум (t*he adjusted range of cumulative sums*) $RS_{k,n}$

$$RS_{k,n} = \frac{R_{k,n}}{S_{k,n}}. \tag{4.34}$$

У даному випадку величина розмаху $R_{k,n}$ нормується значенням емпіричного стандартного відхилення $S_{k,n}$ і тому іноді R/S-аналіз називається методом нормованого розмаху. (Позначення RS нормованого розмаху дало назву цьому методу – R/S-аналіз або R/S-статистика).

6. Для всіх n значень нормованого розмаху відхилень $RS_{k,n}$ ряду k розраховується середнє значення

$$\overline{RS}_k = \frac{1}{n} \sum_{1}^{n} RS_{k,n}. \tag{4.35}$$

7. Фінальний розрахунок для ряду k полягає в обчисленні логарифма величин і m_k:

$$X_k = \log(m_k), \tag{4.36}$$

$$Y_k = \log(\overline{RS}_k). \tag{4.37}$$

Значення X_k і Y_k визначають координати точок на графіку залежності $\log(\overline{RS}_k)$ від $\log(m_k)$ і використовуються для розрахунку показника Херста H методом найменших квадратів.

Логарифм може бути обраний за будь-якою основою, але для прикладу з $N = 2^i$ саме двійковий логарифм дає більш зручний вид графіка для розрахунку цієї залежності, показаної на рис. 4.16.

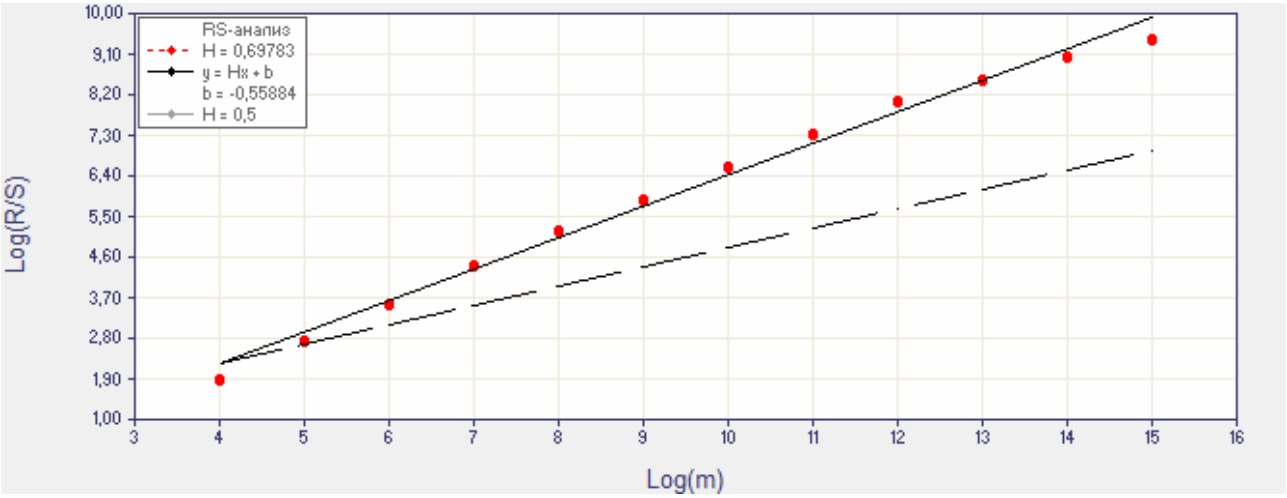


Рисунок 4.16 – Графік залежності $\log(\overline{RS}_k)$ від $\log(m_k)$

Зручність графіка полягає в тому, що за значеннями осі абсцис легко визначається довжина періоду m_k у відповідній точці графіка. В даному випадку зліва направо послідовно $m_k = 2^4, 2^5, 2^6, ..., 2^{15}$. На рис. 4.16 суцільною лінією показана пряма, що згладжує дані розрахунку X_k і Y_k, а штриховою – нахил прямої, що відповідає $H = 0{,}5$.

Розрахунок показника Херста. Для оцінки показника Херста H аналізується залежність нормованого розмаху R/S від довжини періоду m. Для цього методом лінійної регресії розраховані значення Y_k і X_k (дані кожного допоміжного ряду) апроксимуються функцією виду $y = ax + b$. Ця регресія називається методом найменших квадратів, оскільки коефіцієнти a і b обчислюються з умови мінімізації суми квадратів помилок $| b + ax_i - y_i |$. При цьому знаходяться найкращі значення параметрів a і b, які максимально наближають значення функції $y = aX_k + b$ до фактичних значень Y_k.

Розраховані значення Y_k і X_k визначають координати точок графіка залежності $\log(\overline{RS}_k)$ від $\log(m_k)$, показаного на рис. 4.16. Показник Херста H відповідає кутовому коефіцієнту a для прямої, що проходить максимально близько до цих точок або через них. Кутовий коефіцієнт a лінійної функції $y = ax + b$ розраховується так:

$$a = \frac{kg_1 - c_2 g_2}{kc_1 - c_2^2}, \tag{4.38}$$

де

$$c_1 = \sum_{i=1}^{k} X_i^2, \qquad c_2 = \sum_{i=1}^{k} X_i, \qquad g_1 = \sum_{i=1}^{k} X_i Y_i, \qquad g_2 = \sum_{i=1}^{k} Y_i, \tag{4.39}$$

а k – кількість точок на графіку (розрахованих допоміжних рядів). У цьому випадку лінія побудованої регресії проходить через центр ваги вибіркових даних Y_k і X_k.

Зсув b лінійної функції $y = ax + b$ розраховується як $b = \bar{y} - a\bar{x}$, де $\bar{y}$ і $\bar{x}$ – середні значення. Для Y_k і X_k маємо:

$$b = \frac{1}{k}\sum_{i=1}^{k} Y_i - a\frac{1}{k}\sum_{i=1}^{k} X_i . \tag{4.40}$$

Таким чином, за результатами R/S-аналізу, який визначається значеннями Y_k і X_k, розраховується показник Херста H, що дорівнює значенню кутового коефіцієнта a для прямої, що згладжує результати розрахунку Y_k і X_k.

У разі, якщо регресійна залежність між величинами Y_k і X_k досить лінійна, як, наприклад, на рис. 4.16, то розрахунок показника Херста можна виконати спрощено без використання виразів (4.38) і (4.39).

Кутовий коефіцієнт прямої – це коефіцієнт a в рівнянні $y = ax + b$ прямої на координатній площині. Він дорівнює тангенсу кута (що становить найменший поворот від осі x до осі y) між позитивним напрямом осі абсцис і даною прямою лінією.

Тангенс кута нахилу прямої може розраховуватися як відношення протилежного катета до прилеглого катета трикутника, утвореного цією прямою в якості гіпотенузи й уявними відрізками, відкладеними на величину відносного зміщення двох довільних точок цієї прямої по осях x і y, як показано на рис. 4.17.

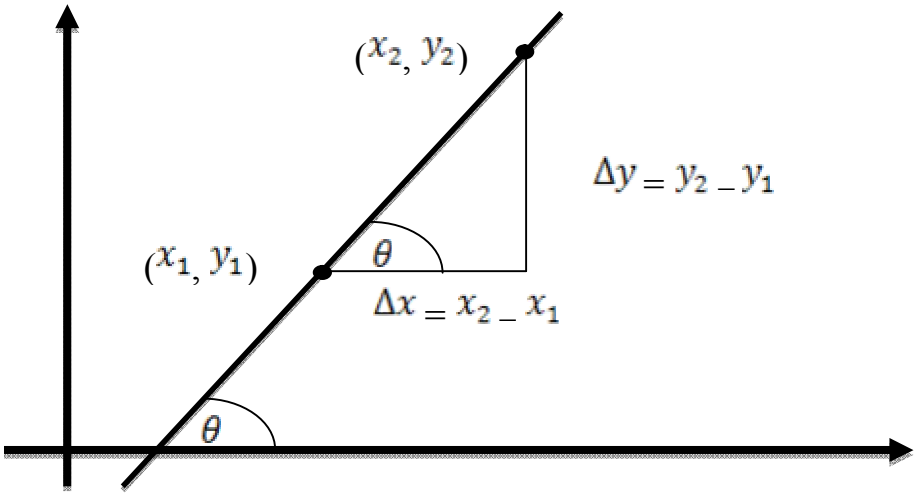


Рисунок 4.17 – Метод визначення кута нахилу прямої

Коефіцієнт a завжди дорівнює відношенню величин відносного зміщення двох довільних точок цієї прямої по осях x і y:

$$a = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}, \tag{4.41}$$

тобто він дорівнює похідній рівняння прямої по x.

З рис. 4.16 видно, що крайні точки графіка дещо віддалені від апроксимуючої прямої, отриманої в результаті регресії даних Y_k і X_k. Отже, з цих даних для розрахунку кутового коефіцієнта a або показника Херста H необхідно використовувати координати інших точок. Наприклад, якщо використовувати з Y_k і X_k тільки результати розрахунку при k = 2 і $k = k$ - 2 (тобто відступити від кожного краю апроксимуючої прямої на дві точки), то результат спрощеного розрахунку буде наближатися до результату розрахунку за формулою (4.38):

$$H = \frac{Y_{k-2} - Y_3}{X_{k-2} - X_3}. \tag{4.42}$$

На рис. 4.18 показаний приклад спрощеного розрахунку коефіцієнта Херста за значеннями Y_k і X_k для третього і десятого допоміжних рядів з дванадцяти (для третьої і десятої точок графіка з дванадцяти, показаних на рис. 4.16). У цих точках довжина періоду обраних допоміжних рядів m_k дорівнює 2^6 і 2^{13} елементів відповідно. При цьому точність цього розрахунку досить висока, оскільки повний розрахунок за формулами (4.38) і (4.39) дає значення H = 0,69783, а за формулою (4.42) коефіцієнт H = 0,70986, що тільки на 1,72 % перевищує результат повного розрахунку.

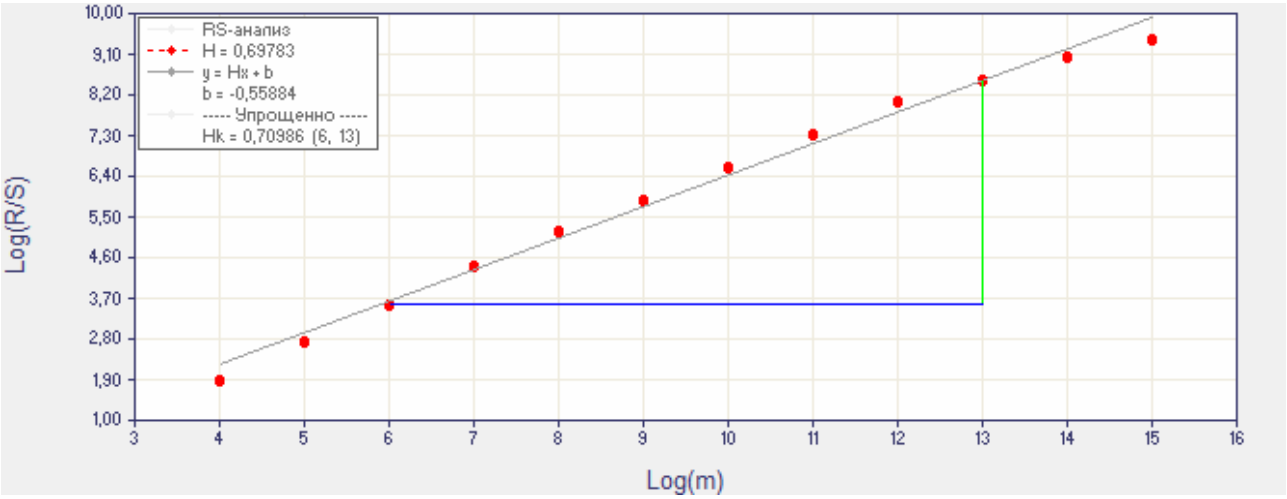


Рисунок 4.18 – Приклад спрощеного розрахунку кутового коефіцієнта

Звідси робимо висновок, що похибка розрахунку спрощеного методу не перевищує 2 ... 5%, що дозволяє використовувати його в умовах реального процесорного часу обробки, тому що істотно скорочено кількість операцій розрахунку.

Таблиця 4.1 – Методи моделювання випадкової величини

Вид функції	Параметр розподілу	Густина розподілу	Спосіб моделювання
Експонентна	λ – інтенсивність	$\lambda e^{-\lambda x}$	$x_i = -\frac{1}{\lambda}\ln U_i$
Релея	b – мода розподілу	$\frac{x}{b}e^{-\frac{x^2}{2b^2}}$	$x_i = b\sqrt{-\ln U_i}$
Вейбулла	a – параметр форми	$\lambda_0 a x^{a-1} e^{-\lambda_0 x^a}$	$x_i = \left(-\frac{1}{\lambda}\ln U_i\right)^{1/a}$
Парето	a – параметр форми b – мода розподілу (b –*мінімальне значення*)	$\frac{a}{b}\left(\frac{b}{x}\right)^{a+1}$	$x_i = \frac{b}{(U_i)^{1/a}}$

Контрольні запитання

1. Яка мета задач аналізу, синтезу та оптимізації систем розподілу інформації?
2. Які типи реального трафіка визначено для телекомунікаційних мереж і у чому різниця між ними?
3. Які параметри трафіка враховано в математичній моделі мультисервісного та пакетного трафіка.
4. Яка імовірнісна функція та з якими параметрами апроксимує стани системи з втратами за гіперекспонентного потоку вимог.
5. Як залежить імовірність втрат у системі з гіперекспонентним потоком вимог від коефіцієнта скупченості навантаження?
6. Яка імовірнісна функція та з якими параметрами апроксимує стани системи з нескінченною чергою за гіперекспонентного потоку вимог?
7. Яку характеристику QoS у системі з нескінченною чергою визначає відношення середнього часу затримки вимоги у системі та у черзі?
8. У чому різниця між поняттями „інтенсивність потоку вимог” та „інтенсивність навантаження” і як їх можна визначити?
9. Наведіть формули розрахунку обслуженого навантаження для систем з втратами, необмеженою та обмеженою чергою.
10. У чому полягає ефект «пачкування» трафіка, на що він впливає і як його можна виміряти?
11. Які імовірнісні функції розподілу параметрів трафіка можна застосувати у математичній моделі трафіка пакетних мереж зв’язку?
12. Які недоліки Ви вбачаєте у ентропійному методі розрахунку характеристик якості обслуговування самоподібного трафіка?
13. Які недоліки розрахунку харакетристик якості обслуговування самоподібного трафіка за формулою Норроса?
14. На що і як впливає показник Херста самоподібного трафіка?
15. Як залежить показник Херста генерованого трафіка від параметру розподілів Парето та Вейбулла реально і наближено?
16. З якої функції розподілу станів системи розрахунок імовірності очікування обслуговування пакета є більш простий?
17. За якою функцією розподілена кількість вимог в одноканальній системі з нескінченною чергою за умов пачкового трафіка?
18. Яка кількість допоміжних рядів та мінімальна довжина періоду допоміжного ряду мають бути для отримання найбільш достовірного результату розрахунку показника Херста?

Список літератури

1. Крылов В.В., Самохвалова С.С. Теория телетрафика и её приложения. – СПб.: БХВ-Петербург. – 2005. – 288 с.: ил.
2. Шнепс М.А. Системы распределения информации. Методы расчета: Справ. пособие. – М.: Связь, 1979. – 344 с., ил.
3. Ложковський А.Г. Теорія масового обслуговування в телекомунікаціях. – Одеса: ОНАЗ ім. О.С. Попова, 2010. – 112 с.
4. Степанов С.Н. Основы телетрафика мультисервисных сетей. – М.: Эко-Трендз. – 2010. – 392 с.: ил.
5. Клейнрок Л. Теория массового обслуживания. Пер. с англ. – М.: Машиностроение. – 1979. – 432 с., ил.
6. Кениг Д., Штойян Д. Методы теории массового обслуживания. Пер. с нем. – М.: Радио и Связь. – 1981. – 128 с., ил.
7. Ложковский А.Г., Салманов Н.С., Вербанов О.В. Моделирование многоканальной системы обслуживания с организацией очереди // Восточно-европейский журнал передовых технологий. – 2007. – №3/6(27). – С.72-76.
8. Ложковський А.Г. Нова методика оцінювання імовірності втрат викликів, наближена до реальних умов // К.: Зв'язок. – 2004. – №3. – С. 52–53.
9. Ложковский А.Г. Метод расчета систем обслуживания с ожиданием при произвольном потоке вызовов // К.: Зв'язок. – 2006. – № 1. – С. 57–60.
10. Ложковский А.Г. Сравнительный анализ методов расчета характеристик качества обслуживания при самоподобных потоках в сети / А.Г. Ложковский // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г. Є. Пухова НАН України. – Вип. 47. – К.: 2008. – С. 187-193.
11. Ложковский А.Г., Ганифаев Р.А. Оценка параметров качества обслуживания самоподобного трафика энтропийным методом // Наукові праці ОНАЗ ім. О.С. Попова. – 2008. – № 1. – С. 57-62.
12. Ложковский А.Г. Расчет одноканальных систем с бесконечной очередью при экспоненциальной длительности обслуживания / А.Г. Ложковский // Наукові праці ОНАЗ ім. О.С. Попова. – 2009. – № 2. – С. 10-13.
13. Ложковский А.Г. Модель трафика в мультисервисных сетях с коммутацией пакетов / А.Г. Ложковский // Наукові праці ОНАЗ ім. О.С. Попова. – 2010. – № 1. – С. 63-67.
14. Ложковский А.Г. Расчет характеристик самоподобного трафика, аппроксимируемого распределением Парето / А.Г. Ложковский, Е.В. Левенберг // Вісник НУ «Львівська політехніка». Серія: „Радіоелектроніка та телекомунікації". – Вип. 885. – 2017. – С. 63-67.
15. Ложковский А.Г. Повышение точности расчета коэффициента самоподобности трафика пакетной сети связи / А.Г. Ложковский, Е.В. Левенберг // Наукові праці ОНАЗ ім. О.С. Попова. – 2017. – №2. – С. 63-68.

Список позначень та скорочень

$C_m(\Lambda)$ – *C*-формула Ерланга
$E_m(\Lambda)$ – *B*-формула Ерланга
FIFO – *First in - first out* (перший обслуговується першим)
LIFO – *Last in first out* (останній обслуговується першим)
SIRO – *Service in random order* (випадкове обслуговування)
QoS – *Quality of Service* (якість обслуговування)
a – параметр форми розподілу
H – показник Херста (коефіцієнт самоподібності)
k – стан системи (кількість зайнятих серверів)
m – кількість серверів системи
P_k – імовірність стану системи у разі зайнятості k серверів
P_w – імовірність очікування
P_B – імовірність втрати (блокування) вимоги
N – середня кількість вимог у системі
Q – середня кількість вимог у черзі (середня довжина черги)
r – кількість місць очікування у черзі
T – середній час перебування вимоги у системі
W – середній час очікування у системі
t_q – середній час очікування у черзі
x – тривалість обслуговування вимоги
S – коефіцієнт скупченості навантаження (пікфактор трафіка)
Λ – інтенсивність вхідного навантаження
Y – інтенсивність обслуженого навантаження
z – тривалість інтервалу часу між вимогами
λ – інтенсивність потоку вимог
μ – інтенсивність обслуговування вимог
ρ – інтенсивність питомого навантаження
ГНН – Година найбільшого навантаження
СМО – Система масового обслуговування
СРІ – Система розподілу інформації
ТМО – Теорія масового обслуговування

Предметний покажчик

B-формула Ерланга	16
C-формула Ерланга	18
Абсолютний пріоритет	26
Базова модель Кендалла	7
Відносний пріоритет	25
Геометричний розподіл	22
Гіперекспонентний розподіл	27, 48
Дисперсія інтенсивності	8, 10
Другий розподіл Ерланга	18
Експонентний розподіл	14
Ентропія розподілу	54
Закон Пуассона	8, 14
Імовірність втрати вимоги	11, 16
Імовірність очікування	11, 18, 65
Інтенсивність вхідного навантаження	9, 14
Інтенсивність обслуговування вимог	7
Інтенсивність обслуженого навантаження	9, 11-13, 19, 22
Інтенсивність питомого навантаження	10
Коефіцієнт асиметрії	49
Коефіцієнт варіації	24
Коефіцієнт використання серверів (*utilization factor*)	41
Математичне сподівання	10
Модель потоку вимог	7, 14, 27, 46
Навантаження	9
Нормальний закон, Гауса	27
Нормальний усічений закон	33
Параметр форми розподілу	56, 58
Пачковий трафік	46
Перший розподіл Ерланга	16
Пікфактор трафіка	10
Показник Херста	50, 67
Пропускна здатність	13
Пуассонівський потік	14
Розподіл імовірностей станів системи	15, 17, 31
Розподіл Вейбулла	50, 58
Розподіл Парето	50, 56
Самоподібний процес	67
Самоподібний трафік	49, 56, 67
Середній час перебування вимоги у системі	11, 12
Середня довжина черги	12, 19, 51
Середня кількість вимог у системі	11, 12, 57
Середній час очікування (затримки) у системі	12, 23, 43
Середній час очікування (затримки) у черзі	19, 42
Скупченість навантаження	8,10
Стан системи	15
Формула Літтла	11,12
Формула Норроса	51
Формула Поллачека-Хінчина	23

Навчальне видання

Анатолій Григорович Ложковський

НОВІ МЕТОДИ ТЕОРІЇ ТЕЛЕТРАФІКА

Для студентів закладів вищої освіти,
які навчаються за спеціальністю 172 „Телекомунікації та радіотехніка”

Редактор Л.А. Кодрул

Підписано до друку 23.10.2018
Формат 60х90 1/16. Зам. № 6283
Тираж 100 прим. Обсяг 5,0 друк. арк.
Віддруковано на видавничому устаткуванні фірми RISO
у друкарні редакційно-видавничого центру ОНАЗ ім. О.С. Попова