

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА**

А. Л. Литвинов

ТЕОРІЯ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ

НАВЧАЛЬНИЙ ПОСІБНИК

**Харків
ХНУМГ ім. О. М. Бекетова
2018**

УДК 519.872(075.8)

Л64

Автор

Литвинов Анатолій Леонідович, доктор технічних наук, професор кафедри прикладної математики і інформаційних технологій Харківського національного університету міського господарства імені О. М. Бекетова

Рецензенти

Нечуйвітер Олеся Петрівна, доктор фізико-математичних наук, професор, завідувач кафедри інформаційних, комп'ютерних і поліграфічних технологій Української інженерно-педагогічної академії;

Федорович Олег Євгенович, доктор технічних наук, професор, завідувач кафедри інформаційних управляючих систем Національного аерокосмічного університету імені М. Є. Жуковського «Харківський авіаційний інститут»

Рекомендовано до друку Вченою радою Харківського національного університету міського господарства імені О. М. Бекетова, протокол № 8 від 06 квітня 2018 р.

Литвинов А. Л.

Л64 Теорія систем масового обслуговування : навч. посібник / А. Л. Литвинов ; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2018. – 141 с.

ISBN 978-966-695-473-5

Викладено базові положення та методи аналізу систем масового обслуговування. Доведена необхідність використання теорії черг для аналізу ймовірнісних процесів у комп'ютерних системах. Наведено математичний апарат і базові методи, що використовуються в теорії масового обслуговування. Проведені викладки доведено до розрахункових формул. Матеріал супроводжується значною кількістю реальних прикладів.

Підручник призначений для студентів і аспірантів, які навчаються за напрямом «Комп'ютерні науки»

УДК 519.872(075.8)

ISBN 978-966-695-473-5

© А. Л. Литвинов, 2018

© ХНУМГ ім. О. М. Бекетова, 2018

ЗМІСТ

ВСТУП	5
1 ЗАСАДИ ТЕОРІЇ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ	12
1.1 Вступ до теорії ймовірностей та математичної статистики	12
1.1.1 Базові поняття теорії ймовірностей	12
1.1.2 Дискретні випадкові величини	16
1.1.3 Неперервні випадкові величини	19
1.1.4 Математична статистика	24
1.2 Засади теорії випадкових процесів	34
1.2.1 Базові поняття випадкових процесів	34
1.2.2 Марківські випадкові процеси з дискретним часом	37
1.2.3 Марківські випадкові процеси з дискретними станами й безперервним часом	42
1.3. Потоки подій і їх характеристики	47
Контрольні питання	53
2 БАЗОВІ ПОНЯТТЯ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ ТА МЕТОДИ ЇХ ДОСЛІДЖЕННЯ	54
2.1 Базові поняття систем масового обслуговування	54
2.2 Показники ефективності систем масового обслуговування	58
2.2.1 Технічні показники ефективності систем масового обслуговування	59
2.2.2 Економічні показники ефективності систем масового обслуговування	60
2.3 Методи дослідження систем масового обслуговування	61
2.3.1 Диференціальний метод	61
2.3.2 Метод фаз	63
2.3.3 Метод вкладених ланцюгів Маркова	67
2.3.4 Метод включення додаткових змінних	72
2.3.5 Формули Літтла	76
Контрольні питання	77
3 ТИПОВІ СИСТЕМИ МАСОВОГО ОБСЛУГОВУВАННЯ	78
3.1 Система масового обслуговування з відмовами	78
3.2 Система з пуассонівським вхідним потоком і експоненціальним часом обслуговування (М/М/1)	82
3.3 Багатоканальна система масового обслуговування з обмеженою чергою	89

3.4 Одноканальна система масового обслуговування з довільними потоками подій.	94
3.5 Система масового обслуговування з кінцевою кількістю джерел запитів	96
3.6 Система масового обслуговування з абсолютним пріоритетом.	101
3.7 Мережі систем масового обслуговування.	106
Контрольні питання.	111
4 ЗАСТОСУВАННЯ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ ПІД ЧАС РОЗРОБЛЕННЯ ТЕХНІЧНИХ СИСТЕМ.	113
4.1 Дослідження й вибір параметрів модуля прийому числа-імпульсної інформації.	113
4.2 Імовірнісне моделювання та оцінювання якості функціонування інформаційно-управляючих систем.	119
4.3 Імовірнісний аналіз і вибір параметрів процесора вводу-виводу управляючих комп'ютерів.	127
Контрольні питання.	133
ПІСЛЯМОВА.	135
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.	137
ДОДАТОК А Таблиця значень функції Лапласа.	140
ДОДАТОК Б Критичні точки розподілу.	141

ВСТУП

Сучасні комп'ютерні системи складаються з низки компонентів, а саме: процесори, оперативна пам'ять, пристрої вводу-виводу інформації (клавіатура, монітор, принтер, миша та ін.), модулі довготривалої пам'яті. Ці компоненти можуть працювати взаємопов'язано й незалежно один від одного, що дає змогу організувати мультипрограмний режим роботи під управлінням операційної системи. Операційна система виконує низку функцій (робіт): управління виконанням програм користувачів, управління вводом-виводом, реагування на сигнали із зовнішнього середовища і надзвичайні ситуації. Здійснення кожної з цих функцій полягає у виконанні відповідних програм, які викликаються за необхідності. Для робіт, які забезпечують базові функції операційної системи, використовують термін «процес». Поняття процесу можна розширити, включивши до нього забезпечення функцій користувачів. Тому виконання програм користувачів - процес. Для здійснення процесу може використовуватися декілька програм, і навпаки: одна програма може бути використана для відпрацювання більш ніж одного процесу. Прикладом може слугувати програма для додавання даних у список, яка може використовуватися будь-яким процесом, що оперує з чергами. До того ж в оперативній пам'яті одночасно можуть виконуватися декілька робочих програм, які ініціюють різні процеси на компонентах комп'ютерної системи з обробки інформації. Процес під час здійснення певної функції може виконуватися на різних пристроях комп'ютерної системи. У теорії операційних систем для всіх таких компонентів комп'ютерної системи використовується термін «ресурс», під яким розуміють увазі будь-який об'єкт, який може розподілятися всередині системи [1, с. 19].

Оскільки центральний процесор обчислювальної системи також є ресурсом, то і його дії стосовно виконання певної програми становить процес. Отже програма – це статичний набір інструкцій, а процес — її відпрацювання на пристроях комп'ютерної системи. Щоб запустити процес на будь-якому ресурсі, відповідна програма ініціює програмний запит. Зазвичай ресурс може обробляти (обслуговувати) в певний момент часу тільки один запит, і якщо під час обслуговування запиту від одного процесу надійде запит на обслуговування від іншого процесу, то він ставиться в чергу до відповідного ресурсу, як це показано на часовій діаграмі (рис. 1).

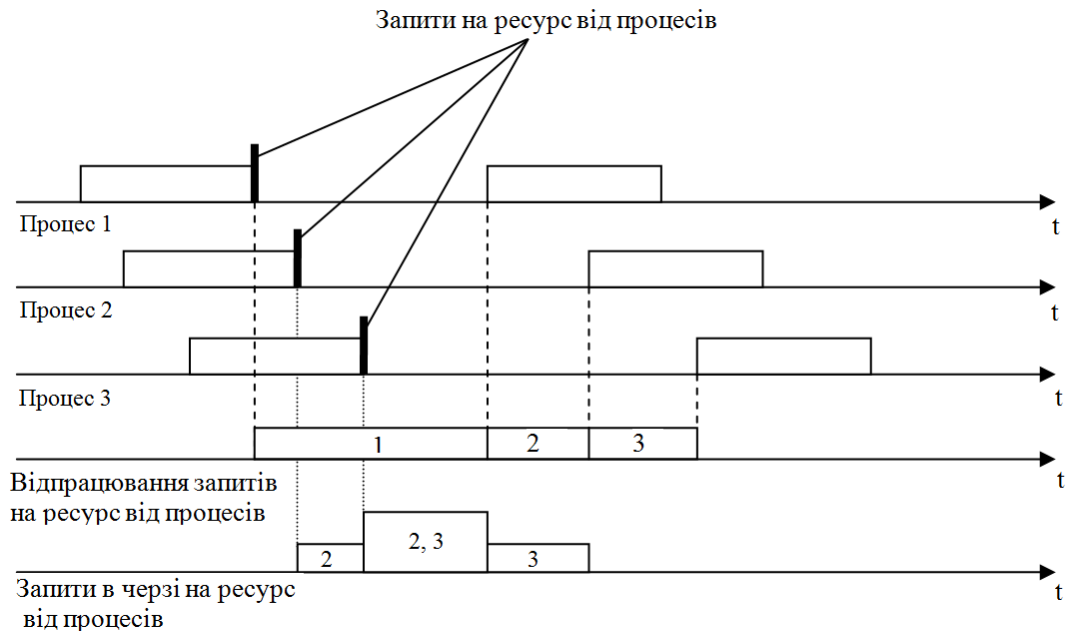


Рисунок 1 – Часова діаграма формування та відпрацювання запитів на ресурси від процесів

Наприклад, якщо надійшло кілька запитів на друк файлів до мережного принтера від комп'ютерів мережі, то він буде їх обробляти послідовно по одному, усі інші будуть очікувати в черги. Оскільки активні процеси переходять від одного ресурсу до іншого, то й до центрального процесора періодично утворюється черга процесів користувачів на відробітку. Таким чином, у комп'ютерній системі, що працює в мультипрограмному режимі, формується множина черг запитів на обслуговування до різних ресурсів (рис. 2).

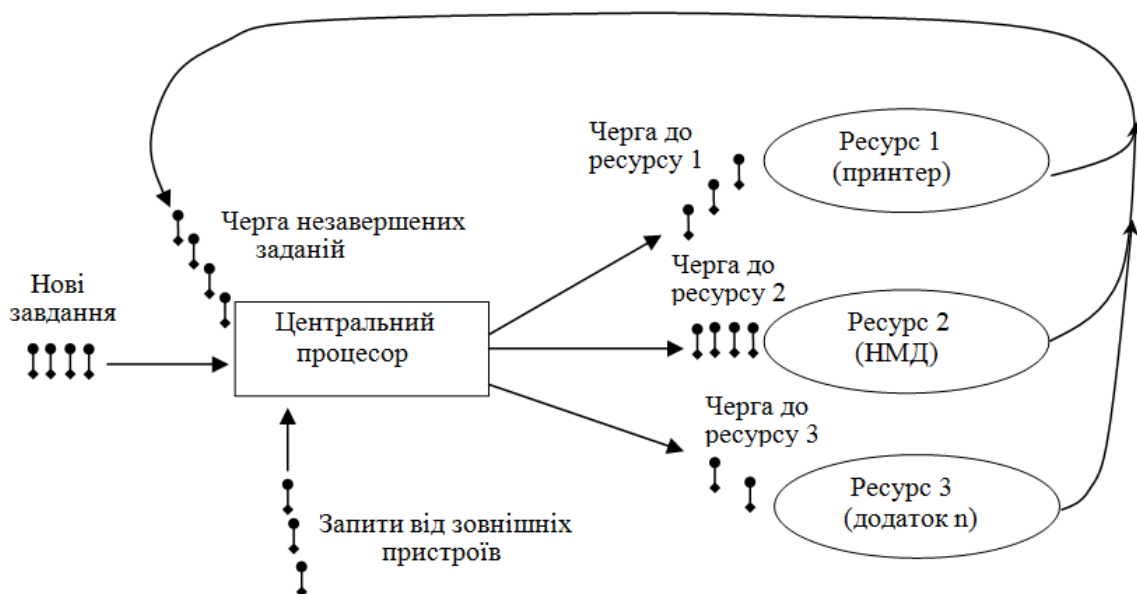


Рисунок 2 – Черги в мультипрограмній системі

Необхідно виокремити запити на обслуговування від зовнішніх пристроїв комп'ютерної системи, таких як миша, клавіатура, пристрої зв'язку з об'єктом у системах реального часу. На час їх обслуговування можуть висуватися певні обмеження, в зв'язку з чим для таких запитів можуть вводитися певні переваги в обслуговуванні – пріоритети. Для пріоритетних запитів можуть також вводитися різні рівні пріоритетності. Так, якщо накопичувач на магнітному диску виставив запит про готовність до зчитування інформації, то найменша затримка в його обслуговуванні призведе до того, що блок інформації просто проскочить повз голівку, яка зчитує інформацію з диска, що обертається.

З рисунку 1 випливає, що запити на обслуговування тим чи іншим ресурсом формуються у випадкові моменти часу; час обслуговування запитів різний і загалом є випадковим. Відповідно, для синхронізації процесів надходження запитів і їх обслуговування, а також для унеможливлення втрати запитів у комп'ютерній системі формуються різноманітні черги запитів, як показано на рисунку 2. З цього малюнка випливає, що мова йде про особливі системи, які вирізняються такими ознаками: наявність джерела запитів, одного або декількох обслуговувальних пристроїв і черги запитів, які очікують на обслуговування. У великих системах запити надходять масово, у зв'язку з цим такі системи отримали назву «системи масового обслуговування» (далі СМО). У зарубіжній літературі подібні системи називаються queuing systems.

Процеси масового обслуговування виникають у найрізноманітніших галузях техніки, науки й у повсякденному житті. Досить прийти в будь-яку поліклініку, і Ви побачите черги на прийом до спеціалістів, які можна вважати обслуговувальними пристроями. Автомобілі, які приїжджають на автозаправку, змушені ставати в чергу, очікуючи на обслуговування. В окремих випадках така черга розтягується на сотні метрів. Процеси масового обслуговування є невіддільною частиною телекомунікаційних і комп'ютерних систем, у яких за запитами від різних джерел передаються масиви інформації.

З рисунку 1 випливає, що запити на обслуговування надходять нерівномірно. Для компенсування цієї нерівномірності розробники закладають в обслуговуючі пристрої додаткові потужності. Найчастіше це практикується в системах, де надмірна затримка в обслуговуванні запитів або їх втрата призводять до катастрофічних наслідків. Прикладом може слугувати робота аеропорту щодо приймання на посадку літаків. Відповідно в окремі

моменти часу обслуговувальний пристрій завантажується максимально, в інші моменти він може простоювати. Також для організації та підтримання черги запитів потрібні відповідні ресурси. Таким чином, функціонування систем масового обслуговування пов'язано з витратами і в ряді випадків вони досить значні. У зв'язку з цим виникла гостра необхідність дослідження процесів масового обслуговування, як наслідок сформувалася наукова дисципліна «Теорія систем масового обслуговування» (далі ТСМО) і відповідна термінологія. Приміром запит на обслуговування може називатися заявкою або вимогою, прилад обслуговування – каналом. Сама система масового обслуговування може складатися з одного або декількох каналів обслуговування. Приміром, якщо комп'ютери на фірмі обслуговує один інженер, а джерелами запитів є оператори, комп'ютери які дають збій або взагалі вийшли з ладу, то така СМО є одноканальною. Якщо обслуговуванням комп'ютерів на фірмі займається кілька інженерів, то ми маємо багатоканальну СМО.

Теорія систем масового обслуговування вирішує, насамперед, завдання аналізу. Робота системи масового обслуговування залежить від виду та внутрішньої структури потоку заявок, що надходять на обслуговування, кількості каналів обслуговування і закону розподілу часу обслуговування заявок кожним каналом. Крім цього, на роботу системи значно впливають алгоритми обслуговування. Алгоритми обслуговування містять усі особливості поведінки заявки, що надійшла в систему. До таких алгоритмів належать, насамперед, правила утворення черги із заявок, які потрібно обслуговувати, і алгоритми вибору заявок із черги. Можуть розглядатися такі правила, коли черга взагалі не допускається. В інших системах масового обслуговування допускається утворення обмеженої (або необмеженої) черги заявок. На особливу увагу заслуговують системи масового обслуговування з «нетерплячими» заявками, коли заявка може покинути чергу (або взагалі систему), не дочекавшись закінчення обслуговування. Найтипівішим прикладом системи масового обслуговування з «нетерплячими» заявками є система протиповітряної оборони: літак (заявка) прагне вилетіти із зони обстрілу протиповітряної оборони, «не дочекавшись» обслуговування» (ураження) [2, с. 5]. Існують і такі правила, коли заявки певного виду повинні обслуговуватися без черги (мати пріоритет в обслуговуванні) або на певний час переривати обслуговування поточної заявки. Прикладом такої системи може слугувати робота операційної системи, що реагує на натискання оператором комп'ютера клавіш клавіатури.

Вибирати заявки з черги можна по-різному. Широко застосовується алгоритм «першим прийшов – першим буде обслужений». Зазвичай його називають FIFO – first input first output. У системах переривання операційних систем практикується алгоритм «останнім прийшов – першим буде обслужений». Його називають LIFO – last input first output, він реалізується за допомогою стека. У деяких системах практикується порядок випадкового вибору заявок із черги.

Набагато складнішими є завдання синтезу. Розробнику систем масового обслуговування доводиться вирішувати низку складних проблем. Насамперед це кількість каналів обслуговування за певної продуктивності кожного каналу. Надмірна кількість каналів призводить до їх простоювання протягом тривалих проміжків часу. Канали обслуговування також можуть виходити з ладу, і це доводиться враховувати під час при розробленні подібних систем. Під час організації черги виникає проблема щодо її розташування й визначення максимальної кількості місць для очікуючих заявок. Передбачається, що заявка, яка надійшла й яка застала чергу зайнятої, втрачається. Теоретично допускається існування нескінченної черги заявок, але на практиці досягти це неможливо, тому розробники повинні передбачати певну ймовірність втрати заявок. У некритичних системах це можливо. Приміром глядач, якому не вдалося придбати квитки на черговий сеанс в кінотеатр, може залишитися вдома або піти в найближче кафе, щоб позбутися стресу.

Окремо розглядаються питання синтезу пріоритетних систем масового обслуговування. Важливо визначити пріоритети. Розрізняють системи з відносними, абсолютними, динамічними та змішаними пріоритетами. При відносних пріоритетах заявка обслуговується до кінця, незважаючи на наявність в черзі пріоритетніших заявок. Абсолютний пріоритет передбачає: якщо надійшла пріоритетніша заявка, то припиняється обслуговування менш пріоритетної заявки. Важливо обрати кількість рівнів пріоритету і призначити пріоритети окремими джерелами заявок.

Ще складніші проблеми виникають тоді, коли декілька систем масового обслуговування об'єднуються в мережу масового обслуговування. Розвиток цього напрямку спричинило створення складних комп'ютерних систем, зокрема й Інтернету. У таких системах заявки обслуговуються на декількох вузлах мережі, можуть змінювати маршрути й алгоритми обслуговування.

Графічно потік запитів на обслуговування можна подати у вигляді точок на часовій осі, як показано на рисунку 3.

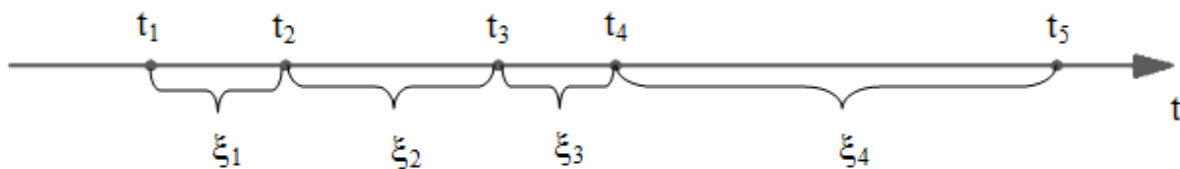


Рисунок 3 — Моменти надходження запитів на часовій осі

ξ_1, ξ_2, ξ_3 – це інтервали часу між надходженням сусідніх запитів. Вони є випадковими величинами, які характеризуються математичним сподіванням, дисперсією, функцією розподілу й іншими характеристиками. Тобто під час вивчення потоків запитів необхідно застосовувати методи теорії ймовірностей і математичної статистики. Те саме стосується і часу обслуговування запитів.

На рисунку 4 зображений одночасний процес надходження запитів на обслуговування (позначено точками t_{ni} на осі t_n), процес закінчення обслуговування запитів (позначено точками t_{oi} на осі t_o) і змінювання кількості запитів у системі L_s для однолінійної СМО.

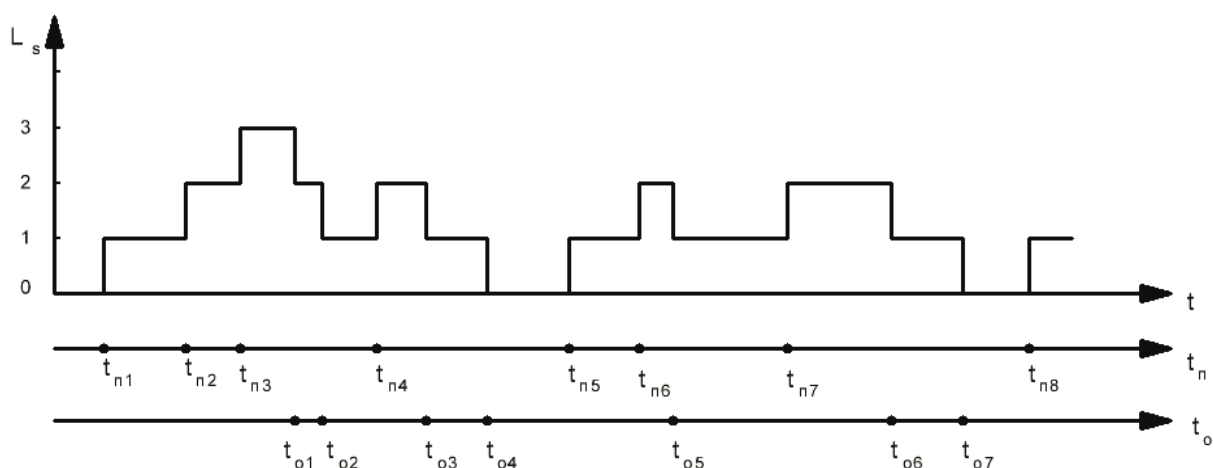


Рисунок 4 – Імовірнісні процеси в однолінійній СМО

З рисунка 4 зрозуміло, що всі процеси в системах масового обслуговування є ймовірнісними, для їхнього дослідження необхідно використовувати теорію випадкових процесів.

Теорія масового обслуговування становить новий напрям у теорії ймовірностей, що сформувався як самостійна наукова дисципліна, внаслідок специфіки застосовуваного математичного апарата й важливості розв'язуваних практичних завдань.

Розроблення практичних завдань масового обслуговування започаткував співробітник Копенгагенської телефонної компанії данський

математик А. К. Ерланг (1878–1929 рр.) у період із 1908 по 1922 роки. У 1909 р. було опубліковано його робота «Теорія ймовірностей і телефонні переговори» та інші, у яких сформульовано перші прикладні завдання телефонії. Ці завдання були пов'язані з необхідністю впорядкувати роботу телефонної мережі й розробити методи оцінювання якості обслуговування споживачів залежно від кількості використовуваних пристроїв. Значний внесок зробив Л. Клейнрок, який використовував ТСМО для дослідження обчислювальних систем [3].

Питання узагальнення методів вирішення різноманітних завдань і розроблення загальної теорії масового обслуговування досліджував радянський математик А. Я. Хінчин. У книзі «Математичні методи теорії масового обслуговування» вперше сформульовані загальні ідеї і методи теорії [4]. Подальший розвиток теорії масового обслуговування пов'язаний з ім'ям радянських математиків Б. В. Гнеденко, І. Н. Коваленко [5], А. Н. Колмогорова, Н. П. Бусленка та ін. Серед зарубіжних авторів потрібно назвати Д. Кендалла, Ф. Паллачека, та Л. Такача.

Автор цього посібника також вніс посильний вклад в розвиток теорії масового обслуговування та її застосування. Приміром, досліджені процеси обслуговування з подвійною обробкою запитів на одному приладі [6] та застосування ТСМО для розрахунку параметрів пристрою зв'язку з об'єктом комп'ютерних систем, що працюють в реальному масштабі часу [7].

1 ЗАСАДИ ТЕОРІЇ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ

Теорія систем масового обслуговування – це комплексна дисципліни, яка базується на низці класичних математичних дисциплін, а саме: теорія ймовірностей, математична статистика, теорія випадкових процесів. Процеси масового обслуговування можуть описуватися системами лінійних рівнянь або системами диференціальних рівнянь. Із огляду на це, дослідник СМО повинен бути обізнаним із цими дисциплінами.

1.1 Вступ до теорії ймовірностей та математичної статистики

В підрозділі коротко наведено базові визначення та формули теорії ймовірностей і математичної статистики. Повний матеріал можна знайти в класичних підручниках, наприклад в навчальному посібнику В. Е. Гмурмана [8]. Аксиоматичні засади теорії ймовірностей викладені в фундаментальній праці академіка А. Н. Колмогорова [9].

1.1.1 Базові поняття теорії ймовірностей

Теорія ймовірностей у абстрактній формі відображає закономірності, притаманні випадковим подіям (явищам) масового характеру. Уявлення про випадкові події складається на підставі життєвого досвіду. Приміром, ми вважаємо, що випадковими є такі події, як виграш в лотерею, відмова приладу, землетрус. Отже на противагу події достовірній, ми називаємо подію **випадковою**, якщо заздалегідь не можна гарантувати її настання, тобто якщо подія характеризується тільки тим, що вона можлива. **Предметом теорії ймовірностей** є вивчення ймовірнісних закономірностей масових випадкових явищ.

Базовими поняттями теорії ймовірностей є **ймовірнісний експеримент** і **подія**. Імовірнісним експериментом (випробуванням) називається забезпечення певної сукупності умов, які або створюються штучно, або виникають незалежно від волі експериментатора, результатом чого є настання однієї події з безлічі можливих. Нехай вироби випускаються партіями по n штук в кожній. Перевірка якості виробу призводить до його руйнування, тому для перевірки партії на якість відбирають m виробів ($m < n$). Експеримент полягає у виборі виробів з партії і їх перевірці. Подією (результатом експерименту) є кількість дефектних виробів.

Події в теорії ймовірностей прийнято позначати прописними буквами латинського алфавіту: A , B , C тощо. Події класифікуються за низкою ознак: сумісні й несумісні, протилежні, рівноможливі, елементарні [8, с. 17]. Над

подіями вводяться дві основні операції: додавання (об'єднання) і множення (перетин). **Сумою** або об'єднанням подій A і B називається подія, яка полягає

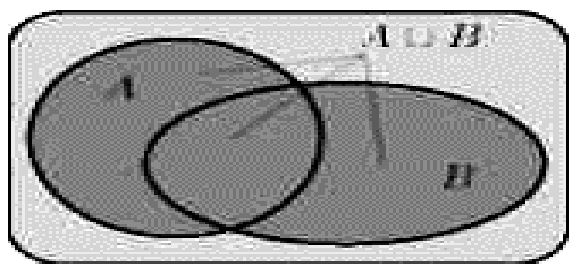


Рисунок 1.1 – Сума двох подій

в тому, що сталося або A , або B , або обидві події одночасно. Позначаються вони так: $A \cup B$ або $A + B$. Сума складених подій складається з елементарних подій, кожна з яких належить хоча б одній з цих подій, як це зображено на рисунку 1.1.

Приклад. Подія A — при киданні кубика випало парне кількість очок, а подія B — випало не більше двох очок.

Тоді $A + B = \{w_1, w_2, w_4, w_6\}$.



Рисунок 1.2 – Добуток двох подій

Добутком, або **перетином** двох подій називається подія, яка полягає в спільній появі цих подій. Позначається: $A \cap B$ або AB . Добуток двох подій складається з елементарних подій, кожна з яких належить обом подіям, як це зображено на рисунку 1.2.

Приклад. Подія A – при киданні кубика випало парне кількість очок, а подія B – випало не менше двох очок. Тоді $AB = \{w_1, w_3, w_5\} \cdot \{w_2, w_3, w_4, w_5, w_6\} = \{w_3, w_5\}$

Елементарні події повної групи попарно несумісних подій називаються **сприятливими** щодо здійснення складеної події A , якщо здійснення будь-якої з цих подій спричиняє здійснення події A .

Приклад. При підкиданні грального кубика події w_5, w_6 сприяють здійсненню події B – випадінню не менше п'яти очок. З останнім визначенням пов'язане класичне визначення ймовірності.

Ймовірністю події A називається відношення кількості m елементарних подій, що сприяють події A , до кількості n всіх елементарних подій, що утворюють повну групу подій:

$$P(A) = \frac{m}{n}. \quad (1.1)$$

Позначення P від англійського слова probability, що означає ймовірність. Із визначення ймовірності випливають такі її наступні властивості: ймовірність достовірної події становить одиниці, ймовірність неможливої події дорівнює нулю, ймовірність випадкової події є додатним числом, яке міститься між нулем і одиницею, тобто $0 < P(A) < 1$. Таким чином, **ймовірність є мірою випадковості**. Формула (1.1) використовується для безпосереднього обчислення ймовірностей в найпростіших випадках.

Приклад 1. Кинуту гральний кубик. Знайдемо ймовірність події, якщо випало не менше трьох очок.

Розв'язок. Загальна кількість елементарних подій - $n=6$, кількість елементарних подій, що сприяють події A , $m=4$ ($\{w_3, w_4, w_5, w_6\}$), тоді ймовірність події A $P(A) = 4/6 = 0,667$.

Якщо під час обчислення ймовірності події A ніяких інших обмежень, крім умов S , не накладається, то таку ймовірність називають безумовною; якщо ж накладаються інші додаткові умови – ймовірність події A називають умовною і позначають, як $P(A/B)$ або $P_B(A)$ – ймовірність настання події A за умови, що вже настала подія B .

Безпосереднє обчислення ймовірностей застосовується вкрай обмежено і може використовуватися тільки в найпростіших випадках. Випадкові події, ймовірності яких визначені заздалегідь, поєднуються в більш складні події, для розрахування ймовірностей яких використовуються базові теореми теорії ймовірностей і низка стандартних схем комбінування випадкових подій.

Теорема добутку ймовірностей. Ймовірність добутку двох подій A і B дорівнює добутку ймовірності однієї з них і умовної ймовірності другої, обчисленої з припущенням, що перша подія вже настала:

$$P(A \cdot B) = P(A)P(B/A) = P(B)P(A/B). \quad (1.2)$$

Теорема додавання ймовірностей. Ймовірність суми двох подій A і B дорівнює сумі ймовірностей цих подій без ймовірності їх одночасної появи:

$$P(A + B) = P(A) + P(B) - P(AB). \quad (1.3)$$

Зазвичай обидві теореми доводиться застосовувати одночасно. До того ж зазвичай подія, ймовірність якої потрібно визначити, подається у вигляді суми декількох несумісних подій (варіантів цієї події), кожна з яких, зі свого боку, є добутком подій.

Приклад 2. Два стрільці здійснюють по одному пострілу в мішень. Ймовірність поразки мішені для першого стрільця дорівнює 0,8, для другого – 0,7. Визначити ймовірність того, що мішень буде вражена однією кулею.

Розв'язок. Нехай подія A – мішень вражена однією кулею, B – перший стрілок влучив у мішень, C – другий стрілок влучив у мішень; \bar{B} і \bar{C} – події, протилежні до подій B і C відповідно. Очевидно, що $A = B\bar{C} + \bar{B}C$ і $P(A) = P(B\bar{C} + \bar{B}C) = P(B\bar{C}) + P(\bar{B}C) = P(B)P(\bar{C}) + P(\bar{B})P(C)$. Оскільки $P(B) = 0,8$, $P(C) = 0,7$, то $P(\bar{B}) = 1 - 0,8 = 0,2$, $P(\bar{C}) = 1 - 0,7 = 0,3$. Отже, $P(A) = 0,8 \cdot 0,3 + 0,2 \cdot 0,7 = 0,24 + 0,14 = 0,38$.

Якщо потрібно визначити ймовірність події A , яка може статися з однією з n подій, які називаються гіпотезами й утворюють повну групу несумісних подій, то варто використати **формулу повної ймовірності**

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i). \quad (1.4)$$

Потрібно взяти до уваги, що події H_1, H_2, \dots, H_n у цьому експерименті утворюють повну групу несумісних подій і внаслідок експерименту неодмінно відбудеться тільки одна з них. Очевидно, що $P(H_1) + P(H_2) + \dots + P(H_n) = 1$.

Приклад 3. Нехай два заводи виготовляють однакові вироби, які надходять на склад і там змішуються. Імовірність браку для виробів першого заводу становить 0,1, другого – 0,2. Визначити ймовірність того, що обраний навмання виріб виявиться бракованим, якщо перший завод виготовляє 60 % виробів, а другий – 40 %.

Розв'язок. Оберемо для розгляду події: A – обраний навмання виріб забракований, H_1 – виріб виготовлений на першому заводі, H_2 – виріб виготовлений на другому заводі. За умовою завдання $P(H_1) = 0,6$; $P(H_2) = 0,4$; $P(A/H_1) = 0,1$; $P(A/H_2) = 0,2$.

Отже $P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = 0,6 \cdot 0,1 + 0,4 \cdot 0,2 = 0,14$.

Зазвичай в умовах попередньої завдання доводиться визначити ймовірність $P(H_k/A)$ того, що подія A , яка вважається такою, що відбулася, відбулася одночасно з подією H_k . Цю ймовірність визначають за **формулою Байєса**:

$$P(H_k / A) = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^n P(H_i)P(A/H_i)}. \quad (1.5)$$

Приклад 4. За умовою попередньої задачі необхідно визначити таку ймовірність: якщо обраний навмання виріб виявився забракованим, то він виготовлений на другому заводі.

Розв'язок:

$$P(H_2 / A) = \frac{P(H_2)P(A / H_2)}{P(H_1)P(A / H_1) + P(H_2)P(A / H_2)} = \frac{0,4 \cdot 0,2}{0,6 \cdot 0,1 + 0,4 \cdot 0,2} = 0,556.$$

Більшість додатків теорії ймовірностей пов'язана з випадковими величинами. Випадкова величина – це величина, яка в експерименті набуває одного зі своїх значень, до того ж заздалегідь не відомо, якого з них. Наприклад, кількість студентів, присутніх на заняттях; кількість викликів, які надійшли на телефонну станцію протягом певного проміжку часу. Загальноприйнято, що випадкові величини позначаються великими літерами, а їх можливі значення — відповідними малими літерами латинського алфавіту з індексами. Наприклад, X, Y, Z і, відповідно, x_i, y_j, z_k . Розрізняють дискретні (перервні) і безперервні випадкові величини.

Випадкова величина повністю описується законом розподілу, що встановлює однозначну відповідність серед можливих значень випадкової величини і її ймовірностей.

1.1.2 Дискретні випадкові величини

Дискретні випадкові величини можуть набувати тільки кінцевої або тільки рахункової множини значень. Наприклад, кількість бракованих виробів в перевіреній партії, кількість влучень при n пострілах. Одним із способів подання закону розподілу дискретної випадкової величини є подання його як ряду розподілу, записаного у вигляді таблиці, в якій перелічено можливі значення випадкової величини в порядку їх зростання й імовірності, що їм відповідають (табл. 1. 1).

Таблиця 1.1. – Подання дискретної випадкової величини

x	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Тут $p_i = P(X = x_i), i = 1, 2, \dots, n$, до того ж $\sum_{i=1}^n p_i = 1$, оскільки події $X = x_i (i = 1, 2, \dots, n)$ утворюють повну групу несумісних подій. Остання рівність використовується для контролю правильності побудови ряду розподілу.

Для наочності ряд розподілу часто зображують графічно у вигляді багатокутника розподілу: по осі абсцис відкладають значення випадкової

величини, по осі ординат – імовірності цих значень, а отримані точки з'єднують відрізками прямих, як це показано на рисунку 1.3.

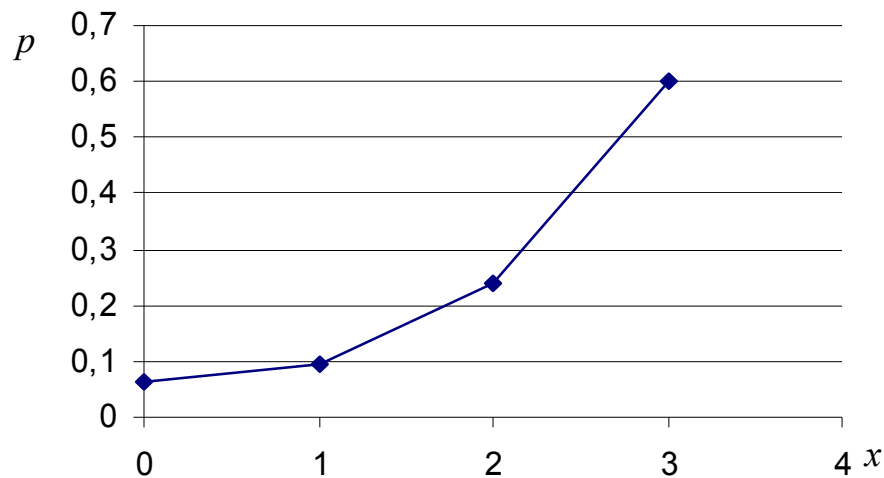


Рисунок 1.3 – Багатокутник розподілу

Подання дискретної випадкової величини у вигляді таблиці або багатокутника розподілу стає неоглядним у разі великої кількості даних, тому для опису всієї сукупності значень дискретної випадкової величини використовують ряд узагальнювальних не випадкових числових характеристик: математичне сподівання, дисперсія, середньоквадратичне відхилення, коефіцієнт варіації, мода.

Математичним сподіванням випадкової величини називається її середнє значення. Математичне сподівання дискретної випадкової величини X позначають як $M(X)$, $M[X]$, M_x , m_x або просто m , якщо відомо, до якої величини воно належить, і обчислюють за такою формулою:

$$m_x = \sum_{i=1}^n x_i p_i. \quad (1.6)$$

Дисперсією випадкової величини X називається математичне сподівання квадрата відхилення випадкової величини щодо математичного сподівання. Дисперсія випадкової величини X позначають як D_x або просто D , якщо відомо, до якої величини воно відноситься, і обчислюють за такою формулою:

$$D_x = \sum_{i=1}^n (x_i - m_x)^2 p_i. \quad (1.7)$$

Після нескладних перетворень отримують ще один вираз для обчислення дисперсії:

$$D_x = \sum_{i=1}^n x_i^2 p_i - m_x^2 = \mu_{2x} - m_x^2, \quad (1.8)$$

де $\mu_{2x} = \sum_{i=1}^n x_i^2 p_i$ – другий початковий момент – ще одна характеристика дискретної випадкової величини X , яка використовується в складних наукових додатках. Вираз (1.8) уможливило зменшення кількості обчислювальних операцій порівняно з виразом (1.7), що істотно при рахуванні вручну.

Дисперсія виражається в квадратах одиниці виміру випадкової величини X , що може призвести до плутанини. Наприклад, якщо випадкова величина вимірюється в метрах, то дисперсія вимірюватиметься в квадратних метрах, у яких вимірюється площа. Отже, окрім дисперсії, для оцінювання розкиду випадкової величини щодо математичного сподівання використовується середньоквадратичне відхилення, яке зазвичай позначають як σ_x або σ і обчислюють за формулами

$$\sigma_x = \sqrt{D_x} = \sqrt{\sum_{i=1}^n (x_i - m_x)^2 p_i} = \sqrt{\sum_{i=1}^n x_i^2 p_i - m_x^2} = \sqrt{\mu_{2x} - m_x^2}. \quad (1.9)$$

Примітка. У зарубіжній літературі і в системах комп'ютерної математики, наприклад Maple, замість терміна середньоквадратичне відхилення використовується термін стандартне відхилення, що має таке саме значення.

Для оцінювання ступеня розкиду випадкової величини щодо математичного очікування в безрозмірних величинах використовується коефіцієнт варіації, який обчислюють за формулою:

$$v_x = \frac{\sigma_x}{m_x}. \quad (1.10)$$

Що більше значення коефіцієнта варіації, то відповідно більший розкид і менша вирівняність випадкової величини. Зазвичай коефіцієнт варіації використовують для порівняння розсіювання двох і більше ознак, що мають різні одиниці виміру.

Модою дискретної випадкової величини X називається її найімовірніше значення.

Математичне сподівання має такі **властивості**:

1. Математичне сподівання постійної величини C дорівнює самій постійній величині: $M(C) = C$.

2. Постійний множник можна винести за знак математичного сподівання: $M(CX) = CM(X)$.

3. Математичне сподівання алгебраїчної суми двох випадкових величин дорівнює алгебраїчній сумі математичних сподівань відповідних випадкових величин: $M(X \pm Y) = M(X) \pm M(Y)$.

4. Математичне сподівання добутку двох випадкових величин дорівнює добутку математичних сподівань відповідних випадкових величин: $M(X \cdot Y) = M(X) \cdot M(Y)$.

Дисперсія має такі властивості:

1. Дисперсія постійної величини C дорівнює нулю: $D_x(C) = 0$.

2. Постійний множник виноситься в квадраті за знак дисперсії: $D(CX) = C^2 D(X)$.

3. Дисперсія алгебраїчної суми двох випадкових величин дорівнює сумі дисперсій відповідних випадкових величин: $D(X \pm Y) = D(X) + D(Y)$.

4. Якщо до випадкової величини алгебраїчно додати константу, то дисперсія не зміниться: $D(X \pm C) = D(X)$.

1.1.3 Неперервні випадкові величини

Неперервні випадкові величини можуть набувати будь-якого значення з деякого інтервалу. Наприклад, величини відхилення розмірів деталей від номінальних, помилка зважування тіла на аналітичних вагах, швидкість рухомого тіла в певний момент часу. Неперервні випадкові величини описуються або функцією розподілу, або щільністю розподілу.

Функція розподілу випадкової величини X – це функція $F(x)$, яка дорівнює ймовірності того, що ця випадкова величина в експерименті набуде значення менше за x , тобто

$$F(x) = P(X < x). \quad (1.11)$$

Функція розподілу має такі властивості, що випливають з її визначення: $0 \leq F(x) \leq 1$, тобто її значення належать відрізьку $[0; 1]$. $F(+\infty) = 1$.

$F(-\infty) = 0$. Функція розподілу – неспадна функція.

Щодо останнього, із властивостей функції розподілу випливає: ймовірність того, що випадкова величина X набуде значення з інтервалу (a, b) , дорівнює приросту функції розподілу на цьому інтервалі:

$$P(a \leq X < b) = F(b) - F(a). \quad (1.12)$$

Типовий вигляд графіка функції розподілу зображено на рисунку 1.4.

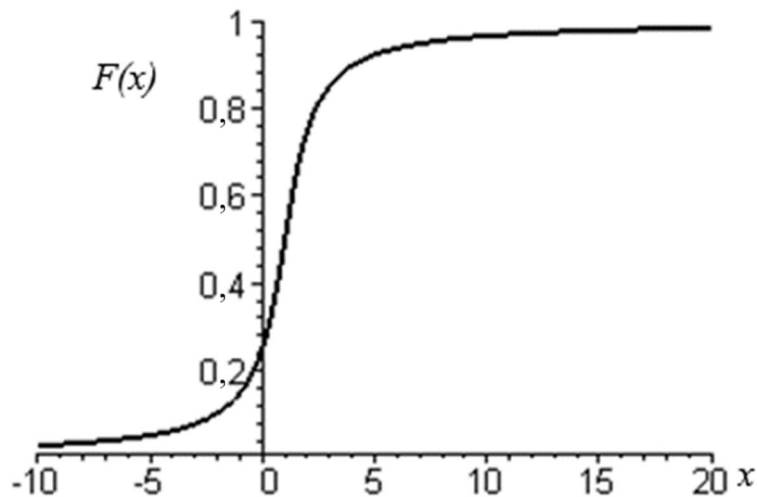


Рисунок 1.4 – Типовий вигляд графіка функції розподілу

Щільність розподілу неперервної випадкової величини X – це похідна від її функції розподілу: $f(x) = F'(x)$. Звідси, $F(x) = \int_{-\infty}^x f(x)dx$. Типовий вигляд графіка щільності розподілу неперервної випадкової величини (логнормальний розподіл) зображено на рисунку 1.5.

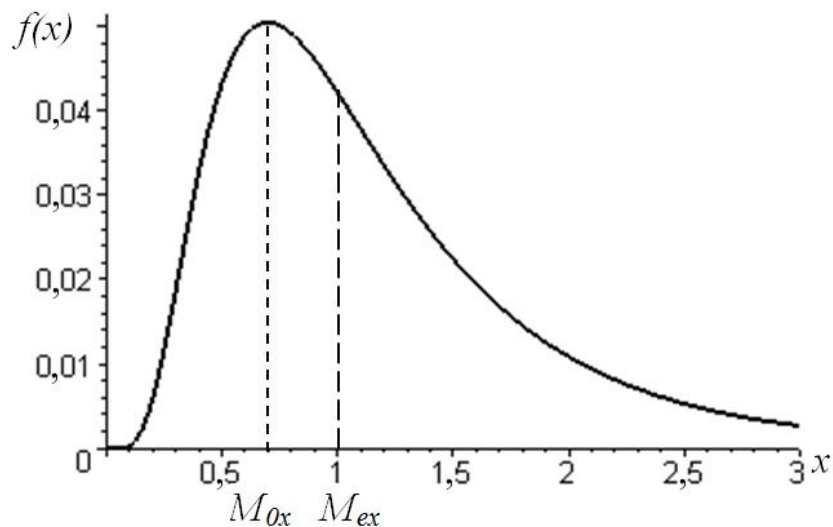


Рисунок 1.5 – Типовий вид графіка щільності розподілу

Із властивостей щільності розподілу випливає: ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу (a, b) , дорівнює визначеному інтегралу від щільності розподілу, обраному в межах від a до b :

$$P(a < x < b) = \int_a^b f(x)dx . \quad (1.13)$$

Звідси, площа криволінійної трапеції, обмеженою віссю $0, x$ і кривою щільності розподілу, дорівнює одиниці.

Як і для дискретних випадкових величин, для неперервних широко використовуються узагальнювальні не випадкові числові характеристики: математичне сподівання, дисперсія, середньоквадратичне відхилення, коефіцієнт варіації, які мають те саме значення.

Математичне сподівання обчислюється за формулою

$$m = \int_{-\infty}^{\infty} xf(x)dx. \quad (1.14)$$

Дисперсія неперервної випадкової величини X обчислюється за формулою:

$$D = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx. \quad (1.15)$$

Середньоквадратичне відхилення й коефіцієнт варіації обчислюються аналогічно до обчислення дискретних випадкових величин: $\sigma = \sqrt{D}$, $v = \sigma / m$.

Варто зауважити, що для більшості законів розподілу вирази для цих характеристик уже знайдені й наведені у відповідних підручниках і довідниках, наприклад у [8, 10].

Хоча потенційно можливих законів розподілу дуже багато, тільки їх невелика частина широко використовується на практиці. Особливо здебільшого використовуються закони нормального, показового й рівномірного розподілів.

Нормальним називають розподіл ймовірностей неперервної випадкової величини X , який описується щільністю розподілу

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}. \quad (1.16)$$

Числові характеристики: математичне сподівання – $m_x = a$, дисперсія – $D_x = \sigma^2$, середньоквадратичне відхилення – σ , мода – $M_{ox} = a$, медіана – $M_{ex} = a$, коефіцієнт асиметрії – $As_x = 0$, коефіцієнт ексцесу – $Ex = 0$.

Графік щільності нормального розподілу називають нормальною кривою (крива Гауса). На рисунку 1.6. зображено графіки щільності нормального розподілу при різних значеннях математичного сподівання – параметр – a і середньоквадратичного відхилення – параметр σ .

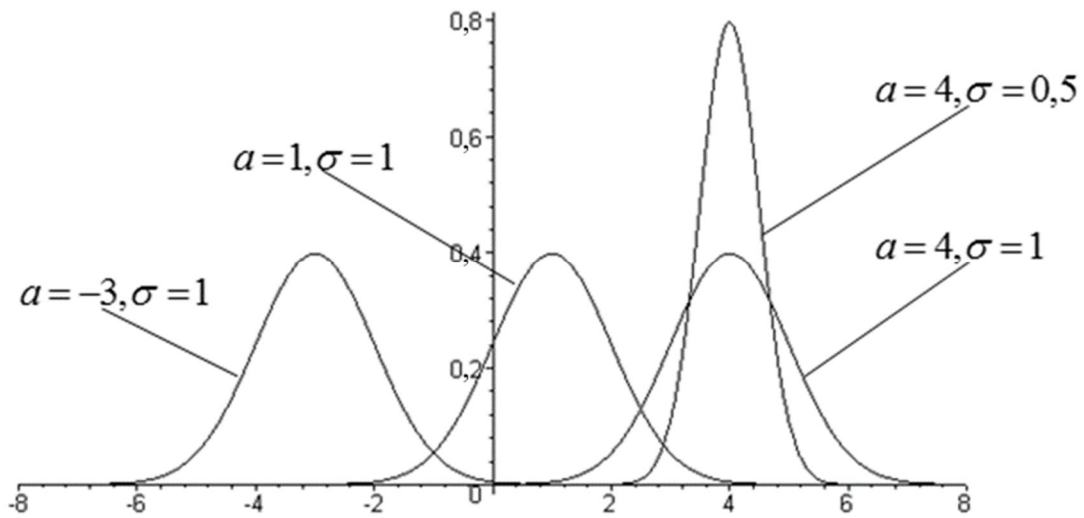


Рисунок 1.6 – Крива Гауса

Для нормально розподіленої випадкової величини X імовірність того, що X набуде значення, яке належить до інтервалу (α, β)

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-(x-a)^2/2\sigma^2} dx = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right), \quad (1.17)$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ – функція Лапласа, значення якої наведено в додатку А.

Експоненціальним (показниковим) називають розподіл ймовірностей неперервної випадкової величини X , яка описується щільністю розподілу:

$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0, \end{cases} \quad (1.18)$$

де λ – постійна додатна величина.

Функція розподілу експоненціального закону

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases} \quad (1.19)$$

Графіки щільності розподілу і функції розподілу для експоненціального закону зображено на рисунку 1.7.

Числові характеристики: математичне сподівання – $m_x = 1/\lambda$, дисперсія – $D_x = 1/\lambda^2$, середньоквадратичне відхилення – $1/\lambda$, мода – $M_{0x} = 0$, медіана – $M_{ex} = \ln 2/\lambda$, коефіцієнт асиметрії – $As_x = 2$, коефіцієнт ексцесу – $Ex = 6$, коефіцієнт варіації – $v_x = 1$.

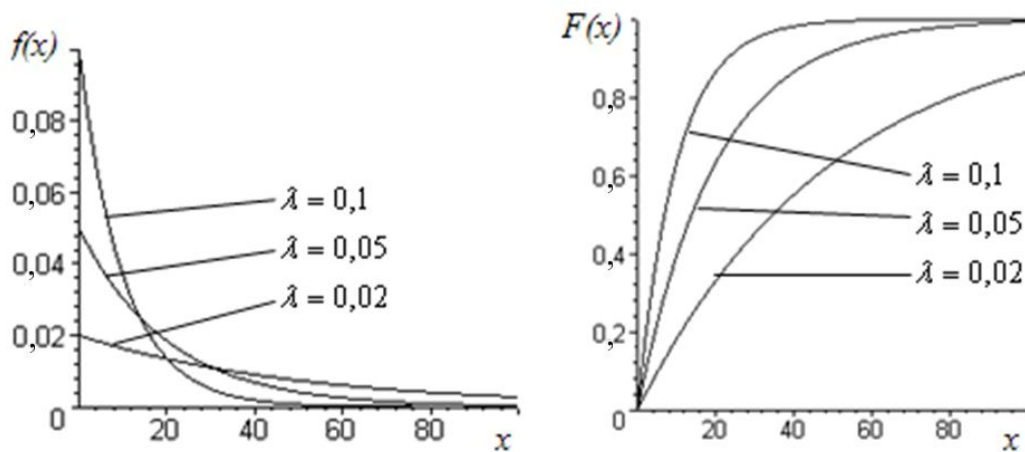


Рисунок 1.7 – Експоненціальний закон розподілу

Для експоненціально розподіленої випадкової величини X імовірність того, що X набуде значення, що належить до інтервалу (α, β) , дорівнює

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} \lambda e^{-\lambda x} dx = e^{-\lambda\alpha} - e^{-\lambda\beta}. \quad (1.20)$$

Експоненціальний розподіл є базовим в теорії масового обслуговування, і більшість досліджених моделей базується на цьому розподілі.

Рівномірним називають розподіл ймовірностей неперервної випадкової величини X , якщо на інтервалі (a, b) , до якого належать усі можливі значення випадкової величини, щільність розподілу зберігає постійне значення, а саме:

$$f(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases} \quad (1.21)$$

Функція розподілу рівномірного закону

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases} \quad (1.22)$$

Графіки щільності розподілу і функції розподілу для рівномірного закону зображено на рисунку 1.8.

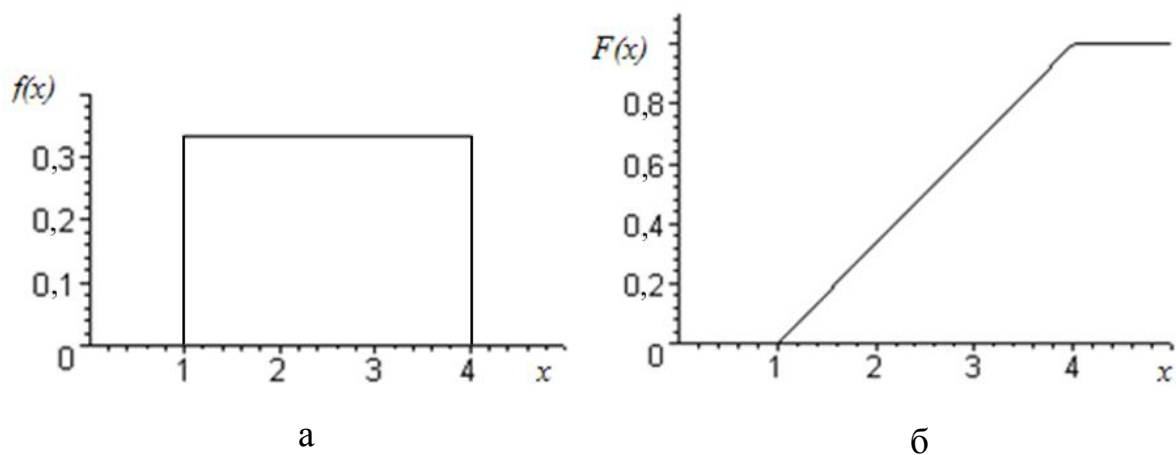


Рисунок 1.8 – Рівномірний закон розподілу:

а – графік щільності розподілу, б – графік функції розподілу

Числові характеристики: математичне сподівання – $m_x = (b + a)/2$, дисперсія – $D_x = (b - a)^2/12$, середньоквадратичне відхилення – $(b - a)/\sqrt{12}$, медіана – $M_{ex} = (b + a)/2$, коефіцієнт асиметрії – $As_x = 0$, коефіцієнт ексцесу – $Ex = -1,2$, коефіцієнт варіації – $v_x = (b - a)/\sqrt{3}(b + a)$.

Для розрахування ймовірності попадання випадкової величини X , розподіленої рівномірно, в інтервал (α, β) можна використати формулу (1.12) або (1.13).

1.1.4 Математична статистика

Математичні закони теорії ймовірностей є математичними виразами (моделями) реальних закономірностей, які існують у масових випадкових явищах. Дійсно, частота випадання герба при багаторазовому підкиданні монети за ймовірністю прагне до відповідної ймовірності:

$$p^* = \frac{m_{\text{герба}}}{n_{\text{киданій}}} \xrightarrow{\text{по ймовірності при } N \rightarrow \infty} p = \frac{m_{\text{сприятливе}}}{n_{\text{загальне}}} = \frac{1}{2}.$$

У теорії ймовірностей було використано різні закони розподілу та їх характеристики. Постає закономірне питання: на підставі чого обираються ці закони? Відповідь одна: усі закони розподілу обираються на підставі досвіду і практики. Кожне дослідження у сфері випадкових подій базується, або на експерименті або на спостереженні за випадковими подіями, внаслідок чого формуються так звані статистичні дані. Розроблення методів реєстрації, опису та аналізу статистичних експериментальних даних становить предмет спеціальної науки – математичної статистики.

Головним завданням математичної статистики є аналіз статистичних даних залежно від цілей дослідження, а саме:

а) оцінювання невідомої ймовірності події та оцінювання невідомої функції розподілу; б) оцінювання параметрів розподілу, різновид якого відомий; в) перевірка статистичних гіпотез про різновид невідомого розподілу або про величину параметрів розподілу, різновид якого відомий.

Нехай потрібно вивчити сукупність однорідних об'єктів щодо деякої якісної або кількісної ознаки, яка характеризує ці об'єкти. Певні значення ознаки у математичній статистиці називають варіантами. Наприклад, якщо є партія продукції, то якісною ознакою може бути сортність одиниці продукції або її колір, а кількісною – контрольована вага одиниці продукції. Уся сукупність об'єктів, які необхідно дослідити, називається генеральною сукупністю.

Зазвичай із генеральної сукупності відбирають обмежену кількість об'єктів і вивчають їх. Вибірковою сукупністю, або вибіркою називають сукупність випадково відібраних з генеральної сукупності об'єктів. Обсягом сукупності (вибіркової чи генеральної) називають кількість об'єктів цієї сукупності. Наприклад, якщо з 1 000 деталей для обстеження відібрано 100 деталей, то обсяг генеральної сукупності (V) дорівнює 1 000, а обсяг вибірки (N) – 100.

Під час побудови моделей за експериментальними даними в теорії масового обслуговування головну увагу приділяють виду закону розподілу випадкової величини та його параметрам. З цією метою на етапі обстеження системи будують інтервальний розподіл і так звану гістограму щільності відносних частот. Для цього інтервал, у якому укладаються всі спостережувані значення ознаки – H , розбивають на декілька часткових інтервалів m завдовжки $h = H / m$, знаходять для кожного часткового інтервалу суму частот варіанти n_i , які потрапили в i -й інтервал і w_i – відносні частоти, що дорівнюють відношенню $w_i = n_i / N$. Результати такої обробки можна подати у вигляді таблиці інтервального розподілу (табл. 1.2).

Таблиця 1.2 – Інтервальний розподіл значень ознаки

x_i, x_{i+1}	x_1, x_2	x_2, x_3	x_{m-1}, x_m
n_i	n_1	n_2	n_m
w_i	w_1	w_2	w_m

Зауважимо, що, що $w_1 + w_2 + \dots + w_m = 1$.

Гістограмою щільності відносних частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є часткові інтервали завдовжки h , а висоти дорівнюють щільностям відносних частот ρ_i , які обчислюються за формулою:

$$\rho_i = \frac{n_i}{N \cdot h}. \quad (1.23)$$

Якщо вважати ρ_i функцією від x , то це буде **емпірична щільності розподілу**, а гістограма щільності відносних частот – графіком. Для побудови гістограми щільності відносних частот на осі абсцис відкладають часткові інтервали, а над ними на відстані ρ_i проводять відрізки, паралельні осі абсцис. Площа i -го часткового прямокутника $h \cdot \frac{n_i}{N \cdot h} / h = w_i$ – відносна частота варіант, що потрапили в i -й інтервал. Отже, площа гістограми відносних частот дорівнює сумі всіх відносних частот, тобто одиниці.

Приклад 5. Зафіксовані інтервали часу, між сусідніми запитами, які надходять в систему – ξ_i . Побудувати гістограму відносних частот випадкових інтервалів часу, які увійшли до такої вибірки:

7,35 8,97 70,5 28,0 5,27 18,3 17,6 3,9 17,5 21,3 59,7
 1,73 22,4 63,1 7,82 34,2 7,39 28,3 86,2 10,2 3,30 7,70 13,9
 ,701 63,5 2,38 65,1 24,2 ,116 38,3 37,3 30,8 74,7 23,4 59,3
 20,1 25,1 76,6 10,9 20,1 1,05 25,9 5,16 67,0 33,1 21,1 11,0
 43,5 26,9 27,7 31,5 ,30 9,01 20,6 14,3 49,3 6,85 2,54 62,4
 15,5 44,7 60,4 65,8 17,5 4,58 119, 6,04 8,74 18,7 1,61 40,0
 26,8 35,8 26,6 16,0 38,5 ,389 110, 17,9 1,89 27,9 5,24 6,62
 25,6 1,41 47,4 13,1 3,25 31,6 58,1 21,0 39,7 43,9 27,6 24,0
 18,0 7,78 21,7 1,93 26,7

Розв’язок. За вибіркою знаходимо мінімальне значення варіант – $\xi_{\min} = 0,12$ і максимальне значення варіант – $\xi_{\max} = 119$. Таким чином усі варіанти містяться в проміжку $[0; 119]$ завдовжки 119. Розіб’ємо його на 10 рівних по довжині ($h = 119/10 = 11,9$) часткових півінтервалів: $[0; 11,9)$, $[11,9; 23,8)$, $[23,8; 35,7)$, $[35,7; 47,6)$, $[47,6; 59,5)$, $[59,5; 71,4)$, $[71,4; 83,3)$, $[83,3; 95,2)$, $[95,2; 107,1)$, $[107,1; 119]$.

Кожне зі значень вибірки віднесемо до одного з цих часткових півінтервалів і підсумуємо кількість варіант, які потрапили в перший, другий, третій, ..., десятий півінтервали. Отримаємо:

$$n_1 = 33, n_2 = 21, n_3 = 19, n_4 = 10, n_5 = 3, n_6 = 9, n_7 = 2, n_8 = 1, n_9 = 0, n_{10} = 2.$$

Обсяг вибірки N дорівнює 100. За формулою (1.23) розраховуємо щільності відносних частот для кожного півінтервалу:

$$\rho_1 = 0,02773; \rho_2 = 0,01765; \rho_3 = 0,01597; \rho_4 = 0,0084; \rho_5 = 0,00252;$$

$$\rho_6 = 0,00756; \rho_7 = 0,00168; \rho_8 = 0,0084; \rho_9 = 0,00; \rho_{10} = 0,00168.$$

На осі абсцис відкладемо часткові інтервали завдовжки 11,9, а над ними проводимо відрізки, паралельні до осі абсцис на відстані розрахованих ρ_i . Шукана гістограма відносних частот зображена на рисунку 1.9.

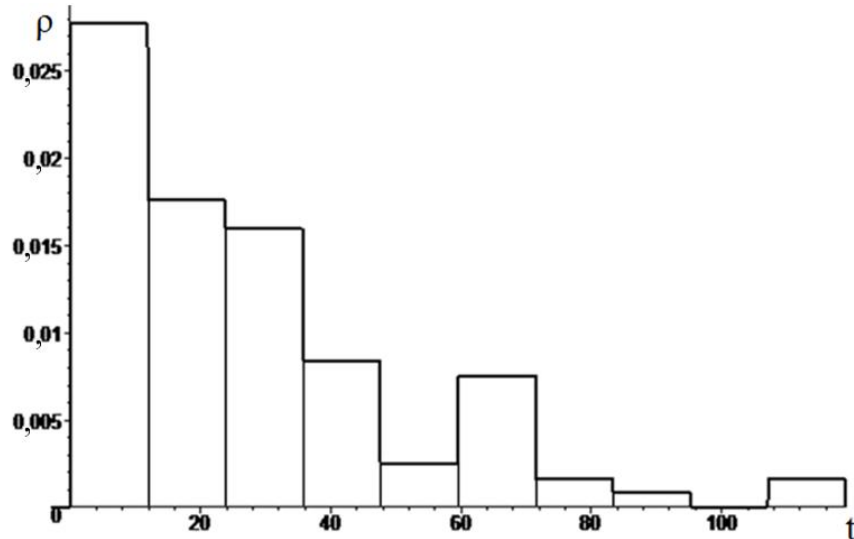


Рисунок 1.9 – Гістограма відносних частот

За виглядом гістограми відносних частот можна висунути гіпотезу про той чи інший закон розподілу, якому підпорядковуються дані вибірки. Приміром, за рисунком 1.9 можна припустити, що випадкова величина – інтервали часу між моментами надходження в систему запитів підпорядковується експоненціальному закону (див. рис. 1.7).

Побудова статистичних розподілів вибірки і їх графічне зображення – це тільки перший крок на шляху дослідження статистичного ряду даних. Наступний кроком є знаходженні числових характеристик, які узагальнено відображають найважливіші особливості статистичного розподілу вибірки. Слід враховувати, що будь-яка числова характеристика, обчислена на підставі обмеженої вибірки, містить елемент випадковості й змінюється від вибірки до вибірки. Наближене значення будь-якого параметра статистичної вибірки називається **оцінкою параметра**. Використання наближеного параметра значення замість точного призводить до погрешностей. Бажано так обирати оцінки, щоб похибки були мінімальні.

До числових характеристик вибірки висувуються певні вимоги:

незміщеність, ефективність і обґрунтованість [8, с. 198]. У теорії масового обслуговування як базові числові характеристики вибірки використовують оцінку для математичного сподівання і оцінку для дисперсії.

Незміщеною і обґрунтованою **оцінкою для математичного сподівання** є середня вибіркова, що обчислюється за формулами:

$$m^* = \bar{x} = \frac{1}{N}(x_1 + x_2 + \dots + x_N) = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1.24)$$

де N – обсяг вибірки при різних варіантах;

$$m^* = \bar{x} = \frac{1}{N} \sum_{i=1}^m n_i x_i, \quad (1.25)$$

якщо є повторювані варіанти, кількість яких m .

Незміщена й обґрунтована **оцінка для дисперсії** обчислюється за формулами:

$$D^* = \frac{1}{N-1} \sum_{i=1}^N (x_i - m^*)^2 \text{ або } D^* = \frac{1}{N-1} \sum_{i=1}^m n_i (x_i - m^*)^2. \quad (1.26)$$

Для оцінки середньоквадратичного відхилення використовується величина σ^* , яка дорівнює квадратному кореню з вибіркової дисперсії:

$$\sigma^* = \sqrt{D^*}.$$

Для порівняння декількох статистичних розподілів з погляду їх мінливості щодо вибіркової середньої використовується коефіцієнт варіації, який дорівнює відношенню оцінок для середньоквадратичного відхилення й математичного сподівання:

$$v^* = \sigma^* / m^*. \quad (1.27)$$

За результатами аналізу гістограми розподілу можна підібрати теоретичний розподіл, що описує генеральну сукупність і який залежить від одного або декількох параметрів. У ряді випадків вид розподілу відомий апіорі. Так, час між відмовами пристрої часто підпорядковане показовому розподілу, відхилення результатів вимірювань від істинного значення – нормальному. Потрібно оцінити параметри розподілу. Для цього використовується метод моментів, запропонований К. Пірсоном і метод найбільшої вірогідності, розроблений Р. Фішером [8, с. 229].

Метод моментів полягає в порівнянні початкових або центральних моментів (математичного сподівання, дисперсії і тощо) теоретичного розподілу, що є функціями параметрів розподілу, до відповідних моментів, обчислених за вибіркою, і у розв'язанні одного або декількох рівнянь щодо

параметрів розподілу.

Приклад 6. На підставі аналізу гістограми, побудованої за вибіркою прикладу 5 і зображеної на рисунку 1.9, зроблено висновок про те, що випадкова величина Ξ (ксі) – інтервали часу між моментами надходження в систему запитів підпорядковується експоненціальному закону з щільністю

розподілу $f(t) = \begin{cases} 0, & t < 0, \\ \lambda e^{-\lambda t}, & t \geq 0. \end{cases}$ За допомогою метода моментів необхідно знайти

точкову оцінку для параметра λ розподілу.

Розв'язок. Із п.1.1.3 відомо, що параметр λ експоненціального розподілу визначається, як $\lambda = 1/m$. За даними прикладу 5 і формулою (1.24) знаходимо оцінку для математичного сподівання вибірки – $m^* = 27,054$, відповідно $\lambda = \lambda^* = 1/27,054 \approx 0,037$. На рисунку 1.10 зображено графік щільності розподілу випадкової величини – інтервалів часу між моментами надходження в систему запитів з оціненим параметром, накладеним на гістограму відносних частот (рис. 1.9).

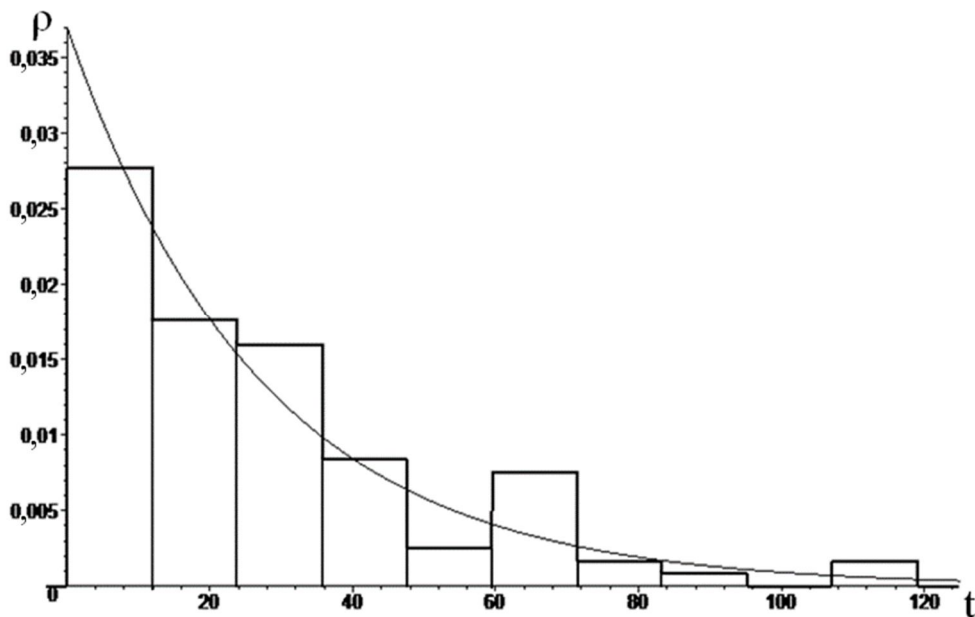


Рисунок 1.10 – Сумісний графік експериментального і теоретичного розподілів

Можна відзначити, що графіки теоретичної і експериментальної щільностей розподілу співпадають.

Як зазначалося вище, на підставі аналізу гістограми відносних частот, побудованої за даними вибірки, було зроблено припущення, що розподіл випадкової величини Ξ – інтервалів часу між моментами надходження в систему запитів підпорядковується експоненціальному закону. Такі

припущення в математичній статистиці називаються **статистичними гіпотезами**.

Статистичною називають гіпотезу про вид невідомого розподілу або про параметри відомих розподілів.

Одночасно з висунутою гіпотезою розглядають протилежну до неї гіпотезу. Якщо висунута гіпотеза буде відкинута, то актуальна протилежна гіпотеза. Із огляду на це такі гіпотези доцільно розрізняти. **Нульовою** (базовою) називають висунуту гіпотезу H_0 .

Конкуруючою (альтернативною) називають гіпотезу H_1 , яка суперечить нульовій.

Приміром, якщо нульова гіпотеза полягає в припущенні, що математичне сподівання m експоненціального розподілу дорівнює 10, то конкуруюча гіпотеза, зокрема, може полягати в припущенні, що $m \neq 10$.

Висунута гіпотеза може бути правильною або неправильною, тому виникає необхідність перевірити її. Оскільки перевірку проводять за допомогою статистичних методів, її називають статистичною. Внаслідок статистичної перевірки гіпотези у двох випадках може бути прийнято неправильне рішення, тобто можуть бути допущені помилки двох видів.

Помилка першого виду полягає в тому, що буде відкинута правильна гіпотеза, **помилка другого виду** полягає в тому, що буде прийнята неправильна гіпотеза.

Підкреслимо, що наслідки цих помилок можуть бути різними. Приміром, якщо відкинута правильне рішення «продовжувати будівництво житлового будинку», то ця помилка першого виду спричинить матеріальну шкоду; якщо ж прийнято неправильне рішення «продовжувати будівництво» незважаючи на небезпеку обвалу будівництва, то ця помилка другого виду може спричинити загибель людей. Можна навести приклади, коли помилка першого виду приводить до більш важких наслідків порівняно з помилкою другого виду.

Правильне рішення може бути прийнято також у двох випадках:

- 1) гіпотеза приймається, до того ж і в дійсності вона правильна;
- 2) гіпотеза відкидається, до того ж і в дійсності вона є неправильною.

Якщо базова гіпотеза містить припущення про параметри відомого розподілу, то статистичний критерій її перевірки називається параметричним. Якщо в гіпотезі йдеться про невідомий розподіл, то відповідний критерій називається критерієм згоди.

Визначення. Імовірність припуститися помилки першого виду прийнято

позначати через α ; її називають рівнем значущості. Здебільшого приймають рівні значущості такі: 0,1; 0,05; 0,01 або 0,001. Якщо, приміром, прийнятий рівень значущості дорівнює 0,05, то це означає, що в п'яти випадках зі ста є ризик припуститися помилки першого виду (відкинути правильну гіпотезу).

Для перевірки нульової гіпотези використовують спеціально підібрану випадкову величину, точний або наближений розподіл якої відомий. Цю величину позначають через U або Z , якщо вона розподілена нормально, F або ν^2 – за законом Фішера–Снедекора, T – за законом Стьюдента, χ^2 – згідно із законом «хі-квадрат» тощо. Для спрощення позначимо цю величину через K .

Статистичним критерієм називають випадкову величину K , яку використовують для перевірки нульової гіпотези.

Для перевірки гіпотези за даними вибірок обчислюють часткові значення величин, які належать до критерію і таким чином отримують часткове (що спостерігається) значення критерію.

Спостережуваним значенням ($K_{\text{спост}}$) називають значення статистичного критерію, обчислене за вибірками. Приміром, якщо за двома вибірками знайдено виправлені вибіркові дисперсії $s_1^2 = 20$ і $s_2^2 = 5$, то спостережуване значення критерію F , що дорівнює відношенню двох дисперсій – $F_{\text{спост}} = s_1^2 / s_2^2 = 20/5 = 4$.

Після вибору певного критерію множину всіх його можливих значень розбивають на дві непересічні підмножини: одна з них містить значення критерію, при яких нульова гіпотеза відкидається, а інша – при яких вона приймається.

Критичною областю називають сукупність значень критерію, при яких нульову гіпотезу відкидають.

Областю прийняття гіпотези (областю допустимих значень) називають сукупність значень критерію, при яких гіпотезу приймають.

Головний принцип перевірки статистичних гіпотез можна сформулювати так: якщо спостережуване значення критерію належить до критичної області, гіпотезу відкидають, якщо до області прийняття гіпотези – приймають. Оскільки критерій K – одномірна випадкова величина, усі її можливі значення належать до деякого інтервалу. Отже критична область і область прийняття гіпотези також є інтервалами, тому існують точки, які їх розділяють.

Критичними точками (границями) k_{kp} називають точки, що відокремлюють критичну область від області прийняття гіпотези.

Розрізняють однобічну (права однобічна або ліва однобічна) та двобічну критичні області.

Правою однобічною називають критичну область, яка визначається нерівністю $K > k_{кр}$, де $k_{кр}$ – додатне число (рис. 1.11, а).

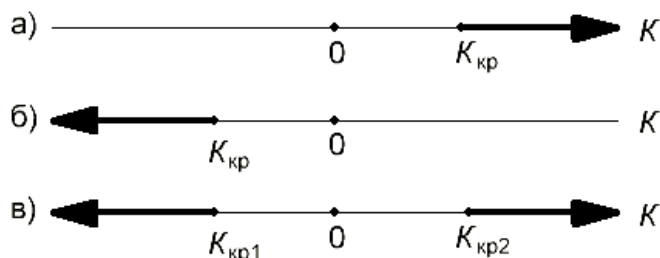


Рисунок 1.11 – Різновиди критичних областей

Лівобічною називають критичну область, яка визначається нерівністю $K < k_{кр}$, де $k_{кр}$ – від’ємне число (рис. 1.11, б).

Однобічною називають правобічну або лівобічну критичну область.

Двобічною називають критичну область, яка визначається нерівностями $K < k_{п1}, K > k_{п2}$, де $k_{п2} > k_{п1}$ – критичні точки (рис. 1.11, в).

Зокрема, якщо критичні точки симетричні відносно нуля, двобічна критична область (із припущенням, що $k_{кр} > 0$): визначається нерівностями або рівносильною нерівністю $|K| > k_{кр}$.

Знаходження критичної області розглянемо на прикладі використання правобічної критичної області. Насамперед задають достатньо малу ймовірність – рівень значущості. Потім шукають критичну точку $k_{кр}$ виходячи з вимоги, що за умови справедливості нульової гіпотези ймовірність того, що критерій K набуде значення більше за $k_{кр}$, буде дорівнювати прийнятому рівню значущості $P\{K > k_{кр}\} = \alpha$.

Для кожного критерію складено відповідні таблиці, за якими й знаходять критичну точку, що задовольняє ці вимоги.

Коли критичну точку вже знайдена, за даними вибірки обчислюють спостережуване значення критерію $K_{спост}$ і, якщо виявиться, що $K_{спост} > K_{кр}$, нульову гіпотезу відкидають; якщо ж $K_{спост} < K_{кр}$, підстав для відкидання нульової гіпотези немає.

Відшукування лівобічної і двобічної критичних областей зводиться (як і для правобічної) до знаходження відповідних критичних точок і перевірки, у якій області розташовується $K_{спост}$.

Зазвичай статистичній перевірці підлягає гіпотеза про розподіл

генеральної сукупності $F(x)$. Статистичні критерії для перевірки цієї гіпотези мають спеціальне найменування – **критерії згоди**. Один з таких критеріїв – критерій χ^2 (хі-квадрат), запропонований американським статистиком К. Пірсоном, назване його ім'ям [8, с. 329].

У цьому критерії мірою розбіжності розподілу, обраного за гіпотезою (теоретичний розподіл), щодо реального слугує величина $\chi_{\text{спост}}^2$:

$$\chi_{\text{спост}}^2 = \sum_{i=1}^m \frac{(n_i - Np_i)^2}{Np_i} = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i}, \quad (1.28)$$

де $n'_i = Np_i$ – теоретичні частоти значень випадкової величини; n_i – емпіричні частоти значень випадкової величини; m – кількість різних значень або груп значень випадкової величини.

Величина $\chi_{\text{спост}}^2$ – випадкова величина, що має розподіл хі-квадрат з числом ступенів свободи k :

$$k = m - s - 1, \quad (1.29)$$

де m – кількість груп емпіричного розподілу; s – кількість параметрів теоретичного розподілу оцінених за емпіричним розподілом.

За обраним рівнем значущості α і числом ступенів свободи $k = m - s - 1$ за додатком Б критичних точок розподілу χ^2 знаходять значення $\chi_{\text{кр}}^2$, при якому $P\{\chi_{\text{спост}}^2 > \chi_{\text{кр}}^2\} = \alpha$. Якщо виявиться, що обчислене за вибіркою значення величини $\chi_{\text{спост}}^2$ потрапило в область допустимих значень, тобто $\chi_{\text{спост}}^2 < \chi_{\text{кр}}^2$, то вважається, що вибірка підтверджує гіпотезу про те, що генеральна сукупність має функцію розподілу $F(x)$. Якщо ж чисельне значення $\chi_{\text{спост}}^2$ опиниться в критичній області, тобто $\chi_{\text{спост}}^2 > \chi_{\text{кр}}^2$, то сформульована гіпотеза відкидають.

Приклад 7. При рівні значущості 0,05 перевірити гіпотезу про те, що за даними вибірки прикладу 5 розподіл випадкової величини Ξ – інтервали часу між моментами надходження в систему запитів підпорядковується експоненціальному закону.

Розв'язок. У прикладі 5 було отримано параметр експоненціального закону $\lambda = 0,037$. Знайдемо теоретичні частоти, скориставшись формулою (1.20) розрахування ймовірності потрапляння випадкової величини, розподіленої за експоненціальним законом, в інтервал:

$$n'_i = N \cdot P(a_{i-1} < \Xi < a_i) = N(\exp(-\lambda a_{i-1}) - \exp(-\lambda a_i)).$$

Отримаємо:

$$\begin{aligned}
n'_1 &= N \cdot P(0 < \Xi < 11,9) = 100 \cdot (\exp(-0,037 \cdot 0) - \exp(-0,037 \cdot 11,9)) \approx 35,6; \\
n'_2 &= N \cdot P(11,9 < \Xi < 23,8) = 100 \cdot (\exp(-0,037 \cdot 11,9) - \exp(-0,037 \cdot 23,8)) \approx 22,9; \\
n'_3 &= N \cdot P(23,8 < \Xi < 35,7) \approx 14,7; n'_4 = N \cdot P(35,7 < \Xi < 47,6) \approx 9,5; \\
n'_5 &= N \cdot P(47,6 < \Xi < 59,5) \approx 6,1; n'_6 = N \cdot P(59,5 < \Xi < 71,4) \approx 3,9; \\
n'_7 &= N \cdot P(71,4 < \Xi < 83,3) \approx 2,5; n'_8 = N \cdot P(83,3 < \Xi < 95,2) \approx 1,6; \\
n'_9 &= N \cdot P(95,2 < \Xi < 107,1) \approx 1,1; n'_{10} = N \cdot P(107,1 < \Xi < 119) \approx 0,7.
\end{aligned}$$

Порівняємо емпіричні й теоретичні частоти, використовуючи критерій Пірсона:

а) обчислимо спостережуване значення критерію Пірсона:

$$\begin{aligned}
\chi^2_{\text{спост}} &= \sum_{i=1}^{10} \frac{(n_i - n'_i)^2}{n'_i} = \frac{(33 - 35,6)^2}{33,6} + \frac{(21 - 22,9)^2}{22,9} + \frac{(19 - 14,7)^2}{14,7} + \frac{(10 - 9,5)^2}{9,5} + \\
&+ \frac{(3 - 6,1)^2}{6,1} + \frac{(9 - 3,9)^2}{3,9} + \frac{(2 - 2,5)^2}{2,5} + \frac{(1 - 1,6)^2}{1,6} + \frac{(0 - 1,1)^2}{1,1} + \frac{(2 - 0,7)^2}{0,7} \approx 13,72;
\end{aligned}$$

б) за формулою (1.29) визначимо кількість ступенів свободи:
 $k = 10 - 1 - 1 = 8$;

в) за таблицею додатка Б критичних точок розподілу χ^2 для рівня значущості $\alpha = 0,05$ і кількості ступенів свободи $k = 8$ знайдемо критичну точку правобічною критичної області: $\chi^2_{kp}(0,05; 8) = 15,5$.

Оскільки $\chi^2_{\text{спост}} < \chi^2_{kp}$, то прийmemo гіпотезу про експоненціальний (показниковий) розподіл випадкової величини Ξ – інтервали часу між моментами надходження в систему запитів на обслуговування.

1.2 Засади теорії випадкових процесів

1.2.1 Базові поняття випадкових процесів

Однолінійна система масового обслуговування (далі СМО) функціонує так. Запит на обслуговування, який надійшов у систему, коли обслуговуючий пристрій вільний, приймається на обслуговування. Решта запитів, що надійшли в систему, коли обслуговуючий пристрій зайнятий, розміщуються в чергу. Після закінчення обслуговування чергового запиту, обслуговуючий пристрій вибирає запит з черги й починає його обслуговувати. Якщо запитів немає, обслуговуючий пристрій простоює, очікуючи надходження чергового запиту. Тобто в СМО як під час надходження запитів, так і після закінчення безперервно змінюється кількість запитів в черги й у системі (кількість запитів у черзі і на обслуговуванні).

Як відомо, під процесом розуміють послідовно змінювання станів об'єкта в часі [11]. Таким чином, у системах масового обслуговування відбуваються різні процеси. На рисунку 1.12 наведено часові діаграми базових процесів в однолінійній СМО.

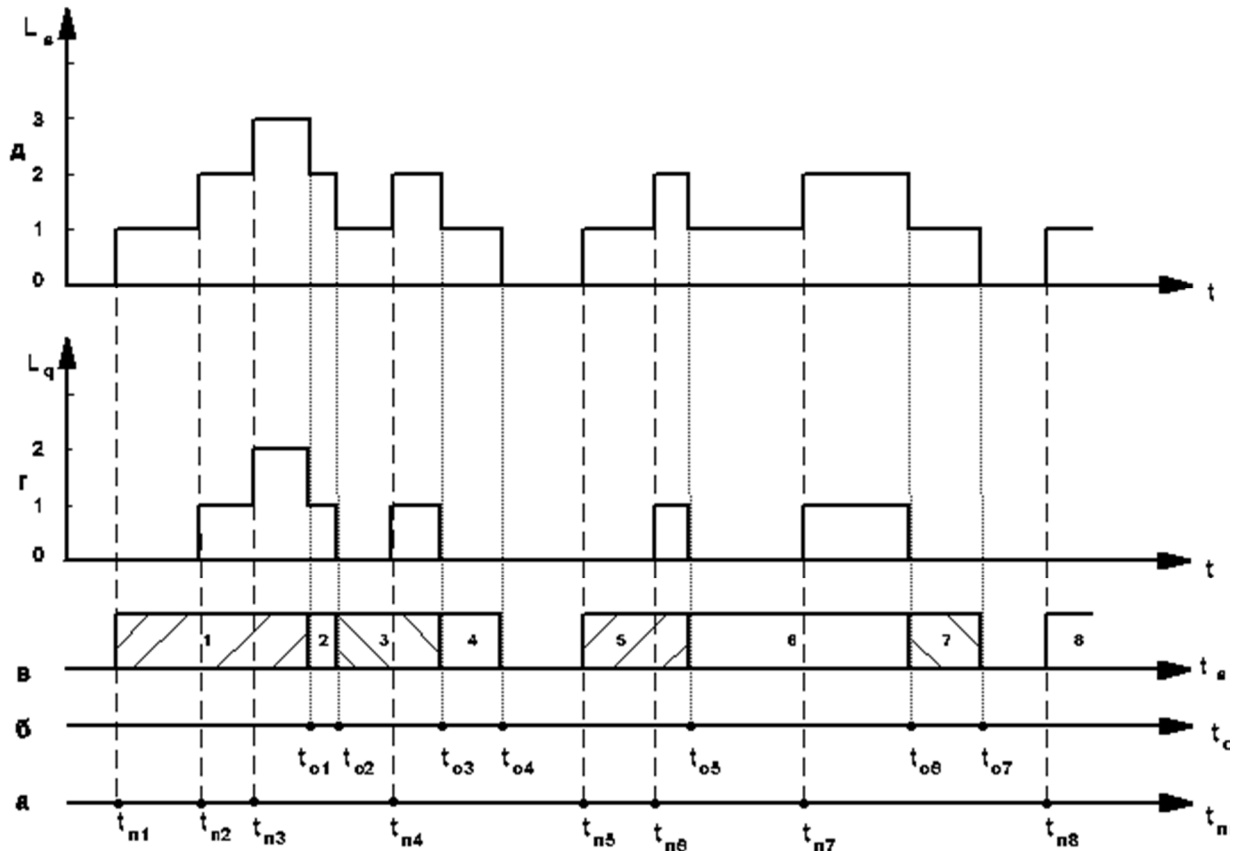


Рисунок 1.12 – Процеси в однолінійній СМО

Отже, на рисунку 1.12, а зображено процес надходження запитів у систему. Моменти надходження позначені точками t_{ni} . За кожним запитом, що надійшов, збільшується кількість запитів у системі, а якщо обслуговуючий пристрій зайнято – то й довжина черги. На рисунку 1.12, б на часовій осі зображено моменти закінчення обслуговування запитів (позначені точками t_{oi}). Відповідно зменшується довжина черги й кількість запитів у системі. На рисунку 1.12, в зображено процес послідовного обслуговування запитів на обслуговуючому приладі. Числами позначено номери запитів, які обслуговуються. На рисунку 1.12, г наведено часову діаграму змінювання кількості запитів у черзі, а на рисунку 1.12, д тимчасову діаграму змінювання кількості запитів в системі. З рисунку 1.12 випливає, що стани СМО змінюються у випадкові моменти часу, відповідно всі процеси є ймовірнісними і, щоб їх дослідити, необхідно використовувати теорію випадкових процесів.

Визначення. У математиці випадковим процесом називається функція двох аргументів, значення якої – випадкові величини [12, с. 297]:

$$\zeta(t) = \varphi(n, t), n = 0, 1, \dots, N, t \in T, \quad (1.30)$$

де $N + 1$ – кількість елементарних подій (воно може бути і нескінченним); t – параметр; T – безліч значень параметра T .

Для кожного значення параметра t функція $\varphi(n, t)$ є функцією тільки від n і, отже, становить випадкову величину. Для кожного фіксованого значення аргументу n (тобто для кожної елементарної події) $\varphi(n, t)$ залежить тільки від t і є, таким чином, просто функцією одного дійсного аргументу. Кожна така функція є реалізацією випадкового процесу. Рисунок 1.12, д відображає реалізацію випадкового процесу змінювання кількості вимог в системі, де в кожен момент часу кількість вимог в системі є випадковою величиною. Як параметр t стосовно СМО, використовується час.

Залежно від виду безлічі T значень аргументу t випадкові процеси можна розподілити на чотири види:

- процеси з дискретними станами та дискретним часом;
- процеси з дискретними станами й безперервним часом;
- процеси з безперервними станами й дискретним часом;
- процеси з безперервними станами й безперервним часом.

Приклад 8. Нехай маємо кілька реалізацій випадкових процесів $\zeta(t) = \varphi(n, t) : x_1(t), x_2(t), \dots, x_n(t)$, поданих на рисунку 1.13.

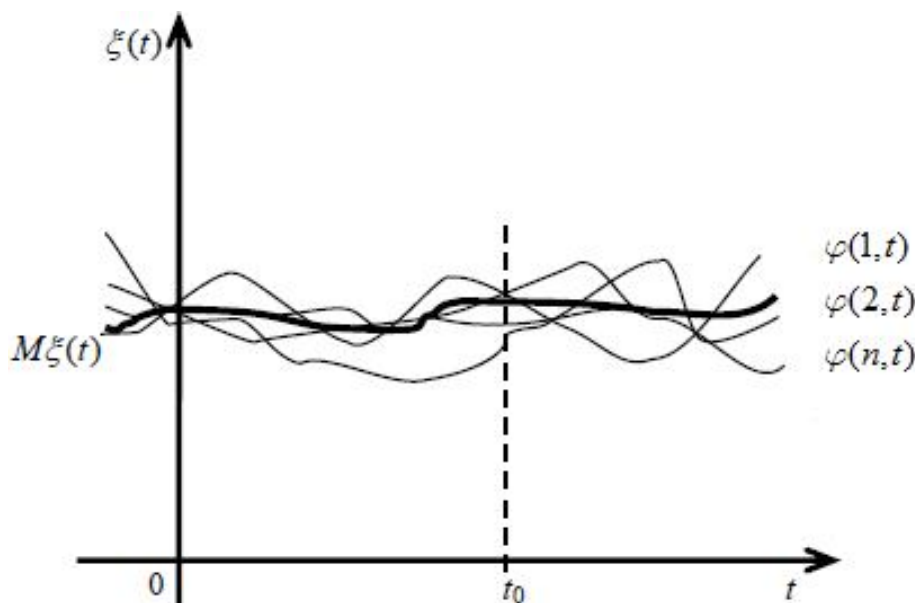


Рисунок 1.13 – Реалізація випадкового процесу $\zeta(t) = \varphi(n, t)$

Жирна лінія позначає функцію, навколо якої можливі різні реалізації. Ця функція називається математичним сподіванням випадкового процесу. Для отримання функції необхідний математичний опис випадкового процесу, до того ж для кожного виду випадкового процесу він може бути іншими.

З аналізу ймовірнісних процесів систем масового обслуговування (рис. 1.12) випливає, що вони становлять процеси з дискретними станами й безперервним часом. Станом системи можна вважати кількість вимог в системі. Залежно від виду СМО воно може бути кінцевим (для систем з обмеженою чергою) або нескінченним. Із рисунка 1.13 зрозуміло, що при конкретному значенні параметра t стан системи змінюється від реалізації до реалізації, а тому можна оцінити тільки його ймовірність або відповідні моменти (математичне сподівання, дисперсія).

Надійшовши в систему, запит перебуває деякий час у черги, на обслуговуванні та в системі загалом. Ці характеристики можна досліджувати, за допомогою теорії випадкових процесів із дискретними станами й безперервним часом.

Найдослідженішими є такі типи випадкових процесів: процеси з незалежними приростами, процеси з некорельованими значеннями, нормальні випадкові процеси, марківські процеси. У теорії масового обслуговування внаслідок специфічності характеру потоку запитів і часу обслуговування більшість задач вирішується із залученням теорії марківських процесів. Її перевагами є досконалий математичний апарат, можливість досліджувати як нестационарні, так і стаціонарні процеси, наочна інтерпретація результатів досліджень

Випадковий процес називається марківським (або процесом без післядії), якщо для кожного моменту часу t ймовірність будь-якого стану системи в майбутньому обумовлюється тільки її станом на сьогодні і не залежить від того, яким шляхом система набула цього стану. Параметр часу t може розглядатися і як дискретний, і як безперервний, унаслідок чого використовуються різні методи дослідження.

1.2.2 Марківські випадкові процеси з дискретним часом

Марківські випадкові процеси, або ланцюги Маркова, названі так на честь нашого співвітчизника, який вперше їх дослідив. Нехай досліджувана система може перебувати в станах $\varphi(n, t), n = 0, 1, 2, \dots, N; t = 0, 1, 2, 3, \dots$. Можливі переходи зі стану $\varphi(i, t)$ в стан $\varphi(j, t)$, $0 \leq i, j \leq N$, вони відбуваються в дискретні моменти часу $t = 0, 1, 2, 3, \dots$. Нехай $p_n(t)$ позначає

ймовірність стану $\varphi(n,t)$ у момент часу t . Тоді вектор $[P(t)] = [p_0(t), p_1(t), \dots, p_N(t)]$ описує ймовірності перебування системи в тому чи іншому стані в дискретний момент часу t . Позначимо через $p_{ij}(t), 0 \leq i, j \leq N$, умовну ймовірність того, що система, перебуваючи в момент t в стані $\varphi(i,t)$, перейде в стан $\varphi(j,t+1)$ в момент часу $t+1$ (так звані перехідні ймовірності), тоді

$$p_j(t+1) = \sum_{i=0}^N p_i(t) \cdot p_{ij}(t), j = 0, 1, 2, \dots, N. \quad (1.31)$$

Рівняння (1.31) і початкові умови повністю описує ланцюг Маркова. Якщо перехідні ймовірності не залежать від часу ($p_{ij}(t) = p_{ij}$), то такий ланцюг називається однорідним (в іншому разі неоднорідним). Здебільшого застосовуються однорідні ланцюги Маркова. Для однорідного ланцюга Маркова рівняння (1.31) можна записати так:

$$p_j(t+1) = \sum_{i=0}^N p_i(t) \cdot p_{ij}, j = 0, 1, 2, \dots, N. \quad (1.32)$$

Матриця

$$[\Pi] = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \dots & p_{0N} \\ p_{10} & p_{11} & p_{12} & \dots & p_{1N} \\ p_{20} & p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ p_{N0} & p_{N1} & p_{N2} & \dots & p_{NN} \end{bmatrix} \quad (1.33)$$

називається однокроковою матрицею перехідних ймовірностей або стохастичною матрицею. Її ймовірності за рядками мають таку властивість:

$$\sum_{j=0}^N p_{ij} = 1. \quad (1.34)$$

Рівняння (1.32) в матричній формі можна записати так:

$$[P(t+1)] = [P(t)] \cdot [\Pi]. \quad (1.35)$$

Вираз (1.35) дає змогу знайти ймовірності стану системи за один крок. Якщо потрібно знайти ймовірності системи за два кроки, то це роблять так:

$$[P(t+2)] = ([P(t)] \cdot [\Pi]) \cdot [\Pi] = [P(t)] \cdot [\Pi]^2. \quad (1.36)$$

За індукцією отримаємо

$$[P(t+k)] = ([P(t)] \cdot [\Pi]) \cdot [\Pi] \cdot [\Pi] \cdot \dots \cdot [\Pi] = [P(t)] \cdot [\Pi]^k, \quad (1.37)$$

звідки впливає найважливіша характеристика ланцюга Маркова – матриця переходу зі стану $\varphi(i,t)$ в стан $\varphi(j,t+k)$ за k кроків

$$[P_{ij}^k] = [\Pi]^k. \quad (1.38)$$

Приклад 9. Нехай є три альтернативні моделі смартфонів: a , b , c . Серед споживачів проведено опитування щодо їхнього ставлення до цих моделей. При цьому визначалася частка (частота) споживачів, які надають перевагу тим чи іншим моделям: $p_a(0) = 0,6$; $p_b(0) = 0,3$; $p_c(0) = 0,1$. За місяць споживачів опитують щодо того, чи продовжують вони надавати перевагу обраним моделям, чи бажають змінити тип моделі відповідно до її переваг. При цьому було отримано частоти, які можна вважати умовними ймовірностями переходу:

$$p_{aa} = 0,2; p_{ab} = 0,5; p_{ac} = 0,3; p_{ba} = 0,4; p_{bb} = 0,1; p_{bc} = 0,5;$$

$$p_{ca} = 0,5; p_{cb} = 0,2; p_{cc} = 0,3.$$

За місяць проводилися нове опитування і т. д. Припустивши, що переваги споживачів не змінюються в часі, отримали однорідний ланцюг Маркова з матрицею перехідних ймовірностей :

$$[\Pi] = \begin{bmatrix} 0,2 & 0,5 & 0,3 \\ 0,4 & 0,1 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{bmatrix}.$$

Досліджуваний процес для наочності зручно надати у вигляді графа (рис. 1.14).

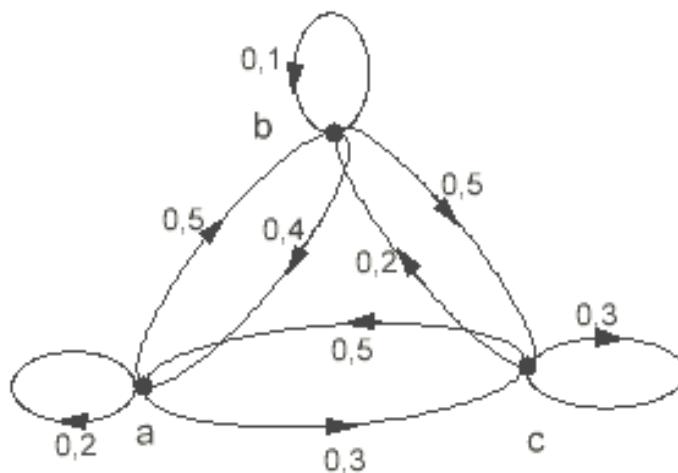


Рисунок 1.14 – Граф станів системи

За формулою (1.35) визначимо переваги споживачів за місяць:

$$[P(t+1)] = [P(t)] \cdot [\Pi] = [0,6; 0,3; 0,1] \cdot \begin{bmatrix} 0,2 & 0,5 & 0,3 \\ 0,4 & 0,1 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{bmatrix} = [0,29; 0,35; 0,36].$$

Отже за модель *a* висловилися 29 % усіх споживачів, за модель *b* – 35 %, за модель *c* – 36%. За два місяці отримаємо:

$$[P(t+2)] = [P(t+1)] \cdot [\Pi] = [0,29; 0,35; 0,36] \cdot \begin{bmatrix} 0,2 & 0,5 & 0,3 \\ 0,4 & 0,1 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{bmatrix} = [0,378; 0,252; 0,37].$$

За три місяці отримаємо:

$$[P(t+3)] = [P(t)] \cdot [\Pi]^3 = [0,6; 0,3; 0,1] \cdot \begin{bmatrix} 0,2 & 0,5 & 0,3 \\ 0,4 & 0,1 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{bmatrix}^3 = [0,361; 0,288; 0,351].$$

За шість місяців –

$$[P(t+6)] = [P(t)] \cdot [\Pi]^6 = [0,6; 0,3; 0,1] \cdot \begin{bmatrix} 0,2 & 0,5 & 0,3 \\ 0,4 & 0,1 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{bmatrix}^6 = [0,362; 0,281; 0,357].$$

Можна зробити висновок про те, що ймовірності прагнуть до своїх границь при $k \rightarrow \infty$. Ця властивість називається ергодичною, згідно з нею система, яка описується ланцюгом Маркова, з часом переходить у сталий, стаціонарний стан, для існування якого необхідні певні умови.

Уведемо деякі поняття. Якщо стохастична матриця $[\Pi]$ має вигляд $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$, де *A* і *B* квадратні підматриці, то вона називається розкладною, в іншому разі – нерозкладною. Система, що перебуває в стані, *A*, ніколи не перейде в стан, *B* і навпаки. Приклад розкладної матриці:

$$\begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & p_{43} & p_{44} \end{bmatrix}.$$

Якщо стохастична матриця $[\Pi]$ має вигляд $\begin{bmatrix} 0 & C \\ D & 0 \end{bmatrix}$, де *C* і *D* – квадратні

підматриці, то вона називається періодичною, в іншому разі неперіодичною. Система періодично буде переходити зі стану, C в стан D , і навпаки.

Якщо ланцюг Маркова є нерозкладним або неперіодичним, то для системи, яку він описує, існує стаціонарний стан [13, с. 151] і

$$\lim_{k \rightarrow \infty} [P(t+k)] = \lim_{k \rightarrow \infty} [p_0(t+k), p_1(t+k), \dots, p_N(t+k)] = [\pi_0, \pi_1, \pi_2, \dots, \pi_N] \quad (1.39)$$

Позначимо стаціонарні ймовірності $\pi_1, \pi_2, \dots, \pi_N$ або (в матричному вигляді) – $[p(\infty)] = [\pi_1, \pi_2, \dots, \pi_N]$. Для визначення стаціонарних ймовірностей використаємо вираз (1.35), який можна записати так:

$$[p(\infty)] = [p(\infty)] \cdot [\Pi], \quad (1.40)$$

його потрібно розв'язати відносно $[p(\infty)]$. Запишемо рівняння (1.40) так:

$$[p(\infty)] = [p(\infty)] \cdot [\Pi] \Rightarrow 0 = [p(\infty)] \cdot [\Pi] - [p(\infty)] = 0 \Rightarrow [p(\infty)]([\Pi] - I) = 0,$$

де I одинична матриця. Розкривши матриці, отримаємо:

$$[\pi_0, \pi_1, \dots, \pi_N] \cdot \begin{bmatrix} p_{00} - 1 & p_{01} & \dots & p_{0N} \\ p_{10} & p_{11} - 1 & \dots & p_{1N} \\ \dots & \dots & \dots & \dots \\ p_{N0} & p_{N1} & \dots & p_{NN} - 1 \end{bmatrix} = 0. \quad (1.41)$$

Остаточно отримаємо таку систему рівнянь:

$$\begin{aligned} (p_{00} - 1)\pi_0 + p_{10}\pi_1 + p_{20}\pi_2 \dots + p_{N0}\pi_N &= 0, \\ p_{01}\pi_0 + (p_{11} - 1)\pi_1 + p_{21}\pi_2 \dots + p_{N1}\pi_N &= 0, \\ p_{02}\pi_0 + p_{12}\pi_1 + (p_{22} - 1)\pi_2 \dots + p_{N2}\pi_N &= 0, \\ \dots & \\ p_{0N}\pi_0 + p_{1N}\pi_1 + p_{2N}\pi_2 \dots + (p_{NN} - 1)\pi_N &= 0. \end{aligned} \quad (1.42)$$

У системі рівнянь (1.42) одне з рівнянь необхідно замінити на так звану умову нормування:

$$\pi_0 + \pi_1 + \pi_2 + \dots + \pi_N = 1. \quad (1.43)$$

Визначимо стаціонарні ймовірності для ланцюга Маркова за прикладом 9. Система рівнянь (1.42) для нього матиме такий вигляд:

$$\begin{aligned} (0,2 - 1)\pi_a + 0,4\pi_b + 0,5\pi_c &= 0, \\ 0,5\pi_a + (0,1 - 1)\pi_b + 0,2\pi_c &= 0, \\ 0,3\pi_a + 0,5\pi_b + (0,3 - 1)\pi_c &= 0. \end{aligned}$$

Умова нормування: $-\pi_a + \pi_b + \pi_c = 1$. Замінивши, наприклад, останнє рівняння умовою нормування, отримаємо систему рівнянь відносно π_a, π_b, π_c

$$\begin{aligned} -0,8\pi_a + 0,4\pi_b + 0,5\pi_c &= 0, \\ 0,5\pi_a - 0,9\pi_b + 0,2\pi_c &= 0, \\ 1 \cdot \pi_a + 1 \cdot \pi_b + 1\pi_c &= 1. \end{aligned}$$

Розв'яжемо її за правилом Крамера через визначники:

$$\pi_a = \frac{\begin{vmatrix} 0 & 0,4 & 0,5 \\ 0 & -0,9 & 0,2 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} -0,8 & 0,4 & 0,5 \\ 0,5 & -0,9 & 0,2 \\ 1 & 1 & 1 \end{vmatrix}} = \frac{0,53}{3,12} = 0,363; \pi_b = \frac{\begin{vmatrix} -0,8 & 0 & 0,3 \\ 0,5 & 0 & 0,2 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} -0,8 & 0,4 & 0,5 \\ 0,5 & -0,9 & 0,2 \\ 1 & 1 & 1 \end{vmatrix}} = 0,281; \pi_c = \frac{\begin{vmatrix} -0,8 & 0,4 & 0 \\ 0,5 & -0,9 & 0 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} -0,8 & 0,4 & 0,5 \\ 0,5 & -0,9 & 0,2 \\ 1 & 1 & 1 \end{vmatrix}} = 0,356.$$

Ймовірності, обчислені на шостому кроці прикладу, практично співпадають зі стаціонарними ймовірностями під час дослідження процесу змінювання уподобань споживачів у часі.

1.2.3 Марківські випадкові процеси з дискретними станами й безперервним часом

Як зрозуміло з рисунка 1.12, перехід системи з одного стану в інший відбуваються не у фіксовані, а у випадкові моменти часу. Для опису таких процесів використовують марківські процеси з дискретними станами й безперервним часом.

Такі процеси описують функціонування систем, які набувають в процесі роботи кінцевого числа станів $\varphi(n,t), n=0,1,2,\dots,N$ і переходять з одного стану в інший $-\varphi(k,t), k=0,1,2,\dots,N$ випадково в довільний момент часу t . Отже, час перебування системи в будь-якому стані становить безперервну випадкову величину.

Випадковий процес з безперервним часом називається марківським процесом з дискретними станами й безперервним часом, якщо поведінка системи після довільного моменту часу t обумовлюється тільки станом процесу в момент часу t і не залежить від передісторії процесу, що передує моменту часу t .

Найважливішою характеристикою системи є ймовірність того, що в момент t система буде перебувати в стані $\varphi(n,t), n=0,1,2,\dots,N$ – $p_i(t)$. Очевидно, що для будь-якого моменту часу t сума всіх ймовірностей

дорівнює одиниці. Позначимо через $p_{nk}(t, \Delta t)$ імовірність переходу системи зі стану $\varphi(n, t), n=0, 1, 2, \dots, N$ у стан $\varphi(k, t), k=0, 1, 2, \dots, N$ за нескінченно малий проміжок часу Δt . Позначимо через $\lambda_{nk}(t)$ таку границю:

$$\lambda_{nk}(t) = \lim_{\Delta t \rightarrow 0} \frac{p_{nk}(t, \Delta t)}{\Delta t}. \quad (1.44)$$

Ця характеристика називається інтенсивністю переходу або щільністю ймовірності переходу і загалом залежить від t . Із формули (1.44) випливає, що при невеликому Δt імовірність переходу з точністю до нескінченно малих вищих порядків

$$p_{nk}(t, \Delta t) = \lambda_{nk}(t)\Delta t, \Delta t \rightarrow 0. \quad (1.45)$$

Якщо усі $\lambda_{nk}(t) = \lambda_{nk}$ не залежать від часу, то марківський процес називається однорідним, в іншому разі неоднорідним. Зауважимо, що в теорії масового обслуговування використовуються здебільшого однорідні марківські процеси.

Аналогічно до марківських процесів із дискретним часом, для подання процесів із безперервним часом широко використовуються графи. Кожен стан зображується у вигляді вершини графа; між вершинами позначаються дуги, що відображають можливість переходу системи з n -го стану в k -й за нескінченно малий інтервал Δt ; над дугами проставляють величини інтенсивностей переходів λ_{nk} . Такий граф прийнято називати графом станів і переходів, на підставі цього графа можна вивести диференціальні рівняння Колмогорова [14, с. 98], які повністю описують поведінку системи.

Приклад 10. Нехай система S складається з двох вузлів – A і B , кожен із яких у процесі роботи може відмовити. Можливими є такі стани системи:

- Ψ_0 – працюють обидва вузла;
- Ψ_1 – вузол A відмовив і відновлюється, вузол B – працює;
- Ψ_2 – вузол A працює, вузол B відмовив і відновлюється;
- Ψ_3 – відновлюються обидва вузла, до того ж час закінчення відновлення вузла A не збігається з часом закінчення відновлення вузла B .

Відомо інтенсивності переходів між станами. Граф станів і переходів буде мати такий вигляд поданий на рисунку. 1.15.

Виведемо рівняння для ймовірності стану $\Psi_0 - p_0(t)$. Для цього розглянемо ймовірності різних переходів зі стану Ψ_0 і навпаки в інтервалі часу $(t, t + \Delta t)$.

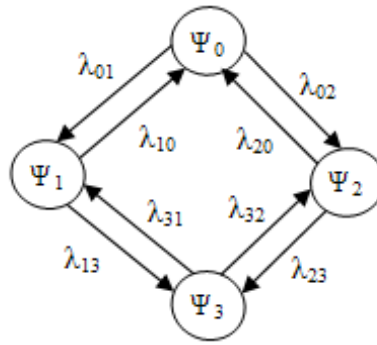


Рисунок 1.15 – Граф станів і переходів для системи з відмовами

У момент часу t система перебувала в стані Ψ_0 , імовірність того, що вона перейшла з нього в який інший стан дорівнює $p_0 \cdot (\lambda_{01} + \lambda_{02})\Delta t$, а того, що вона залишилася в ньому, $- 1 - p_0(t) \cdot (\lambda_{01} + \lambda_{02}) \cdot \Delta t$. Імовірність того, що система в момент часу t перебувала в стані Ψ_1 і перейшла за час Δt в стан Ψ_0 , дорівнює $p_1(t) \cdot \lambda_{10} \cdot \Delta t$. Імовірність того, що система в момент часу t перебувала в стані Ψ_2 і перейшла в стан Ψ_0 , дорівнює $p_2(t) \cdot \lambda_{20} \cdot \Delta t$. Тоді, використовуючи формули додавання й множення ймовірностей, одержимо:

$$p_0(t + \Delta t) = (1 - (\lambda_{01} + \lambda_{02})\Delta t)p_0(t) + \lambda_{10} \cdot p_1(t) \cdot \Delta t + \lambda_{20} \cdot p_2(t) \cdot \Delta t.$$

Розкриваємо дужки, перенесемо $p_0(t)$ з правої частини в ліву, розділимо усе рівняння на Δt і знайдемо границю лівої і правої частин при $\Delta t \rightarrow 0$. Отримаємо:

$$\lim_{\Delta t \rightarrow 0} \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -(\lambda_{01} + \lambda_{02})p_0(t) + \lambda_{10} \cdot p_1(t) + \lambda_{20} \cdot p_2(t).$$

Ліва частина становить собою похідну від $p_0(t)$ по t . Остаточно отримаємо:

$$p_0'(t) = -(\lambda_{01} + \lambda_{02})p_0(t) + \lambda_{10} \cdot p_1(t) + \lambda_{20} \cdot p_2(t).$$

Узагальнюючи процес складання диференціального рівняння для ймовірності стану $\Psi_0 - p_0(t)$, можна за розміченим графам станів і переходів скласти мнемонічне правило складання диференціального рівняння для довільного стану $\Psi_i - p_i(t)$ [14, 135]:

похідна ймовірності будь-якого стану Ψ_i дорівнює добутку суми інтенсивностей переходу зі стану Ψ_i в інші стани та ймовірності стану Ψ_i , узятого з від'ємним знаком, плюс сумі добутків інтенсивностей переходу з інших станів у стан Ψ_i на ймовірності відповідних станів.

Для інших станів диференціальні рівняння для ймовірностей станів будуть мати такий вигляд:

$$\begin{aligned}
p_1'(t) &= -(\lambda_{10} + \lambda_{13})p_1(t) + \lambda_{01} \cdot p_0(t) + \lambda_{31} \cdot p_3(t), \\
p_2'(t) &= -(\lambda_{20} + \lambda_{23})p_2(t) + \lambda_{02} \cdot p_0(t) + \lambda_{32} \cdot p_3(t), \\
p_3'(t) &= -(\lambda_{31} + \lambda_{32})p_3(t) + \lambda_{13} \cdot p_1(t) + \lambda_{23} \cdot p_2(t).
\end{aligned}$$

У початковий момент часу система була працездатною, тому

$$p_0(0) = 1, p_1(0) = 0, p_2(0) = 0, p_3(0) = 0.$$

Як результат отримаємо систему диференціальних рівнянь:

$$\begin{aligned}
p_1'(t) &= -(\lambda_{10} + \lambda_{13})p_1(t) + \lambda_{01} \cdot p_0(t) + \lambda_{31} \cdot p_3(t), \\
p_2'(t) &= -(\lambda_{20} + \lambda_{23})p_2(t) + \lambda_{02} \cdot p_0(t) + \lambda_{32} \cdot p_3(t), \\
p_3'(t) &= -(\lambda_{31} + \lambda_{32})p_3(t) + \lambda_{13} \cdot p_1(t) + \lambda_{23} \cdot p_2(t), \\
p_3'(t) &= -(\lambda_{31} + \lambda_{32})p_3(t) + \lambda_{13} \cdot p_1(t) + \lambda_{23} \cdot p_2(t)
\end{aligned} \tag{1.46}$$

із початковими умовами

$$p_0(0) = 1, p_1(0) = 0, p_2(0) = 0, p_3(0) = 0.$$

Використовуючи технологію розв'язання систем лінійних диференціальних рівнянь [15, с. 54] при конкретних значеннях λ_{ij} , можна знайти аналітичний розв'язок цієї системи. Приміром, при

$$\lambda_{01} = 0,8; \lambda_{02} = 1,5; \lambda_{10} = 1,6; \lambda_{13} = 2,0; \lambda_{20} = 1,3; \lambda_{23} = 2,5; \lambda_{31} = 3,8; \lambda_{32} = 4,6$$

розв'язок системи (1.46) буде виглядати так:

$$\begin{aligned}
p_0(t) &= 0,8544 \sin(0,2541t + 0,8943)e^{-3,4892t} + 0,01e^{-11,12t} + 0,32788, \\
p_1(t) &= -0,9290 \sin(0,2541t + 0,2094)e^{-3,48t} - 0,03e^{-11,12t} + 0,22471, \\
p_2(t) &= -0,5642 \sin(0,2541t + 2,6592)e^{-3,48t} - 0,041e^{-11,12t} + 0,30371, \\
p_3(t) &= -0,1244 \sin(0,2541t + 1,0199)e^{-3,48t} + 0,062e^{-11,12t} + 0,14385.
\end{aligned}$$

Графіки залежностей функцій $p_0(t), p_1(t), p_2(t), p_3(t)$ від часу наведено на рисунку 1.16.

Зрозуміло, що в разі необмеженого зростання t функції $p_i(t)$ перестають залежати від часу й система переходить в стаціонарний режим. Тобто, $\lim_{t \rightarrow \infty} p_i(t) = p_i, i = 0, 1, 2, 3$, де p_i – константи. З аналізу виразів для функцій $p_i(t)$ випливає, що при $t \rightarrow \infty$ вони починають дорівнювати відповідним константам в їхній правій частині:

$$p_0 = 0,32788, p_1 = 0,22471, p_2 = 0,30371, p_3 = 0,14385.$$

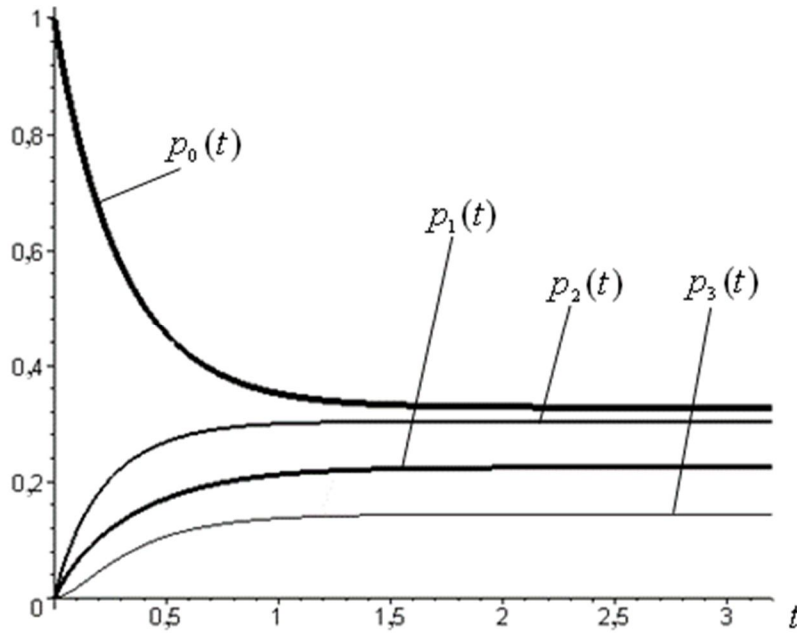


Рисунок 1.16 – Залежності ймовірностей станів системи від часу

Ймовірності станів системи в стаціонарному режимі можна обчислити більш простим методом. З огляду на те що $\lim_{t \rightarrow \infty} p'_i(t) = \lim_{t \rightarrow \infty} p_i = 0$, система рівнянь (1.46) змінюється так:

$$\begin{aligned}
 &-(\lambda_{01} + \lambda_{02})p_0 + \lambda_{10} \cdot p_1 + \lambda_{20} \cdot p_2 = 0, \\
 &-(\lambda_{10} + \lambda_{13})p_1 + \lambda_{01} \cdot p_0 + \lambda_{31} \cdot p_3 = 0, \\
 &-(\lambda_{20} + \lambda_{23})p_2 + \lambda_{02} \cdot p_0 + \lambda_{32} \cdot p_3 = 0, \\
 &-(\lambda_{31} + \lambda_{32})p_3 + \lambda_{13} \cdot p_1 + \lambda_{23} \cdot p_2 = 0,
 \end{aligned} \tag{1.47}$$

тобто вона перейде в систему лінійних алгебраїчних рівнянь відносно $p_i, i=0,1,2,3$, розв'язати яку значно простіше ніж систему диференціальних рівнянь. Щоб знайти ймовірності стаціонарних станів $p_i, i=0,1,2,3$ необхідно розв'язати систему лінійних алгебраїчних рівнянь (1.47), замінивши одне з рівнянь умовою нормування $p_0 + p_1 + p_2 + p_3 = 1$. Як результат отримаємо таку систему лінійних алгебраїчних рівнянь щодо ймовірностей:

$$\begin{aligned}
 &-(\lambda_{01} + \lambda_{02})p_0 + \lambda_{10} \cdot p_1 + \lambda_{20} \cdot p_2 = 0, \\
 &-(\lambda_{10} + \lambda_{13})p_1 + \lambda_{01} \cdot p_0 + \lambda_{31} \cdot p_3 = 0, \\
 &-(\lambda_{20} + \lambda_{23})p_2 + \lambda_{02} \cdot p_0 + \lambda_{32} \cdot p_3 = 0, \\
 &p_0 + p_1 + p_2 + p_3 = 1.
 \end{aligned} \tag{1.48}$$

При $\lambda_{01} = 0,8; \lambda_{02} = 1,5; \lambda_{10} = 1,6; \lambda_{13} = 2,0; \lambda_{20} = 1,3; \lambda_{23} = 2,5; \lambda_{31} = 3,8; \lambda_{32} = 4,6$ розв'язок системи лінійних алгебраїчних рівнянь буде виглядати так:

$$p_0 = 0,327\ 89, p_1 = 0,224\ 70, p_2 = 0,303\ 56, p_3 = 0,143\ 85$$

Цей результат з точністю до 10^{-4} збігається з результатом, отриманим за розв'язком системи диференціальних рівнянь при $t \rightarrow \infty$.

З аналізу графіків рисунку 1.16 випливає, що більшість часу експлуатації система, описувана марківським процесом, перебуває в стаціонарному стані. Отже подібні системи зазвичай досліджують при стаціонарному стані, який описується системою лінійних алгебраїчних рівнянь. Якщо складено розмічений граф станів і переходів, то для нього можна скласти систему лінійних алгебраїчних рівнянь щодо стаціонарних ймовірностей. У цій системі для ймовірностей кожного стану $\Psi_i = p_i$ складається лінійне алгебраїчне рівняння за таким правилом:

у правій частині розміщується нуль, а в лівій – добуток суми інтенсивностей переходу зі стану Ψ_i в інші стани на ймовірність p_i , узятого зі знаком мінус, плюс сума добутків інтенсивностей переходу з інших станів в стан Ψ_i та ймовірності відповідних станів.

Математично це можна записати так:

$$-(\lambda_{ie} + \lambda_{id} + \dots + \lambda_{is}) \cdot p_i + \sum_k \lambda_{ki} \cdot p_k = 0. \quad (1.49)$$

Одне з рівнянь необхідно замінити на умову нормування

$$\sum_k p_k = 1. \quad (1.50)$$

Сформовану систему лінійних алгебраїчних рівнянь доцільно розв'язувати за допомогою програмного додатка, з відповідними можливостями, наприклад системи комп'ютерної математики Matlab.

1.3 Потоки подій і їх характеристики

Процес надходження запитів для обслуговування в СМО є ймовірнісним і утворює потік однорідних або неоднорідних подій, які настають через випадкові проміжки часу (рис. 1.17, де t_i – моменти надходження запитів, що утворюють потік випадкових подій; ξ_i – інтервали часу між сусідніми запитами).

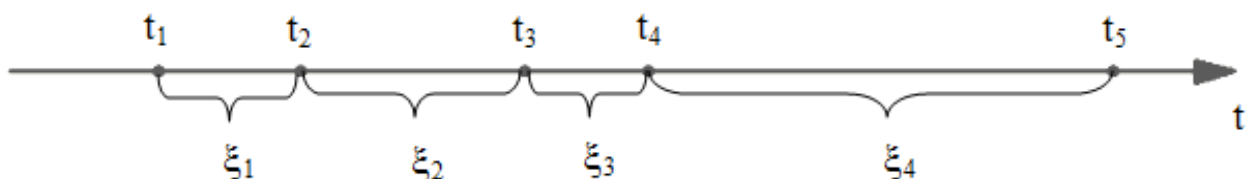


Рисунок 1.17 – Процес надходження запитів для обслуговування

Базовим параметром потоку є інтенсивність потоку $\lambda(t)$ – математичне сподівання кількості подій за одиницю часу [16]:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{m_k(t, t + \Delta t)}{\Delta t}, \quad (1.51)$$

тобто, це границя відношення математичного сподівання кількості подій – $m_k(t, t + \Delta t)$ на інтервалі $(t, t + \Delta t)$ до довжини цього інтервалу, що наближається до нуля.

Якщо величини ξ_i є незалежними і однаково розподіленими, то такий потік називається потоком з обмеженою післядією (потік Пальма). Таким чином, базовим описом потоку є функція розподілу $A(t) = P\{\xi < t\}$. Такі потоки є головними в теорії масового обслуговування.

Кожен запит, що надійшов на обслуговування, перебуває на приладі обслуговування випадковий час η , що описується функцією розподілу $B(t) = P\{\eta < t\}$. Якщо припустити, що на вході до приладу обслуговування весь час перебувають запити, то на виході обслуговуючого приладу сформується потік обслужених запитів з функцією розподілу $B(t)$. Отже, обслуговування відбувається так, ніби на запит, що надійшов на обслуговування, спрямовується потік обслуговування. Така постановка завдання дуже зручна методично, оскільки уможливує узагальнений розгляд процесів надходження й обслуговування. Таким чином, у системі масового обслуговування функціонують два видів потоків – надходження й обслуговування, що характеризуються різними функціями розподілу.

Випадкові часові інтервали між настанням подій у потоці можуть характеризуватися різними законами розподілу. Однак у більшості робіт щодо теорії масового обслуговування, особливо прикладних, розглядається пуассонівський (найпростіший) потік, у якому ймовірність того, що в проміжку часу t відбудеться рівно k вимог, задається формулою Пуассона [4, с. 16]:

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad (1.52)$$

де $\lambda > 0$ – параметр потоку.

Найпростіший потік характеризується трьома головними властивостями: стаціонарність, відсутність післядії й ординарність.

Випадковий потік називається стаціонарним, якщо ймовірність надходження певної кількості запитів протягом певного відрізка часу залежить від його величини й не залежить від початку його відліку на осі

часу. Це означає: якщо на часовій осі відкласти рівні, які не перетинаються, інтервали часу τ (рис. 1.18), то ймовірність появи в цих інтервалах певної кількості запитів для цього потоку залежить від величини τ й не залежить від розміщення цього інтервалу на часовій осі (від моментів часу t_1, t_2, t_3, \dots).

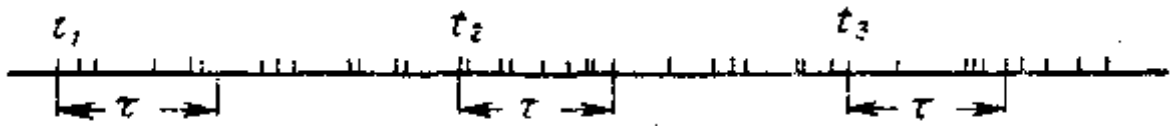


Рисунок 1.18 – Стаціонарний потік однорідних подій

Для стаціонарного потоку $\lambda(t) = \lambda = 1/m$, (1.53)

де m – математичне сподівання часових проміжків між сусідніми подіями потоку, яке обчислюється за формулою (1.14).

Таким чином, два найпростіші потоки різняться тільки своїми параметрами, тому для завдання найпростішого потоку досить задати лише його параметр λ .

Відсутність післядії полягає в тому, що ймовірність надходження за відрізок часу τ певної кількості вимог не залежить від того, скільки вимог вже надійшло в систему, тобто не залежить від передісторії досліджуваного явища. Відсутність післядії передбачає взаємну незалежність перебігу процесу за проміжки часу, які не перекриваються.

Ординарність потоку вимог означає практичну неможливість появи двох і більше вимог в один і той самий момент часу.

Таким чином, найпростіший потік – це стаціонарний, ординарний потік без післядії.

Обчислимо інтенсивність найпростішого потоку. У проміжку Δt можуть відбуватися як одна так і нескінченна кількість подій. За формулою

$$(1.6), \quad m_k(t, t + \Delta t) = \sum_{k=1}^{\infty} k p_k(\Delta t) = \sum_{k=1}^{\infty} k \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t} = e^{-\lambda \Delta t} \sum_{k=1}^{\infty} k \frac{(\lambda \Delta t)^k}{k!} = \lambda \Delta t.$$

Звідси $\lambda = m_k(t, t + \Delta t) / \Delta t = \lambda \Delta t / \Delta t = \lambda$, тобто інтенсивність найпростішого потоку співпадає з його параметром.

Обчислимо функцію розподілу випадкової величини ξ – інтервали часу між сусідніми подіями пуассонівського потоку $A(t) = P\{\xi < t\}$. Функція $A(t) = P\{\xi < t\}$ є ймовірністю того, що в проміжку часу t , що прилягає до моменту настання чергової події, настане хоча б одна подія потоку.

Ймовірність цього – $\sum_{k=1}^{\infty} p_k(t)$. Оскільки $\sum_{k=0}^{\infty} p_k(t) = 1$, то

$A(t) = 1 - p_0(t) = 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t}$. Остаточно – $A(t) = 1 - e^{-\lambda t}$, тобто функція розподілу випадкової величини – інтервали часу між сусідніми подіями пуассонівського потоку задається експоненціальним розподілом. Щільність розподілу – $a(t) = A'(t) = \lambda e^{-\lambda t}$, математичне сподівання – $m = 1/\lambda$, дисперсія – $D = 1/\lambda^2$, середньоквадратичне відхилення – $\sigma = 1/\lambda$, коефіцієнт варіації – $v = \sigma/m = 1$.

Експоненціальний розподіл має одну властивість: якщо проміжок часу, розподілений експоненціально, триває час τ , то це не впливає на функцію розподілу решти проміжку T . Він буде таким самим, як і функція розподілу всього проміжку, тобто: якщо проміжок часу різниться експоненціальним розподілом, то будь-які відомості про час перебігу на деякому проміжку не впливають на функцію розподілу в частині, що залишилася. Доведено, що тільки експоненціальний розподіл має таку властивість. Це значно спрощує математичні обчислення під час аналізу систем масового обслуговування.

Найпростіший потік в теорії масового обслуговування відіграє таку саму роль, як і нормальний закон розподілу випадкових величин у теорії ймовірностей. У разі підсумовування (взаємного накладання) великої кількості ординарних стаціонарних потоків фактично з будь-якими розподілами формується потік, наближений до найпростішого [16, с. 11].

Найпростіший потік відіграє важливу роль в теорії моделювання. По-перше, найпростіші й близькі до найпростіших потоки подій дуже часто зустрічаються на практиці. По-друге, навіть при потоці подій, що відрізняється від найпростішого, зазвичай можна знайти задовільні за точністю результати, замінивши потік будь-якої структури найпростішим потоком із тією самою середньою інтенсивністю.

У практиці моделювання для опису реальних потоків, крім експоненціального, застосовують такі розподіли:

1. Ерлангівський з функцією розподілу:

$$A(t) = 1 - \sum_{i=1}^k \frac{(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t}. \quad (1.54)$$

Ерлангівський розподіл формується шляхом підсумовування k інтервалів, кожен з яких розподілений за експоненціальним законом із параметром λ . Отже, потік подій, описуваний ерлангівським розподілом k -го порядку, формується з пуассонівського шляхом видалення всіх подій, що

містяться між l -ю і $(l + k)$ -ю подіями. При $k = 1$ ерлангівський розподіл перетворюється на експоненціальний.

Знайдемо щільність ерлангівського розподілу $a(t)$:

$$\begin{aligned} a(t) = A'(t) &= -\sum_{i=1}^k \left(\frac{(i-1)\lambda^{i-1}t^{i-2}}{(i-1)!} e^{-\lambda t} - \frac{\lambda^i t^{i-1}}{(i-1)!} e^{-\lambda t} \right) = \\ &= -e^{-\lambda t} \sum_{i=1}^k \left(\frac{(i-1)\lambda^{i-1}t^{i-2}}{(i-1)!} - \frac{\lambda^i t^{i-1}}{(i-1)!} \right) = -e^{-\lambda t} \left(0 + \lambda + \lambda^2 t + \frac{1}{2} \lambda^3 t^2 + \frac{1}{6} \lambda^4 t^3 + \dots \right) + \\ &+ e^{-\lambda t} \left(\lambda + \lambda^2 t + \frac{1}{2} \lambda^3 t^2 + \frac{1}{6} \lambda^4 t^3 + \dots + \frac{1}{(k-1)!} \lambda^k t^{k-1} \right). \end{aligned}$$

Остаточно отримаємо

$$a(t) = \frac{1}{(k-1)!} \lambda (\lambda t)^{k-1} e^{-\lambda t}. \quad (1.55)$$

Математичне сподівання – $m = k / \lambda$, середньоквадратичне відхилення $\sigma = \sqrt{k} / \lambda$, коефіцієнт варіації – $v = \sqrt{k} / k \leq 1$.

За допомогою цього розподілу досить точно описуються потоки подій, що мають коефіцієнт варіації v , менший за одиницю.

При $k \rightarrow \infty$ ерлангівський потік перетворюється на регулярний, коли події настають через рівні проміжки часу.

На рисунку 1.19 зображено графіки щільності ерлангівського розподілу при $\lambda = 0,1$ і різних k .

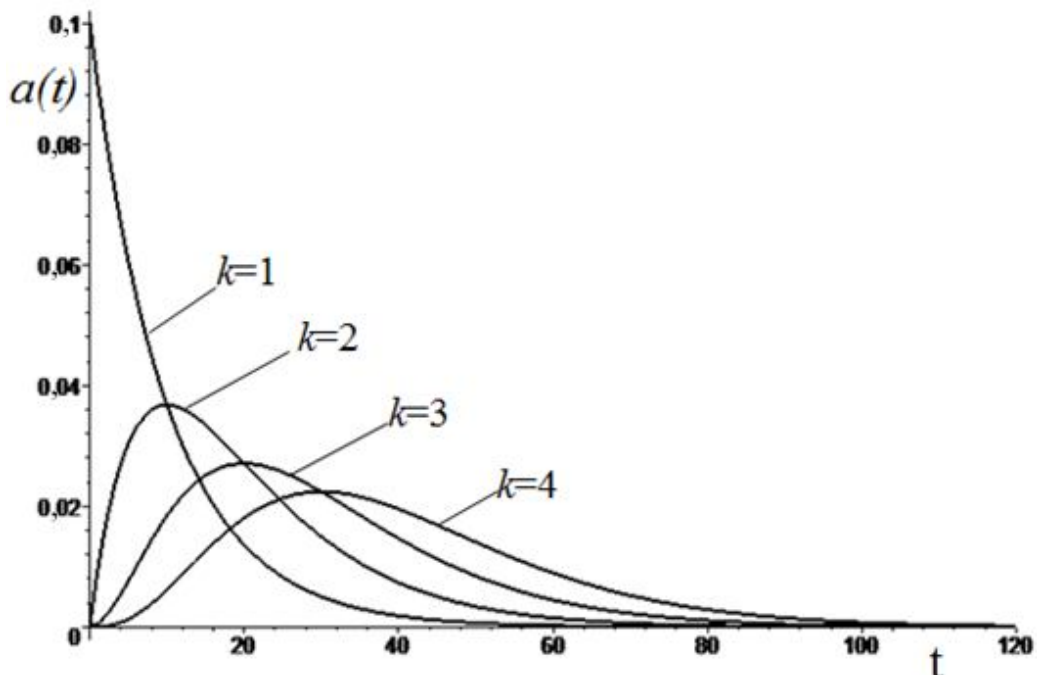


Рисунок 1.19 – Графіки щільності ерлангівського розподілу

На відміну від експоненціального розподілу ($k = 1$), графік щільності ерлангівського розподілу має максимум (моду). З ростом k мода зміщується вправо.

2. Гіперекспоненціальний – з функцією розподілу

$$A(t) = 1 - \sum_{i=1}^k a_i e^{-\lambda_i t}, \quad \sum_{i=1}^k a_i = 1 \quad (1.56)$$

і щільністю розподілу

$$a(t) = \sum_{i=1}^k a_i \lambda_i e^{-\lambda_i t}, \quad \sum_{i=1}^k a_i = 1. \quad (1.57)$$

Математичне сподівання – $m = \sum_{i=1}^k \frac{a_i}{\lambda_i}$,

$$\text{коефіцієнт варіації} - v = \left[2 \sum_{i=1}^k \frac{a_i}{\lambda_i^2} - \left(\sum_{i=1}^k \frac{a_i}{\lambda_i} \right)^2 \right]^{0,5} / \sum_{i=1}^k \frac{a_i}{\lambda_i}, \quad v \geq 1.$$

Рекомендовано використовувати гіперекспоненціальний розподіл другого порядку, що має тільки два параметри λ і φ із функцією розподілу:

$$A(t) = 1 - \varphi e^{-2\varphi\lambda t} - (1 - \varphi) e^{-2(1-\varphi)\lambda t}, \quad 0 < \varphi \leq 0,5, \quad (1.58)$$

Тобто $a_1 = \varphi$, $a_2 = 1 - \varphi$, $\lambda_1 = 2\varphi\lambda$, $\lambda_2 = 2(1 - \varphi)\lambda$. При цьому математичне

сподівання – $m = 1/\lambda$, коефіцієнт варіації – $v = \left[1 + \frac{(1 - 2\varphi)^2}{2\varphi(1 - \varphi)} \right]^{0,5}$.

Якщо задано математичне сподівання потоку m і його коефіцієнт варіації v , то параметри λ і φ можна визначити за формулами

$$\lambda = 1/m, \quad \varphi = 0,5 - 0,5 \sqrt{\frac{v^2 - 1}{v^2 + 1}}. \quad (1.59)$$

Потік подій, який описується гіперекспоненціальним розподілом формується, якщо з імовірністю a_i , $\sum a_i = 1$, черговий інтервал часу між настанням сусідніх подій обирається розподіленням за експоненціальним законом з параметром λ_i . При $k = 1$ гіперекспоненціальний розподіл перетворюється на експоненціальний. За допомогою гіперекспоненціального розподілу досить точно описуються потоки подій, що мають коефіцієнт варіації, більший за одиницю.

Контрольні питання

1. Що називається подією; випадковою подією?
2. Які події називаються складними, а які елементарними?
3. Що називається об'єднанням подій; перетином подій?
Які події називаються спільними; несумісними?
4. Які події називаються протилежними?
5. Що таке ймовірність подій?
6. Що називається випадковою величиною?
7. Які випадкові величини називаються дискретними, а які – неперервними?
8. Що називається: законом розподілу випадкової величини; рядом розподілу; функцією розподілу?
9. Якими властивостями характеризується функція розподілу?
10. Що називається щільністю розподілу?
11. Якими властивостями характеризується щільність розподілу?
12. Що називається: математичним сподіванням випадкової величини; дисперсією випадкової величини; середньоквадратичним відхиленням випадкової величини?
13. Який розподіл називається нормальним?
14. Чому дорівнює математичне сподівання, дисперсія, коефіцієнт асиметрії та варіації випадкової величини, розподіленої нормально?
15. Який розподіл називається експоненціальним (показниковим)?
16. Чому дорівнює математичне сподівання, дисперсія, коефіцієнт асиметрії та варіації випадкової величини, розподіленої за експоненціальним законом?
17. Дайте визначення випадковому процесу.
18. Назвіть головні види випадкових процесів.
19. Подайте визначення марківського процесу.
20. Що описує граф станів марківського процесу? Як за ним скласти систему рівнянь для однорідного ланцюга Маркова?
21. Як скласти систему диференціальних рівнянь для ймовірностей станів марківського процесу з дискретними станами й безперервним часом?
22. Як скласти систему лінійних рівнянь для ймовірностей станів марківського процесу з дискретними станами й безперервним часом?
23. Назвіть головний параметр потоку подій.
24. Який потік подій називається найпростішим?
25. Якими властивостями характеризується найпростіший потік?

2 БАЗОВІ ПОНЯТТЯ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ ТА МЕТОДИ ЇХ ДОСЛІДЖЕННЯ

2.1 Базові поняття систем масового обслуговування

Формально під системою масового обслуговування (далі СМО) розуміють складну систему, що складається з одного або декількох джерел запитів (заявок, вимог) на виконання певних дій (обслуговування), декількох приладів обслуговування (ліній обслуговування, каналів обслуговування), що виконують ці дії відповідно до певних правил (дисципліни обслуговування) за запитами, що надійшли в систему (рис. 2.1).

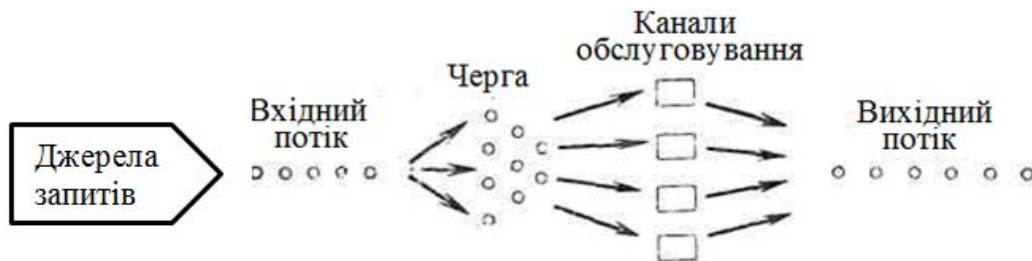


Рисунок 2.1 – Структура системи масового обслуговування

Характерною особливістю СМО є імовірнісність процесів, що відбувається, і можливість утворення черги запитів на обслуговування. Розглянемо базові поняття СМО.

Джерело запитів. Джерело запитів визначається як зовнішня щодо СМО система, із якої запити надходять в неї для обслуговування. Джерело називають нескінченним або кінцевим залежно від того, нескінченна чи кінцева кількість запитів міститься в ньому. Якщо джерело містить кінцеву, але досить велику кількість запитів, то його зазвичай вважають нескінченним. Наприклад, хоча кількість користувачів інформаційно-пошукової системи Google кінцева, припускають, що вони утворюють нескінченне джерело.

Вхідний потік. Запити, що надходять з нескінченного джерела, прибувають в канал обслуговування в моменти часу $t_0 < t_1 < \dots < t_k < \dots$. Інтервали часу $\xi_k = t_k - t_{k-1}$ ($k > 1$) між послідовними моментами надходження запитів є випадковими величинами. Передбачається, що ξ_k утворюють послідовність незалежних і однаково розподілених випадкових величин з функцією розподілу $A(t) = P\{\xi_k < t\}$. У математичній теорії масового обслуговування використовують обмежений набір законів розподілу для опису вхідних потоків. Головними є пуассонівський (найпростіший), який задається формулою Пуассона (1.52).

Процес обслуговування. Нехай η_k – тривалість обслуговування k -го запиту. Передбачається, що η_k ($k = 1, 2, \dots$) – незалежні однаково розподілені випадкові величини з функцією розподілу $B(t) = P\{\eta_k < t\}$. Функція $B(t)$ називається розподілом тривалості обслуговування. Припускаємо, що існує щільність імовірності $b(t) = B'(t)$. Як і для опису вхідних потоків запитів у теорії масового обслуговування використовують обмежений набір законів розподілу для опису часу обслуговування. Найпоширенішим є експоненціальний закон розподілу з функцією розподілу $B(t) = 1 - \exp(-\mu t)$. Параметр μ інтерпретується як інтенсивність обслуговування. Середній час обслуговування дорівнює $1/\mu$. Крім експоненціального розподілу, використовують ерлангівський і гіперекспоненціальний розподіли, які отримують шляхом перетворення експоненціального закону.

Канали обслуговування. СМО може мати один або більше каналів (ліній, приладів) обслуговування. СМО з одним каналом називаються одноканальними або однолінійними, тоді як системи обслуговування, що містять більше каналів обслуговування, називаються багатоканальними або багатолінійними. Прилади обслуговування можуть бути однорідними й неоднорідними. У СМО з однорідними приладами всі прилади обслуговують запити однаково. У СМО з неоднорідними приладами прилади відрізняються один від одного деякими параметрами, наприклад інтенсивністю обслуговування.

Дисципліна обслуговування. Правило, згідно з яким запити вибираються для обслуговування, складають дисципліну обслуговування. Умовно всі дисципліни обслуговування по наданих переваг в обслуговуванні діляться на дві групи: безпріоритетні і пріоритетні. Кожна з цих груп ділиться на ряд підгруп.

Безпріоритетні дисципліни обслуговування поділяються на дисципліну обслуговування в порядку надходження, у зворотному порядку, із випадковим вибором з черги та циклічну дисципліну обслуговування.

Дисципліна обслуговування в порядку надходження (у міжнародній нотації FIFO – First Input First Output) є найпоширенішою, більшість досліджень щодо теорії масового обслуговування виконуються для цієї дисципліни. Вона широко використовується в операційних системах.

Дисципліна обслуговування в зворотному (інверсному) порядку (у міжнародній нотації LIFO – Last Input First Output) використовується в операційних системах під час обробки переривань і організації стеків.

Циклічні дисципліни обслуговування використовуються під час обробці інформації в режимі поділу часу.

При пріоритетних дисциплінах обслуговування з черги на обслуговування спочатку вибираються заявки з вищим пріоритетом. Вони поділяються на дисципліни з фіксованими й динамічними пріоритетами [17, с. 63].

При дисципліні обслуговування з відносним пріоритетом не дозволяється переривання обслуговування запиту на каналі.

Якщо в систему з дисципліною обслуговування з абсолютним пріоритетом надійде запит з пріоритетом вищим, ніж той що обслуговується, то він припинить обслуговування цього запиту й надійде на обслуговування. Системи з абсолютним пріоритетом розрізняють за кількістю рівнів пріоритету, а також за алгоритмами для обслуговування перерваних запитів.

При дисциплінах обслуговування з динамічним пріоритетом пріоритет конкретних запитів змінюється залежно від змінювання деяких величин, наприклад часу очікування в черги.

За наявністю певної ознаки системи масового обслуговування можна класифікувати так:

1. *За кількістю вимог, що надходить за одиницю часу, на системи з ординарним і неординарним потоками вимог.* Якщо імовірність надходження двох і більше вимог одночасно дорівнює нулю або має настільки мале значення, що ним можна знехтувати, то отримаємо систему з ординарним потоком вимог. Наприклад, потік вимог – літаки, що надходить на злітно-посадкову смугу аеродрому, можна вважати ординарним.

2. *За зв'язком між вимогами – на системи без післядії і з післядією.* Якщо ймовірність надходження вимог у систему в деякий момент часу не залежить від того, скільки вимог уже надійшло, тобто не пов'язана з передісторією досліджуваного процесу, то отримаємо систему без післядії, в іншому разі – із післядією. Прикладом системи з післядією може слугувати потік студентів, що здають залік викладачу.

3. *За реакцією вимоги на зайнятість каналів – на системи з відмовами й очікуваннями.* Якщо вимога, яка надійшла на обслуговування, застала всі канали зайнятими і змушена залишити систему, то отримаємо систему з відмовами.

4. *Системи з очікуванням* розподіляються на системи з обмеженим і необмеженим очікуванням. Якщо вимога залишає систему, коли черга набула певного розміру, то отримаємо систему з обмеженим очікуванням.

Прикладом може слугувати самоскид з розчином. Якщо час очікування настільки великий, що розчин може затвердіти, то самоскид доречно розвантажити в іншому місці. Якщо вимога, яка надійшла, застала всі канали зайнятими і змушена очікувати своєї черги доти, доки вона не буде обслужена, то отримаємо систему з очікуванням без обмежень. Приклад: літак, що перебуває на аеродромі й очікує звільнення злітної смуги.

5. *За способом вибору вимог на обслуговування* розподіляються так:

- із пріоритетом вимог;
- у процесі надходження вимог;
- із випадковим вибором вимог;
- остання вимога обслуговується першою.

Якщо система масового обслуговування охоплює декілька категорій вимог і за якимись ознаками визначається порядок їх вибору на обслуговування, то отримаємо систему з пріоритетом вимог. Приміром, під час надходження виробів на будівельний майданчик насамперед монтують ті, що обумовлені будівельною технологією.

Якщо канал, що звільнився, обслуговує вимогу, яка надійшла в систему раніше за інші, то отримаємо систему обслуговування вимог у процесі їх надходження. Наприклад, покупець, що підійшов першим до продавця, обслуговується першим.

Якщо вимоги з черги в канал обслуговування надходить випадково, то отримаємо систему з випадковим вибором вимог на обслуговування. Приклад: вибір слюсарем-сантехніком однієї з декількох заявок від мешканців, із часом надходження яких він не ознайомлений.

Якщо для обслуговування обирається остання вимога, що надійшла, то отримаємо систему з вибором «останній обслуговується першим». Приміром, під час укладання будівельних виробів штабелями зручніше обирати зі штабеля (черги) виріб, покладений останнім.

6. *За часом обслуговування* вимоги – на системи з *детермінованим і випадковим часом* обслуговування. Якщо інтервал часу між моментом надходження вимоги в канал обслуговування і моментом виходу вимоги з каналу постійний, то отримаємо систему з детермінованим часом обслуговування, в іншому разі – із випадковим. Наприклад, миття автомобілів становить систему обслуговування з детермінованим часом обслуговування.

7. *За кількістю каналів обслуговування* – на одно- й багатоканальні системи. Приміром під час монтажу будинку може використовуватися один

підіймальний кран (один канал обслуговування) або декілька (багато каналів обслуговування).

8. За кількістю етапів обслуговування – на одно- й багатофазні системи. Якщо канали обслуговування розташовуються послідовно й неоднорідні, то отримаємо багатофазну систему обслуговування. Прикладом такої системи може слугувати обслуговування автомобілів на СТО (миття, діагностика, заміна фільтрів тощо).

9. За однорідністю вимог – на системи з однорідними й неоднорідними потоками вимог. Приміром, якщо під навантаження прибувають фургони однієї вантажопідйомності, то отримаємо систему з однорідним потоком вимог, якщо різної – із неоднорідним.

10. За завантаженістю каналів – на впорядковані й невпорядковані системи. У впорядкованих системах обслуговуючі канали завантажені нерівномірно. Вимога, що надійшла, обслуговується чіткого визначеним каналом із наявних вільних, а саме каналом з найменшим номером (вважається, що всі канали пронумеровані). У невпорядкованих системах усі канали однакові й вимога, що надійшла, обслуговується одним із вільних каналів без будь-яких переваг.

Для скороченого позначення розімкнутих безпріоритетних невпорядкованих систем масового обслуговування Д. Кендал запропонував таке позначення : $A/B/n/m$. A і B описують, відповідно, закони розподілу інтервалів часу між подіями вхідного потоку запитів і часом обслуговування, n задає кількість каналів обслуговування, m – кількість місць очікування.

Певні закони розподілу позначають так: M – експоненціальний розподіл, E або E_k – ерланговський розподіл; H або H_r – гіперекспоненціальний розподіл; D – регулярний потік подій; G – розподіл загального вигляду. Приміром, запис $D/H_2/3/10$ позначає систему з трьома каналами обслуговування, постійним (детермінованим) часом між запитами вхідного потоку, час обслуговування розподілено за гіперекспоненціальним законом другого порядку й кількістю місць для очікування – 10. Якщо параметр m відсутній, то це означає, кількість місць для очікування нескінченна.

2.2 Показники ефективності систем масового обслуговування

Показники ефективності систем масового обслуговування поділяються на *технічні*, що характеризують якість і умови роботи обслуговуючої системи, і *економічні*, що відображають економічні особливості системи.

Показники першої групи зазвичай формують на підставі отриманих із розрахунків значень ймовірностей станів системи. Показники другої групи – на підставі показників першої групи.

2.2.1 Технічні показники ефективності систем масового обслуговування

Серед технічних показників можна відокремити такі:

1) Імовірність відмови обслуговування. Імовірність того, що вимога, яка надходить у систему, відмовиться приєднатися до черги і *втратить*ся системою – P_{vidm} . Цей показник для системи масового обслуговування з відмовами дорівнює імовірності того, що в системі міститься стільки вимог, скільки є каналів обслуговування:

$$P_{\text{vidm}} = P_n, \quad (2.1)$$

де n – кількість каналів обслуговування.

Для системи з обмеженою довжиною черги P_{vidm} дорівнює імовірності того, що в системі міститься $n + m$ вимог:

$$P_{\text{vidm}} = P_{n+m}, \quad (2.2)$$

де m – допустима довжина черги.

Протилежним показником є імовірність обслуговування вимоги:

$$P_{\text{obsl}} = 1 - P_{\text{vidm}}; \quad (2.3)$$

2) середня кількість вимог, що очікують на обслуговування:

$$L_q = \sum_{i=n+1}^{n+m} (i - n) p_i, \quad (2.4)$$

де p_i – імовірність того, що в системі міститься i вимог;

3) відносна й абсолютна пропускні здатності системи, які визначають за такими формулами:

– відносна пропускна здатність:

$$Q = 1 - P_{\text{vidm}}; \quad (2.5)$$

– абсолютна пропускна спроможність:

$$A = \lambda \cdot Q; \quad (2.6)$$

4) середня кількість зайнятих обслуговуванням каналів для систем G/G/n/m (тобто n – обслуговуючі пристрої, m – місця для чекання):

$$\bar{n} = \sum_{i=1}^{n-1} i p_i + n \sum_{i=n}^{n+m} p_i. \quad (2.7)$$

Для систем масового обслуговування з відмовами середню кількість зайнятих обслуговуванням каналів можна знайти за формулою:

$$\bar{n} = \sum_{i=1}^n i \cdot p_i; \quad (2.8)$$

5) загальна кількість вимог, що містяться в системі L_s . Цей показник визначають так:

– для систем масового обслуговування з відмовами:

$$L_s = \bar{n}; \quad (2.9)$$

– для систем масового обслуговування з обмеженою довжиною черги:

$$L_s = \bar{n} + L_q; \quad (2.10)$$

б) середній час очікування вимогами початку обслуговування. Якщо відома функція розподілу ймовірності часу очікування вимогою початку обслуговування $W(t) = P(T_q < t)$, то середній час очікування вимогами початку обслуговування T_q визначається, як математичне очікування випадкової величини T_q :

$$\tau_q = M[T_q] = \int_0^{\infty} t dW(t). \quad (2.11)$$

При показовому (експоненціальному) законі розподіли вимог у вхідному потоці τ_q можна визначити за формулою

$$\tau_q = \frac{L_q}{\lambda}. \quad (2.12)$$

2.2.2 Економічні показники ефективності систем масового обслуговування

Показники, що характеризують економічні особливості систем масового обслуговування, зазвичай формують відповідно до певного виду системи та її призначення. Одним із загальних показників є економічна ефективність системи:

$$G = \lambda P_{obsl} C_s T - E, \quad (2.13)$$

де C_s – середній економічний ефект, отриманий під час обслуговуванні однієї вимоги; T – розглянутий інтервал часу; E – величина втрат у системі.

Останню величину (втрати) можна визначити так:

– для систем з відмовами

$$E = (c_e \bar{n} + \lambda c_{zb} P_{vidm} + c_{prost} \bar{n}_{vi ln}) T, \quad (2.14)$$

де c_e – вартість експлуатації одного каналу за одиницю часу; c_{zb} – вартість збитків унаслідок виходу вимог із системи за одиницю часу; c_n – вартість одиниці часу простоювання каналу обслуговування; \bar{n}_{vln} – середня кількість вільних каналів (що простоюють), $\bar{n}_{vln} = n - \bar{n}$;

– для систем з очікуванням:

$$E = (c_e \bar{n} + c_q L_q \lambda + c_{zb} \bar{n}_{vln}) T, \quad (2.15)$$

де c_q – вартість втрат, пов'язаних із простоюванням вимоги в черзі за одиницю часу.

2.3 Методи дослідження систем масового обслуговування

Головним завданням дослідження СМО є встановлення залежностей базових характеристик, а саме: середній час очікування, імовірність відмови обслуговування від параметрів системи – інтенсивностей обслуговування на приладах, кількості приладів обслуговування, дисципліни обслуговування. Різноманітність СМО унеможливорює використання для аналізу одного методу. Хоча на практиці для певної групи СМО зазвичай використовується один метод, що дає змогу мінімізувати обчислення й скоротити час визначення базових залежностей.

2.3.1 Диференціальний метод

Диференціальний метод можна застосовувати для дослідження СМО з пуассонівськими потоками подій, тобто час обслуговування й інтервали часу між надходженням запитів підпорядковані експонентному розподілу. У цьому разі процес змінювання кількості запитів у системі є марківським і його можна дослідити за методикою, викладеною в підрозділі 1.3. Послідовність дослідження така:

– із аналізу системи обслуговування відокремлюють стани системи, можливі переходи між станами та їх інтенсивності і складають граф станів і переходів (рис 1.15);

– із кожним станом зв'язується відповідна ймовірність і складається система диференціальних рівнянь для ймовірностей станів системи за методикою, викладеною в підрозділі 1.2.3 (звідси й назва методу). Складену систему доцільно розв'язувати й досліджувати за однією з систем комп'ютерної математики, наприклад Maple.

Якщо дослідження системи у стаціонарному режимі передбачається заздалегідь, то знову (за методикою підрозділу 1.2.3) складається система

лінійних алгебраїчних рівнянь для стаціонарних ймовірностей станів. Зазвичай розв'язок вдається звести до розрахункових формул.

Використання диференціального методу для аналізу СМО розглянемо на системі обслуговування з груповим вибором запитів для обслуговування [18]. Ця СМО складається з накопичувача, у який надходить пуассонівський потік запитів інтенсивністю λ . Через деякі інтервали часу запити, що накопичилися вибираються для обслуговування. Ємкість накопичувача – k запитів. Інтенсивність вибору запитів з накопичувача – μ_i ($i = 0, 1, \dots, k$), тобто вона обумовлюється кількістю запитів у ньому. Цю СМО можна використовувати, наприклад, як модель роботи узгоджувального буфера для інформаційної системи.

Нехай s_i – стан, коли в накопичувачі міститься i запитів, імовірність цього $p_i, i = 1, 2, \dots, k$. Ця СМО відповідно до надходження нового запиту переходить зі стану s_i у стан s_{i+1} , інтенсивність цього переходу – λ . Під час вибору запитів із накопичувача відбувається перехід зі стану s_i ($i = 1, 2, \dots, k$) у стан s_0 . Інтенсивність цього переходу – μ_i . Граф станів і переходів для цієї СМО при $k = 4$ зображений на рисунку 2.2.

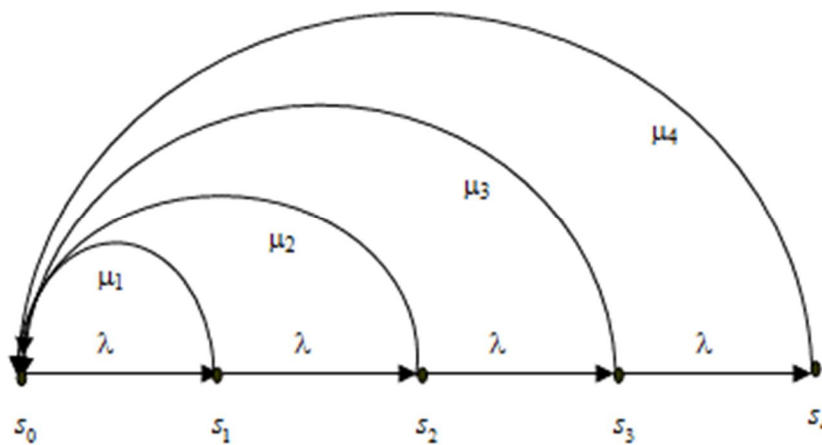


Рисунок 2.2 – Граф станів і переходів для системи з груповим вибором

Склавши для досліджуваної СМО систему диференціальних рівнянь, отримаємо:

$$\begin{aligned}
 p'_0(t) &= -\lambda p_0(t) + \sum_{i=1}^k \mu_i p_i(t), \\
 p'_i(t) &= -(\lambda + \mu_i) p_i(t) + \lambda p_{i-1}(t), \quad 1 \leq i \leq k, \\
 p'_k(t) &= -\mu_k p_k(t) + \lambda p_{k-1}(t),
 \end{aligned}
 \tag{2.16}$$

Початкові умови – $p_0(0) = 1, p_i(0) = 0, i \neq 0$.

Для стаціонарного стану система (2.16) перетворюється на таку:

$$\begin{aligned} -\lambda p_0 + \sum_{i=1}^k \mu_i p_i &= 0, \\ -(\lambda + \mu_i) p_i + \lambda p_{i-1} &= 0, \quad 1 \leq i \leq k, \\ -\mu_k p_k + \lambda p_{k-1} &= 0. \end{aligned} \quad (2.17)$$

Умова нормування –

$$p_0 + p_1 + p_2 + \dots + p_k = 1. \quad (2.18)$$

З рівнянь системи (2.17), починаючи з другого, послідовно отримаємо:

$$\begin{aligned} p_1 &= \frac{\lambda}{(\lambda + \mu_1)} p_0, & p_2 &= \frac{\lambda}{(\lambda + \mu_2)} p_1 = \frac{\lambda^2}{(\lambda + \mu_1)(\lambda + \mu_2)} p_0, \\ p_i &= \frac{\lambda^i}{\prod_{e=1}^i (\lambda + \mu_e)} p_0, & p_k &= \frac{\lambda^k}{\mu_k \prod_{e=1}^{k-1} (\lambda + \mu_e)} p_0. \end{aligned}$$

Підставивши значення p_i , виражені через p_0 , в умову нормування (2.18), отримаємо:

$$p_0 = \left[1 + \sum_{i=1}^{k-1} \frac{\lambda^i}{\prod_{e=1}^i (\lambda + \mu_e)} + \frac{\lambda^k}{\mu_k \prod_{e=1}^{k-1} (\lambda + \mu_e)} \right]^{-1}.$$

Базові характеристики досліджуваної системи:

- імовірність відмови щодо обслуговування – $P_{vidm} = p_k$;
- середня кількість запитів у накопичувачі – $L_q = \sum_{i=1}^k i p_i$;
- середній час перебування запитів у накопичувачі – $\tau_q = \sum_{i=1}^k \frac{p_i}{\mu_i}$.

За цими виразами можна обирати обсяг буфера, щоб імовірність його переповнення містилася в заданих межах.

2.3.2 Метод фаз

Метод фаз дає змогу досліджувати СМО більш загального вигляду, порівняно з пуассонівським за допомогою введення додаткових станів. Це дає змогу використати для такої розширеної системи методику складання рівнянь для додаткових станів, розглянуту в підрозділі 1.2.3. Метод фаз зазвичай використовується, коли потоки подій в СМО описуються ерлангівським або гіперекспоненціальним розподілами. Для методу фаз можна виокремити чотири базові способи переходу до розширеної системи рівнянь:

1. Час обслуговування розподілено за ерлангівським розподілом k -го порядку (1.54) з параметром μ . У цьому разі канал обслуговування умовно замінюється k окремими фіктивними каналами (фазами), розташованими послідовно. Кожен запит, прийнятий на обслуговування, послідовно проходить через ці k фаз, перебуваючи на кожній фазі період часу, що підпорядковується експоненціальному розподілу з параметром μ . Новий запит не обслуговується до тих пір, доки попередній не пройде всі k фаз обслуговування. СМО в цьому разі описується марківським процесом із станами s_{ij} , $1 \leq j \leq k$. Стан s_{ij} означає, що в системі міститься i запитів, а черговий запит обслуговується на j -й фазі. Цьому процесу відповідають імовірності станів $p_{ij}(t)$. На рисунку 2.3 зображено граф станів і переходів для однолінійної СМО з ерлангівським потоком обслуговування другого порядку $M/E_2/1$, для якого легко складається система рівнянь. Якщо ймовірності $p_{ij}(t)$ будуть знайдено, то ймовірності $p_i(t)$ то можна визначити за формулою $p_i(t) = \sum_{j=1}^k p_{ij}(t)$.

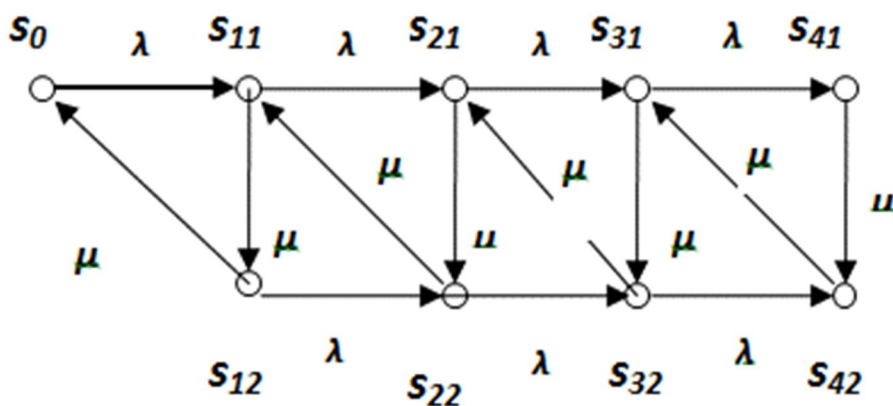


Рисунок 2.3 – Розширений граф станів і переходів системи $M/E_2/1$

2. Час обслуговування розподілено за гіперекспоненціальним розподілом k -го порядку виду (1.56) з параметрами b_i, μ_i . У цьому разі канал обслуговування умовно замінюється окремими фазами обслуговування з параметрами μ_i , розташованими паралельно. З імовірністю b_i кожен запит обслуговується на i -й фазі. Граф станів і переходів для однолінійної СМО з гіперекспоненціальним потоком обслуговування другого порядку $M/H_2/1$ зображено на рисунку 2.4.

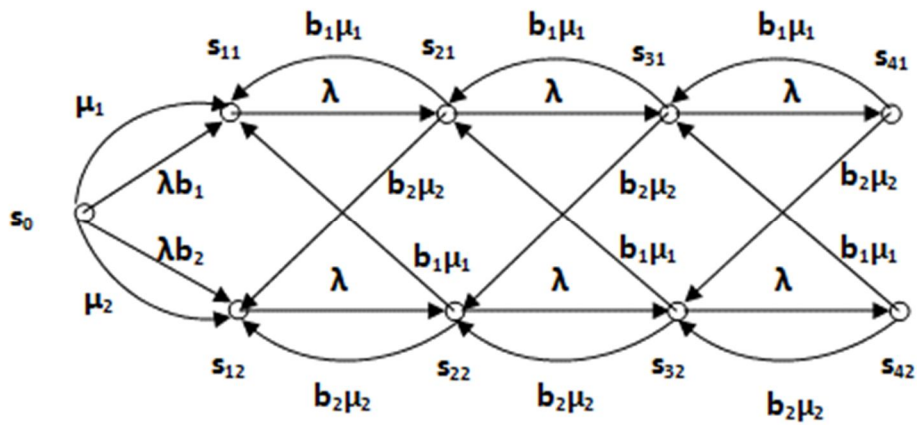


Рисунок 2.4 – Розширений граф станів і переходів системи M/H₂/1

3. Вхідний потік – ерлангівський k -го порядку з параметром λ . У цьому разі до СМО додається фіктивний приймальний пристрій, що складається з k фаз. Кожен запит спочатку надсилається в цей приймальний пристрій, де послідовно затримується на кожній фазі. Час затримки на кожній фазі характеризується експоненціальним розподілом із параметром λ . Тільки після проходження цього фіктивного приймального пристрою запит надходить у систему на обслуговування. СМО буде описуватися марківським випадковим процесом s_{ij} , де j – номер фази надходження вхідного запиту в систему. Граф станів і переходів для однолінійної СМО виду $E_2/M/1$ зображено на рисунку 2.5.

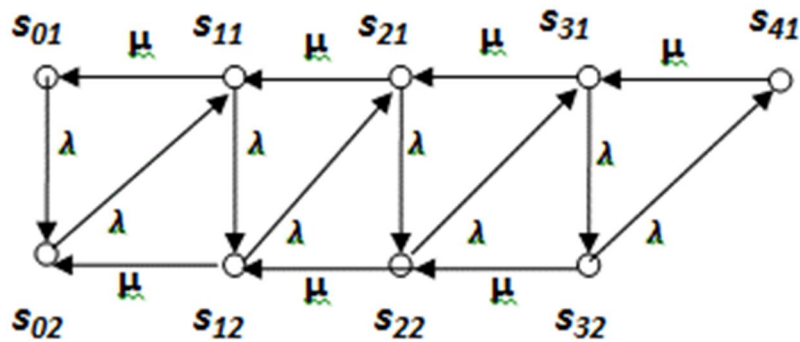


Рисунок 2.5 – Розширений граф станів і переходів системи E₂/M/1

4. Вхідний потік – гіперекспоненціальний k -го порядку з параметрами a_i, λ_i . Для переходу до марківського процесу до СМО додається фіктивний приймальний пристрій, що складається з паралельних фаз. Кожен запит на обслуговування спочатку надсилається в цей приймальний пристрій, де з ймовірністю a_i проходить i -ту фазу, а потім надходить на обслуговування. Граф станів і переходів для СМО виду $H_2/M/1$ зображено на рисунку 2.6.

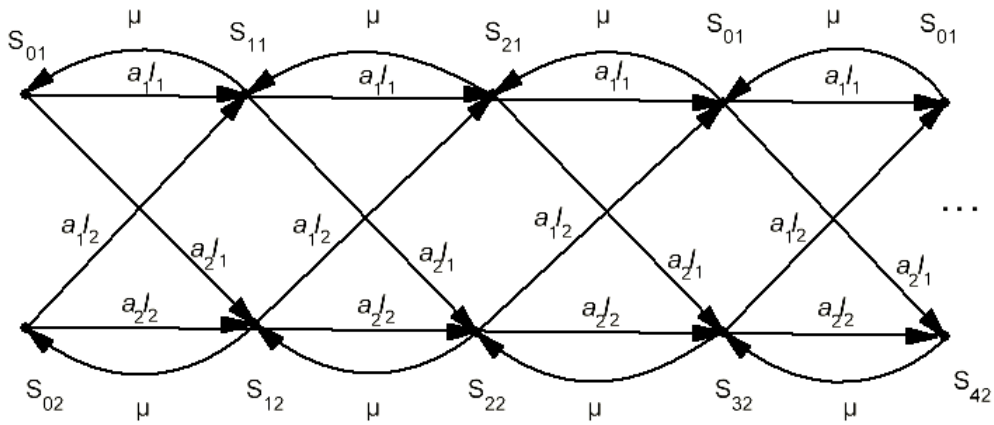


Рисунок 2.6 – Розширений граф станів і переходів системи $H_2/M/1$

Головним недоліком методу фаз є велика розмірність області можливих станів і відповідної системи рівнянь, розв'язати яку аналітично не вдається можливим. Єдиним способом отримання рішення є чисельний, із використанням стандартних підпрограм розв'язку систем рівнянь на комп'ютері.

Щоб проілюструвати використання методу фаз, розглянемо такий приклад. Нехай маємо СМО виду $M/GE/1$ де GE – символ узагальненого розподілу Ерланга, що містить перетворення Лапласа – Стілтєса від функції розподілу $\beta(s) = \prod_{j=1}^k \frac{\mu_j}{(\mu_j + s)}$. Відрізняється він від ерлангівського розподілу

тим, що кожен етап має свою інтенсивність – μ_j . Позначимо через $p_n(t)$ імовірність знаходження в системі n запитів. Уведемо фіктивні фази обслуговування. Тоді кожен стан можна розкласти на k фаз, процес переходів між якими – марківський. Нехай p_{nj} – стаціонарна ймовірність того, що в системі міститься n запитів, черговий запит обслуговується на j -й фазі ($j = \overline{1, k}$). Для ймовірностей справедлива система лінійних рівнянь, отримана за методикою, викладеною в підрозділі 1.2.3:

$$\begin{aligned}
 & -\lambda p_0 + \mu_1 p_{11} = 0, \\
 & -(\lambda + \mu_j) p_{1j} + \mu_{j+1} p_{1j+1} = 0, \\
 & -(\lambda + \mu_k) p_{1k} + \mu_1 p_{21} + \lambda p_0 = 0, \\
 & -(\lambda + \mu_j) p_{nj} + \mu_{j+1} p_{nj+1} + \lambda p_{n-1j} = 0, \quad n > 1, \quad 1 \leq j < k, \\
 & -(\lambda + \mu_k) p_{nk} + \mu_1 p_{n+11} + \lambda p_{n-1k} = 0, \quad n > 1.
 \end{aligned} \tag{2.19}$$

Умова нормування – $p_0 + \sum_{n=1}^{\infty} \sum_{j=1}^k p_{nj} = 1$.

Для розв'язання системи рівнянь (2.19) використаємо метод виробляючих функцій. Помножимо кожне рівняння системи (2.19) для p_{nj} ($1 \leq j \leq k$) на z^n , а для p_0 і p_{nk} – на z^{n+1} і підсумуємо рівняння з однаковим індексом j при однакових членах. Уведемо виробляючі функції $Q_j(z) = \sum_{n=1}^{\infty} p_{nj} z^n$. Отримаємо систему рівнянь щодо $Q_j(z)$:

$$\begin{aligned} -(\lambda + \mu_j)Q_j(z) + \lambda z Q_j(z) + \mu_{j+1}Q_{j+1}(z) &= 0, 1 \leq j < k, \\ -(\lambda + \mu_k)zQ_k(z) + \lambda z^2 Q_k(z) + \mu_1 Q_1(z) &= \lambda z(1-z)p_0. \end{aligned}$$

Розв'язавши її, знайдемо:

$$Q_j(z) = \frac{\lambda z(1-z)p_0 \prod_{i=1}^j (\lambda - \lambda z + \mu_i) \prod_{r=j}^k \mu_r}{\left[\prod_{i=1}^k \mu_i - z \prod_{i=1}^k (\lambda - \lambda z + \mu_i) \right] (\lambda - \lambda z + \mu_j) \mu_j}.$$

Завантаження системи – $\rho = \lambda \sum_{j=1}^k 1/\mu_j$ і $p_0 = 1 - \rho$. Виробляючу функцію розподілу стаціонарних ймовірностей p_n можна визначити так:

$$Q(z) = p_0 + \sum_{j=1}^k Q_j(z). \text{ Тоді } p_n = \frac{1}{n!} \left[\frac{d^n}{dz^n} Q(z) \right]_{z=0}.$$

$$\text{Наприклад, } p_1 = \frac{\lambda(1-\rho)}{\prod_{j=1}^k \mu_j} \sum_{j=1}^k \frac{\prod_{i=1}^k (\lambda + \mu_i) \prod_{r=j}^k \mu_r}{(\lambda + \mu_j) \mu_j}.$$

У цьому прикладі, окрім усього іншого, продемонстровано можливості методу виробляючих функцій, які широко використовуються для аналізу систем різницевих рівнянь.

2.3.3 Метод вкладених ланцюгів Маркова

Цей метод доцільно використовувати, коли лише один із потоків подій у СМО (вхідний потік або потік обслуговування) є марківським або зводиться до нього методом фаз, а інший є довільним. При методі вкладених ланцюгів Маркова випадковий процес у СМО розглядається не в довільні моменти часу, а в спеціально зазначених точках – точках регенерації (або точках відновлення), які утворюють ланцюг Маркова [19, с. 195]. Якщо вхідний потік пуассонівський, то точками регенерації потрібно обирати моменти часу,

коли чергова процедура обслуговування виявилася нещодавно завершеною, а відповідна вимога саме залишає систему. Якщо час обслуговування розподілено за експоненціальним законом, то точками регенерації обираються моменти часу, які співпадають із моментами надходження нових запитів.

Після виокремлення точок регенерації конкретизуються події, що відбуваються навколо цих точок і визначається розподіл ймовірностей кількості подій вхідного пуассонівського потоку або пуассонівського потоку обслуговування за час між сусідніми точками регенерації.

Будується матриця перехідних ймовірностей

$$[P_{ij}] = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \dots & p_{0n} \\ p_{10} & p_{11} & p_{12} & \dots & p_{1n} \\ p_{20} & p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ p_{n0} & p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix}, \quad (2.20)$$

де p_{ij} – імовірність переходу зі стану i в стан j за час між сусідніми точками регенерації. За матрицею (2.20) та методикою підрозділу 1.2.2 можна скласти систему лінійних алгебраїчних рівнянь для ймовірностей станів:

$$p_j = \sum_{i=0}^n p_i \cdot p_{ij}, \quad j = 0, 1, 2, \dots, n, \quad (2.21)$$

яка доповнюється умовою нормування: $\sum_{j=0}^n p_j = 1$.

Розв'язувати систему (2.21) доцільно за допомогою однієї із систем комп'ютерної математики, наприклад Matlab.

Для одноканальної СМО виду M/G/1 з пуассонівським вхідним потоком, який характеризується інтенсивністю λ і функцією розподілу часу обслуговування $B(t)$ отримаємо $p_{ij} = k_{j-i+1}$, $j > 0$; $p_{0j} = k_j$ [20, с. 152], тобто матриця перехідних ймовірностей (2.20) перетвориться на таку:

$$[P_{ij}] = \begin{bmatrix} k_0 & k_1 & k_2 & k_3 & \dots \\ k_0 & k_1 & k_2 & k_3 & \dots \\ 0 & k_0 & k_1 & k_2 & \dots \\ 0 & 0 & k_0 & k_1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix},$$

де $k_m = \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^m}{m!} dB(t), m \geq 0$ і $k_m = 0, m < 0$.

Щоб проілюструвати використання методу вкладених ланцюгів Маркова, розглянемо СМО з груповим вибором запитів (аналогічно до прикладу із підрозділу 2.3.1), але за умови, що вхідний потік запитів описується функцією розподілу $A(t)$, а процес вибору запитів описується експоненціальним розподілом з інтенсивністю μ [21]. За стан вкладеного ланцюга Маркова оберемо кількість вимог у накопичувачі безпосередньо перед надходженням чергової вимоги. Позначимо через p_{ij} ймовірність переходу за час τ між двома точками регенерації. За час τ може надійти запит на перепис вмісту накопичувача, ймовірність цього – $1 - \exp(-\mu\tau)$. Тоді процес перейде зі стану s_i в стан $s_0, i = \overline{0, k}$. Ймовірність цієї події – $p_{i0} = \int_0^{\infty} (1 - e^{-\mu t}) dA(t) = 1 - \alpha(\mu)$, де $\alpha(\mu)$ – перетворення Лапласа-Стілтєса від $A(t)$ за параметром μ .

За час τ обслуговуючий пристрій може не вибрати вмісту накопичувача, ймовірність цього – $\exp(-\mu\tau)$. Отже процес перейде зі стану i у стан $i + 1, i = \overline{0, k - 1}$. Ймовірність цієї події – $p_{i,i+1} = \int_0^{\infty} e^{-\mu t} dA(t) = \alpha(\mu)$.

Якщо процес перебуває в стані k , то з ймовірністю $p_{kk} = \alpha(\mu)$ він залишиться у цьому самому стані, тобто запит, який нещодавно прийшов, буде втрачено. Матриця перехідних ймовірностей виглядає так:

$$\|P_{ij}\| = \begin{pmatrix} 1 - \alpha(\mu) & \alpha(\mu) & 0 & \dots & 0 \\ 1 - \alpha(\mu) & 0 & \alpha(\mu) & \dots & 0 \\ 1 - \alpha(\mu) & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 - \alpha(\mu) & 0 & 0 & \dots & \alpha(\mu) \end{pmatrix}.$$

Усі стани матриці $\|P_{ij}\|$ – неперіодичні, ненульові й нерозкладні. Отже для досліджуваного ланцюга існує стаціонарний стан, що характеризується ймовірностями $p_i, i = \overline{0, k}$ того, що в накопичувачі перед моментом регенерації було $i = \overline{0, k}$ запитів. Для цих ймовірностей справедлива система рівнянь

$$p_j = \sum_{i=0}^k p_i \cdot p_{ij}, j = 0, 1, 2, \dots, k, \quad (2.22)$$

а умовою нормування є

$$\sum_{j=0}^k p_j = 1. \quad (2.23)$$

Розв'язуючи систему рівнянь (2.22) одночасно з умовою нормування, отримуємо:

$$p_j = [1 - \alpha(\mu)] [\alpha(\mu)]^j, 0 \leq j < k; p_k = [\alpha(\mu)]^k.$$

Середня кількість запитів у накопичувачі –

$$L_q = \sum_{j=1}^k j p_j = \alpha(\mu) \{1 - [\alpha(\mu)]^k\} / [1 - \alpha(\mu)].$$

Важливою характеристикою функціонування СМО з накопичувачем є ймовірність відмови щодо обслуговування.

На рисунку 2.7 наведено графіки залежності ймовірності відмови щодо обслуговування p_k як функція об'єму накопичувача k і інтенсивності зчитування μ для трьох типів вхідних потоків при їх середній інтенсивності, що дорівнює восьми:

- а – гіперекспоненціальний вхідний потік з функцією розподілу $A(t) = 1 - [0,3 \exp(-2,79t) + 0,7 \exp(-40t)]$ і коефіцієнтом варіації $\nu = 1,996$;
- б – пуассонівський вхідний потік з функцією розподілу $A(t) = 1 - \exp(-8t)$ коефіцієнтом варіації $\nu = 1$;
- в – ерлангівський вхідний потік з функцією розподілу $A(t) = 1 - \sum_{i=1}^3 \frac{(24t)^{i-1}}{(i-1)!} e^{-24t}$ і коефіцієнтом варіації $\nu = 0,5773$.

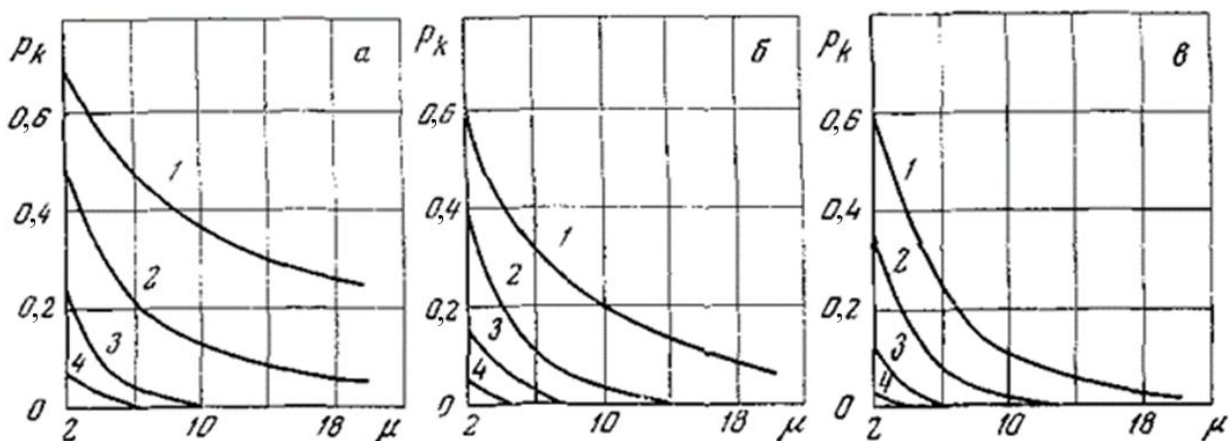


Рисунок 2.7 – Імовірність відмови щодо обслуговування

Зрозуміло, що коефіцієнт варіації істотно впливає на характеристики системи.

Характеристики, одержані за методом вкладених ланцюгів Маркова, справедливі лише для моментів регенерації і не повністю описують особливості СМО. Повну інформацію містить лише характеристики для довільного моменту часу, пов'язані певними співвідношеннями з характеристиками для точок регенерації. Подамо ці співвідношення.

Нехай t_n ($n \geq 1$) – моменти надходження запитів, $z(t)$ – кількість запитів у накопичувачі в момент часу t , $\lambda_p(i)$ – інтенсивність підпоследовності тих моментів часу, для яких кількість запитів у накопичувачі дорівнює i . Тоді

$$p_i = P\{z(t_n - 0) = i\} = \lambda_p(i) / \lambda, \quad (2.24)$$

де λ – середня інтенсивність вхідного потоку запитів.

Позначимо через $N_i(t)$ кількість переходів кількості запитів у накопичувачі типу $i \rightarrow i + 1$ за проміжок часу $(0, t)$, а через $M_i(t)$ – кількість переходів типу $i + 1 \rightarrow 0$ ($i < k$) за той же проміжок часу $(0, t)$. Справедливою буде така рівність:

$$\left| N_i(t) - \sum_{j=i}^{k-1} M_j(t) \right| \leq 1. \quad (2.25)$$

Середньою (E) кількістю переходів $i \rightarrow i + 1$ буде за час t $E[N_i(t)] = \lambda_p(i)t$. Використавши (2.24), отримаємо $E[N_i(t)] = \lambda p_i t$. Середня кількість переходів типу $i + 1 \rightarrow 0$ за час t

$$E[M_i(t)] = \int_0^t \mu x p_{\xi_{j+1}(t)} dx, \quad (2.26)$$

де $p_{\xi_{j+1}(t)}$ – розподіл випадкової частки часу в $(0, t)$, протягом якого в накопичувачі містилося $i + 1$ запитів. Оскільки $E[\xi_{j+1}(t)] = \int_0^t x p_{\xi_{j+1}(x)} dx = \bar{p}_{j+1} t$, де

\bar{p}_{j+1} – імовірність перебування в накопичувачі $j + 1$ запитів в довільний момент часу, то із (2.26) отримаємо: $E[M_i(t)] = \mu \bar{p}_{j+1} t$. Із нерівності (2.25) випливає:

$$\lim_{t \rightarrow \infty} E[N_i(t)]/t = \lim_{t \rightarrow \infty} \sum_{j=i}^{k-1} E[M_j(t)]/t \quad \text{або} \quad \sum_{j=1}^{k-1} \bar{p}_{j+1} = (\lambda / \mu) p_i, \quad 0 \leq i < k.$$

Звідси остаточно отримаємо поєднання ймовірностей стану СМО в довільний момент часу (\bar{p}_i) з ймовірностями стану в точках регенерації (p_i):

$$\bar{p}_0 = 1 - \frac{\lambda}{\mu} p_0; \quad \bar{p}_i = \frac{\lambda}{\mu} (p_{i-1} - p_i), \quad 0 < i < k; \quad \bar{p}_k = \frac{\lambda}{\mu} p_{k-1}.$$

Відповідно скоректуються вирази для характеристик СМО. Приміром імовірність відмови щодо обслуговування – $\bar{P}_{\text{відм}} = \frac{\lambda}{\mu} [1 - \alpha(\mu)] [\alpha(\mu)]^{k-1}$.

2.3.4 Метод включення додаткових змінних

Метод включення додаткових змінних є одним із найуніверсальніших під час дослідження немарківських систем масового обслуговування, але його застосування передбачає використання апарату інтегро-диференціальних рівнянь і розв'язок не завжди вдається отримати в квадратурах [22, с. 76]. Метод включення додаткових змінних дозволяє перетворення вихідного немарківського процесу на марківський шляхом введення додаткових змінних для опису стану системи.

Нехай $n(t)$ – стан вихідної системи в момент часу t , $p_n(t)$ – імовірність цього стану. Якщо вхід системи марківський або його можна звести до марківського, то стан системи доповнюється змінною u – час, що минув із початку обслуговування запиту, що міститься на приладі. Якщо ж, навпаки, процес обслуговування – марківський, то стан системи $n(t)$ доповнюється змінною v – час, що минув із моменту надходження останнього запиту. Як наслідок вихідний процес перетворюється на марківський.

Використання цього методу розглянемо на прикладі системи масового обслуговування з груповим вибором запитів із накопичувача, у якій вхідний потік – пуассонівський з інтенсивністю λ , об'єм накопичувача – k запитів, а інтервали часу між сусідніми зверненнями до накопичувачу за запитами мають розподіл $B(t)$. Звернення здійснюється обслуговуючим пристроєм після оброблення попередньої групи запитів. Позначимо через $r(u)$ миттєву інтенсивність зчитування запитів з накопичувача, яка визначається як $r(u) = B'(u)[1 - B(u)]^{-1}$, через μ – середню інтенсивність зчитування запитів, через $\beta(s) = \int_0^{\infty} e^{-st} dB(t)$ – перетворення Лапласа – Стілтєса від $B(t)$.

Нехай $[\delta]$ – стан, коли в системі немає запитів, $[n, u]$ – стан, коли в накопичувачі n запитів, канал обслуговування обробляє чергову групу запитів і час, що минув після попереднього звернення до накопичувача, дорівнює u . Позначимо через $p_{\delta}(t)$ імовірність стану $[\delta]$, через $p_n(t)$ – імовірність того, що в накопичувачі міститься n запитів, а через $p_n(t, u)$ – щільність імовірності за u того, що в накопичувачі міститься n запитів, а час, що минув після

попереднього звернення до накопичувача (час обслуговування), дорівнює $u_0, u < u_0 < u + \Delta$. $p_n(t), p_n(t, u)$ поєднані співвідношенням $p_n(t) = \int_0^{\infty} p_n(t, u) du$.

Розглянемо можливі переходи системи з одного стану в інший за інтервал $(t, t + \Delta)$. Зафіксуємо момент часу і визначимо імовірність $p_0(t + \Delta, u + \Delta)\Delta$ – того, що в момент $t + \Delta$ система буде перебувати в стані $[0, u + \Delta]$. Це може статися лише за однієї умови: система в момент t перебувала в стані $[0, u]$, а за час Δ з ймовірністю $1 - r(u)\Delta$ не закінчилося обслуговування групи запитів, які вже перебували на обробці (імовірність цього за визначенням $p_n(t, u)$), і з імовірністю $1 - \lambda\Delta$ в систему не надійшов новий запит. Справедливість другого твердження впливає з того, що якщо за час Δ надійде запит, то він одразу буде обраний для обслуговування і система зі стану $[\delta]$ в момент t перейде в стан $[0, \Delta]$ при $t + \Delta$, хоча це суперечить вихідному твердженню. Таким чином,

$$p_0(t + \Delta, u + \Delta)\Delta = p_0(t, u)\Delta\{[1 - r(u)\Delta][1 - \lambda\Delta]\} + o(\Delta). \quad (2.27)$$

Визначимо ймовірність того, що в момент часу $t + \Delta$ система буде перебувати в стані $[n, u + \Delta]$. Цього можна досягти двома способами:

у момент часу t система перебувала в стані $[n, u]$, за час Δ не надійшов новий запит і не закінчилося обслуговування поточної групи запитів;

у момент часу t система перебувала в стані $[n - 1, u]$, і за час Δ з імовірністю $\lambda\Delta$ в систему надійшов запит.

Таким чином,

$$p_n(t + \Delta, u + \Delta)\Delta = p_n(t, u)\Delta\{[1 - r(u)\Delta][1 - \lambda\Delta]\} + p_{n-1}(t, u)\Delta\lambda\Delta + o(\Delta). \quad (2.28)$$

Визначимо граничні умови в точці $u = 0$. Для цього знову зафіксуємо момент часу t і введемо ймовірність того, що в момент $t + \Delta$ система буде перебувати в стані $[n, \Delta]$, тобто з моменту початку обслуговування чергової групи запитів минув час, що не перевищує Δ . Попередньо введемо рівняння для стану $[0, \Delta]$. Система може перейти в стан $[0, \Delta]$ за час Δ двома способами:

– у момент часу t система перебувала в стані $[\delta]$ і за час Δ з імовірністю $\lambda\Delta$ надійшов новий запит, який відразу ж був обраний на обслуговування;

– у момент часу t система перебувала в стані $[n, u]$ і з імовірністю $r(u)\Delta$ закінчилося обслуговування попередньої групи запитів, унаслідок чого накопичувач обнулився. Імовірність цієї події буде складатися з ймовірностей за всіма можливими значеннями n і всіма можливими значеннями u .

Таким чином,

$$p_0(t + \Delta, \Delta)\Delta = p_\delta(t)\lambda\Delta + \sum_{n=1}^k \int_0^\infty p_n(t, u)r(u)du\Delta. \quad (2.29)$$

Оскільки за час Δ в систему не може надійти більше одного запиту й накопичувач у разі звернення до нього обнуляється, то

$$p_k(t, \Delta) \equiv 0. \quad (2.30)$$

Виведемо ймовірність того, що в момент $t + \Delta$ система буде перебувати в стані $[\delta]$. Цього можна досягти двома способами:

– у момент часу t система перебувала в стані $[\delta]$ і з ймовірністю $1 - \lambda\Delta$ новий запит не надійде;

– система в момент t перебувала в стані $[0, u]$ і за час Δ з ймовірністю $r(u)\Delta$ закінчилося обслуговування чергової групи запитів.

Таким чином,

$$p_\delta(t + \Delta) = p_\delta(t)[1 - \lambda\Delta] + \int_0^\infty p_0(t, u)r(u)du\Delta. \quad (2.31)$$

Розкриємо дужки у виразах (2.27) і (2.28), перенесемо $p_n(t, u)$ в ліву частину, сюди ж додамо $p_n(t, u + \Delta) - p_n(t, u + \Delta)$, усе розділимо на Δ і перейдемо до границі при $\Delta \rightarrow 0$. Отримаємо систему рівнянь

$$\begin{aligned} \frac{\partial}{\partial t} p_0(t, u) + \frac{\partial}{\partial u} p_0(t, u) &= -[\lambda + r(u)]p_0(t, u), \\ \frac{\partial}{\partial t} p_n(t, u) + \frac{\partial}{\partial u} p_n(t, u) &= -[\lambda + r(u)]p_n(t, u) + \lambda p_{n-1}(t, u). \end{aligned} \quad (2.32)$$

Аналогічно із виразів (2.29) – (2.31) виводять граничні умови

$$\begin{aligned} p_0(t, 0)\Delta &= \lambda p_\delta(t) + \sum_{n=1}^k \int_0^\infty p_n(t, u)r(u)du \\ \frac{\partial}{\partial t} p_\delta(t) &= -\lambda p_\delta(t) + \int_0^\infty p_0(t, u)r(u)du, \\ p_n(t, 0) &\equiv 0, \quad 0 < n \leq k. \end{aligned} \quad (2.33)$$

Як вихідні умови оберемо $p_\delta(0) = 1$. Спільну систему рівнянь (2.32), (2.33) можна розв'язати чисельно на комп'ютері, наприклад методом сіток, використовуючи відповідне програмне забезпечення.

Розглянемо стаціонарний режим. При цьому $p_n(t, u) \rightarrow p_n(u)$, $\frac{\partial p_n(u)}{\partial t} = 0$.

Як наслідок отримаємо таку систему рівнянь:

$$\begin{aligned}\frac{\partial}{\partial u} p_0(u) &= -[\lambda + r(u)]p_0(u), \\ \frac{\partial}{\partial u} p_n(u) &= -[\lambda + r(u)]p_n(u) + \lambda p_{n-1}(u), 1 \leq n \leq k\end{aligned}\tag{2.34}$$

і граничні умови:

$$\begin{aligned}-\lambda p_\delta + \int_0^\infty p_0(u)r(u)du &= 0, \\ p_0(0) &= \lambda p_\delta + \sum_{n=1}^k \int_0^\infty p_n(u)r(u)du.\end{aligned}\tag{2.35}$$

Розв'язуючи поетапно систему рівнянь (2.34), отримуємо:

$$\begin{aligned}p_0(u) &= [1 - B(u)]p_0(0)e^{-\lambda u}, \\ p_1(u) &= \lambda u[1 - B(u)]p_0(0)e^{-\lambda u}, \\ p_n(u) &= \frac{(\lambda u)^n}{n!}[1 - B(u)]p_0(0)e^{-\lambda u}, 0 \leq n \leq k.\end{aligned}\tag{2.36}$$

Підставивши вираз (2.36) у перше рівняння системи (2.35), знайдемо

$$p_\delta = \beta(\lambda)p_0(0)/\lambda.$$

Отже,

$$p_n = \int_0^\infty p_n(u)du = \frac{p_0(0)}{\lambda} - \frac{(-1)^n}{n!} \lambda^{n-1} \beta^{(n)}(\lambda),$$

де $\beta^{(n)}(\lambda)$ – значення n -ї похідною від $\beta(s)$ при $s = \lambda$.

Використавши умову нормування $p_\delta + \sum_{n=0}^k p_n = 1$, отримаємо

$$p_0(0) = \lambda \cdot \left\{ k - \sum_{n=1}^k \frac{(-1)^n}{n!} \beta^{(n)}(\lambda) \right\}^{-1}.$$

Для ерлангівського розподілу m -го порядку

$$\beta^{(n)}(\lambda) = (-1)^n \frac{(m+n-1)!}{(m-1)!} \frac{\mu^m}{(\mu + \lambda)^{m+n}}.$$

Для гіперекспоненціального розподілу $\beta^{(n)}(\lambda) = (-1)^n n! \sum_{i=1}^m \frac{b_i \mu_i}{(\mu_i + \lambda)^{n+1}}$.

Базові характеристики досліджуваної системи:

– імовірність відмови щодо обслуговування – $P_{vidm} = p_k$;

– середня кількість запитів у буфері – $L_q = \sum_{i=1}^k ip_i$;

– середній час перебування запитів у буфері – $\tau_q = \sum_{i=1}^k \frac{p_i}{\mu_i}$.

Таким чином, для використання метода включення додаткових змінних потрібно, щоб один із потоків подій СМО був пуассонівським, або зводився до нього, а функція розподілу другого мала скінченне перетворення Лапласа – Стілтєса.

2.3.5 Формули Літтла

Для систем масового обслуговування, які функціонують у стаціонарному режимі, найсприятливішими є формули Літтла, які пов'язують важливіші характеристики СМО.

Нехай СМО, на яку поступає потік запитів з інтенсивністю λ , функціонує в стаціонарному режимі досить великий інтервал часу T . Тоді за цей час у систему надійде в середньому λT запитів. Кожен запит перебуває в черзі середній час τ_q ; тоді середній час W , який перебуватимуть у черзі всі λT запитів, обчислюється за формулою:

$$W = \lambda \tau_q T . \quad (2.37)$$

З іншого боку, якщо L_q – середня кількість запитів, які очікують обслуговування, то сумарний середній час W , який вони перебуватимуть у черзі за час T , обчислюється за формулою:

$$W = L_q T . \quad (2.38)$$

Порівнявши праві частини виразів (2.37) і (2.38) і скоротивши T , отримаємо

$$L_q = \lambda \tau_q . \quad (2.39)$$

Кожен запит перебуває в системі середній час τ_s ; тоді середній час V , який перебувають у системі всі λT запитів, обчислюється за формулою:

$$V = \lambda \tau_s T . \quad (2.40)$$

З іншого боку, якщо L_s – середня кількість запитів у системі, то середній час τ_s , який вони перебуватимуть у системі за час T , обчислюється за формулою:

$$V = L_s T . \quad (2.41)$$

Порівнявши праві частини виразів (2.40) і (2.41) і скоротивши T , отримаємо:

$$L_s = \lambda \tau_s. \quad (2.42)$$

Формули (2.39) і (2.42) були названі формулами Літтла на честь дослідника Дж. Д. К. Літтла, який їх обґрунтував [19, с. 33]. Ці формули відіграють важливу роль у теорії масового обслуговування, оскільки дають змогу за певних умов унеможливити складні перетворення під час виведення характеристик системи.

Контрольні питання

1. Що таке система масового обслуговування?
2. З яких елементів складається система масового обслуговування?
3. Які економічні процеси описуються за допомогою систем масового обслуговування?
4. Якими характеристиками оцінюється система масового обслуговування?
5. Що таке потік подій системи масового обслуговування і як він описується?
6. Як описується процес обслуговування в СМО?
7. Запишіть формулу Літтла.
8. Як описуються системи масового обслуговування в короткій формі?
9. Якими методами досліджуються системи масового обслуговування?
10. Що таке граф станів і переходів системи масового обслуговування?
11. Які критерії ефективності використовуються в системах масового обслуговування?
12. Які закони розподілу ймовірностей використовуються в системах масового обслуговування?
13. У чому полягає сутність диференціального методу дослідження систем масового обслуговування?
14. У чому полягає сутність методу фаз дослідження систем масового обслуговування?
15. У чому полягає сутність методу включення додаткових змінних дослідження систем масового обслуговування?

3 ТИПОВІ СИСТЕМИ МАСОВОГО ОБСЛУГОВУВАННЯ

3.1 Система масового обслуговування з відмовами

СМО з відмовами складається з n каналів обслуговування, на які надходить найпростіший потік запитів із інтенсивністю λ , час обслуговування розподілений за експоненціальним законом із параметром μ , тобто потік обслуговування також найпростіший за параметром μ . Черги не спостерігається, тобто запит, який надійшов коли всі канали зайняті, отримує відмову щодо обслуговування. Ця система є однією з найпростіших і на ній будується теорія систем масового обслуговування. Вперше її розглянув видатний дослідник А. К. Ерланг [4, с. 69]. Типовим прикладом застосування СМО з відмовами є автоматичні телефонні станції. Тут запитом, який надійшов у систему, є звернення клієнта до телефонної станції. Якщо потрібна лінія зв'язку вже зайнята розмовою, абонент отримує відмову. Автоматична телефонна станція дає часті гудки й запит губиться [16, с. 33].

Дослідимо її. Позначимо через S_i стан системи, коли в ній перебуває i , $i = 0, 1, \dots, n$ запитів. Граф станів і переходів для цієї системи при $n = 5$ зображено на рисунку 3.1.

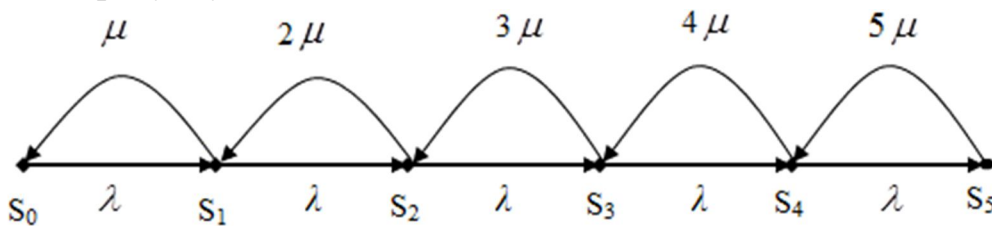


Рисунок 3.1 – Граф станів і переходів СМО з відмовами

Позначимо через $p_i(t)$ ймовірність стану S_i . Використовуючи методику підрозділу 1.2.3, складемо систему диференціальних рівнянь для станів $s_i, i = 0, 1, 2, \dots, n$:

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p'_1(t) &= -(\lambda + \mu) p_1(t) + \lambda p_0(t) + 2\mu p_2(t), \\ p'_i(t) &= -(\lambda + i\mu) p_i(t) + \lambda p_{i-1}(t) + (i+1)\mu p_{i+1}(t), 1 \leq i < n, \\ p'_n(t) &= -n\mu p_n(t) + \lambda p_{n-1}(t). \end{aligned} \quad (3.1)$$

Початкові умови –

$$p_0(0) = 1, p_i(0) = 0, 1 \leq i < n. \quad (3.2)$$

Якщо n невелике, то, використовуючи технологію розв'язку систем лінійних диференціальних рівнянь [15, с. 54] при конкретних значеннях λ і μ ,

можна знайти аналітичний розв'язок цієї системи. Приміром, при $n=2$; $\lambda=1$; $\mu=0,55$ розв'язок системи (3.1) виглядає так:

$$\begin{aligned} p_0(t) &= 0,645 \cdot e^{-1,03t} + 0,12 \cdot e^{-2,61t} + 0,224, \\ p_1(t) &= -0,040 \cdot e^{-1,03t} - 0,366 \cdot e^{-2,61t} + 0,407, \\ p_2(t) &= -0,605 \cdot e^{-1,03t} + 0,25 \cdot e^{-2,61t} + 0,370. \end{aligned}$$

Графіки залежностей функцій $p_0(t), p_1(t), p_2(t)$ від часу наведено на рисунку 3.2.

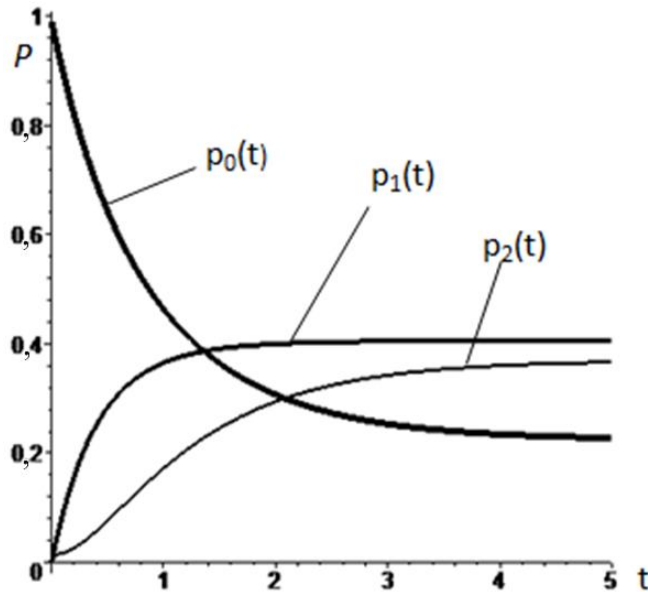


Рисунок 3.2 – Залежності ймовірностей станів системи з відмовами від часу

Зрозуміло, що в разі необмеженого зростання t поведінка функцій $p_i(t)$ перестане залежати від часу й система переходить у стаціонарний режим.

Тобто, $\lim_{t \rightarrow \infty} p_i(t) = p_i$, $i = 0, 1, 2$, де p_i – константи. З аналізу виразів для функцій $p_i(t)$ випливає, що при $t \rightarrow \infty$ вони дорівнюють відповідним константам у правій частині, тобто $p_0 = 0,224$; $p_1 = 0,406$; $p_2 = 0,370$.

Ймовірності станів системи в стаціонарному режимі можна визначити більш простим методом. З огляду на те що $\lim_{t \rightarrow \infty} p'_i(t) = \lim_{t \rightarrow \infty} p'_i = 0$, система рівнянь (3.1) перетворюється так:

$$\begin{aligned} -\lambda p_0 + \mu p_1 &= 0, \\ -(\lambda + \mu) p_1 + \lambda p_0 + 2\mu p_2 &= 0, \\ -(\lambda + i\mu) p_i + \lambda p_{i-1} + (i+1)\mu p_{i+1} &= 0, 1 \leq i < n, \\ -n\mu p_n + \lambda p_{n-1} &= 0. \end{aligned} \tag{3.3}$$

Умова нормування –
$$\sum_{i=0}^n p_i = 1. \quad (3.4)$$

Поділимо кожне рівняння системи (3.3) на μ і введемо параметр $\rho = \lambda / \mu$ навантаження системи по одному каналу. Система рівнянь (3.4) стане такою:

$$\begin{aligned} -\rho p_0 + p_1 &= 0, \\ -(\rho + 1)p_1 + \rho p_0 + 2p_2 &= 0, \\ -(\rho + i)p_i + \rho p_{i-1} + (i+1)p_{i+1} &= 0, 1 \leq i < n, \\ -np_n + \rho p_{n-1} &= 0. \end{aligned} \quad (3.5)$$

Послідовно розв'язуючи систему (3.5) відносно p_0 , отримаємо: $p_i = \frac{\rho^i}{i!} p_0$.

Підставивши цей вираз у (3.4), визначимо $p_0 = \left(1 + \sum_{i=1}^n \frac{\rho^i}{i!}\right)^{-1}$.

Остаточно ймовірність того, що обслуговується i запитів, тобто система перебуває в стані S_i , визначається за виразом

$$p_i = \frac{\rho^i}{i!} \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}, 0 \leq i \leq n. \quad (3.6)$$

Цей вираз було названо формулою Ерланга. Доведено, що ця формула справедлива для будь якого неперервного закону розподілу часу обслуговування на каналі [16, с. 35].

Базові характеристики досліджуваної СМО з відмовами такі:

– імовірність відмови щодо обслуговування $P_{vidm} = p_n = \frac{\rho^n}{n!} \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}$.

Цей вираз означає, що цьому запиту буде відмовлено щодо обслуговування, якщо всі лінії будуть зайняті;

– абсолютна пропускна здатність дорівнює середній кількості запитів, які може обслужити СМО за одиницю часу, тобто це частка вхідного потоку заявок, яка надходить у систему й обслуговується нею:

$$A = \lambda Q = \lambda \left[1 - \frac{\rho^n}{n!} \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1} \right]; \quad (3.7)$$

– середня кількість запитів у системі, а відповідно й середня кількість каналів, зайняти обслуговуванням запитів, обчислюється як математичне сподівання:

$$L_s = \sum_{i=1}^n i p_i = \sum_{i=1}^n i \frac{\rho^i}{i!} \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}; \quad (3.8)$$

– коефіцієнт завантаженості системи $K_z = L_s / n = \sum_{i=1}^n i \frac{\rho^i}{ni!} \left(\sum_{k=0}^n \frac{\rho^k}{k!} \right)^{-1}$.

Доцільно знати залежності базових характеристик системи від n і коефіцієнта завантаженості за всіма каналами $\alpha = \lambda / (n\mu) = \rho / n$. Звідси $\rho = n\alpha$. Тоді базові характеристики будуть такими:

– імовірність відмови щодо обслуговування $P_{vidm} = p_n = \frac{(n\alpha)^n}{n!} \left(\sum_{k=0}^n \frac{(n\alpha)^k}{k!} \right)^{-1}$;

– коефіцієнт завантаженості системи $K_z = L_s / n = \sum_{i=1}^n i \frac{(n\alpha)^i}{ni!} \left(\sum_{k=0}^n \frac{(n\alpha)^k}{k!} \right)^{-1}$.

На рисунку 3.3 зображено графік залежності ймовірності відмови щодо обслуговування від кількості каналів обслуговування при різних ρ .

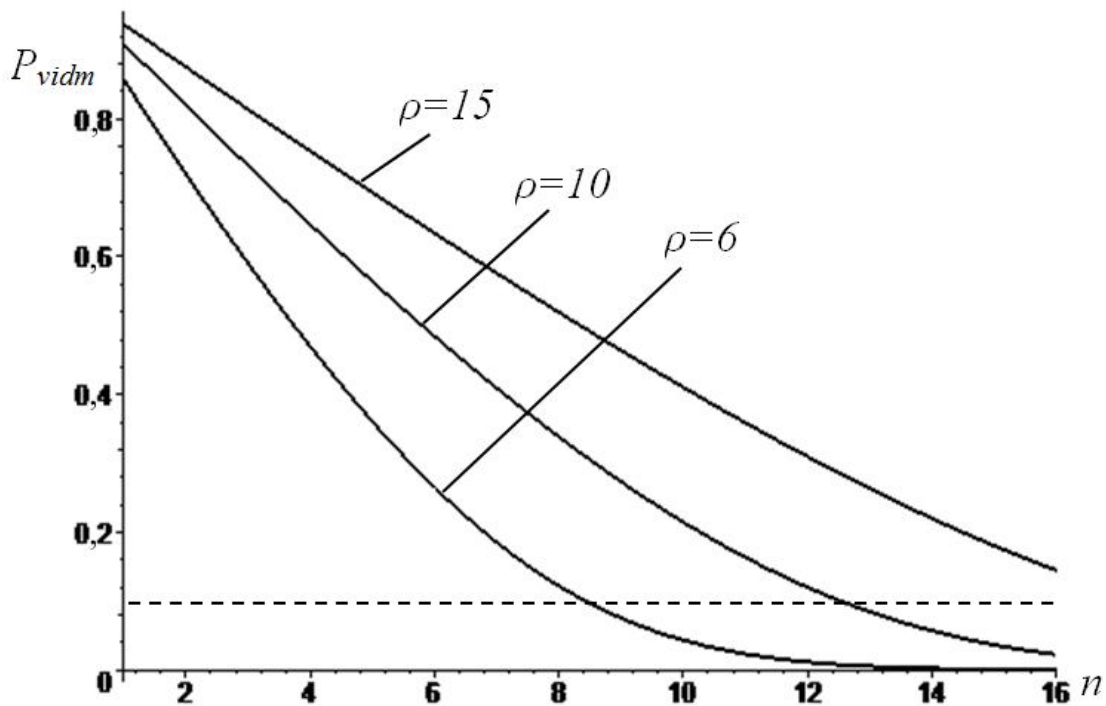


Рисунок 3.3 – Залежність ймовірності відмови щодо обслуговування для СМО з відмовами

На рисунку 3.4 зображено графік залежності коефіцієнта завантаженості системи від кількості каналів обслуговування при різних ρ .

За цими графіками можна обирати кількість каналів обслуговування при різних коефіцієнтах завантаженості одного каналу ρ залежно від

технічних завдань на проектування системи. Приміром, якщо потрібно обрати кількість каналів обслуговування в разі імовірності відмов 0,1 при $\rho = 10$, то з рисунка 3.3 випливає, що їх має повинна бути не менше 13 одиниць.

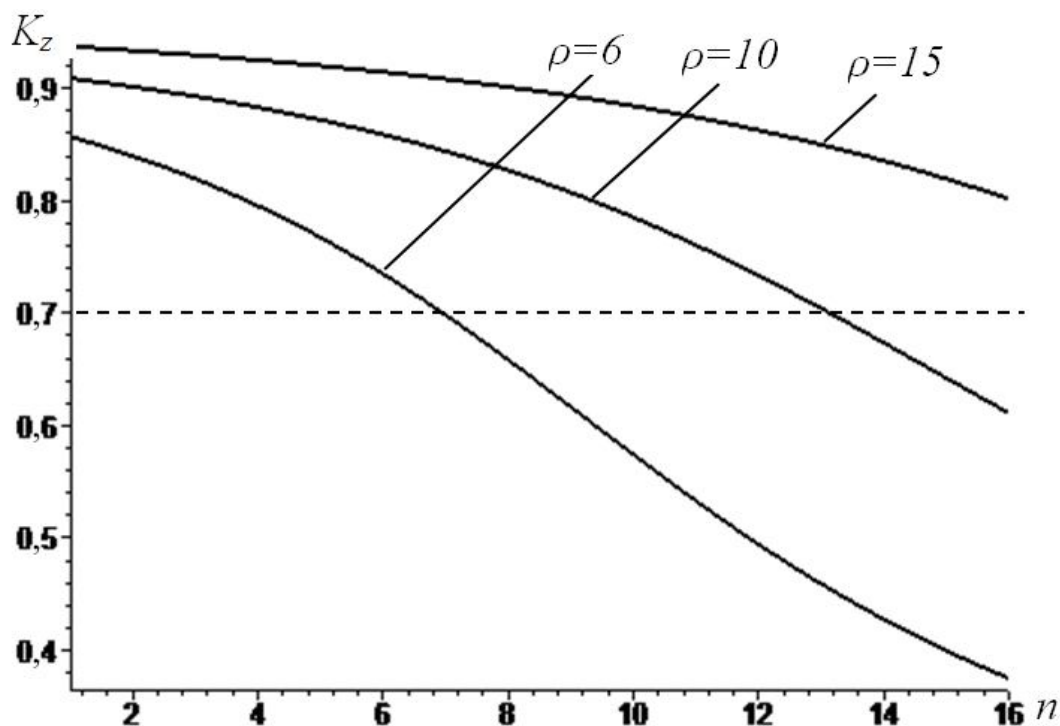


Рисунок 3.4 – Залежність коефіцієнта завантаженості системи для СМО з відмовами

Якщо кількість каналів обирається залежно від коефіцієнта завантаження системи, то потрібно використовувати графіки, подані на рисунку 3.4. Приміром, якщо потрібно обрати кількість каналів обслуговування в разі коефіцієнта завантаженості системи $K_z = 0,7$ при $\rho = 10$, то з рисунка 3.4 випливає, що їх має бути не більше 14 одиниць.

3.2 Система з пуассонівським вхідним потоком і експоненціальним часом обслуговування (М/М/1)

Це одна з найдослідженіших СМО [19, с. 112], для якої отримано в скінчених виразах майже всі характеристики. Нехай в систему надходить пуассонівський вхідний потік $(p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t})$ з інтенсивністю λ , час обслуговування розподілено за експоненціальним законом $(B(t) = 1 - e^{-\mu t})$ з інтенсивністю μ . Черга не обмежена. Граф станів і переходів для цієї системи зображено на рисунку 3.5.

$$\begin{aligned}
p_0(t + \Delta t) &= p_0(t) + \Delta t[-\lambda p_0(t) + \mu p_1(t)], \\
p_1(t + \Delta t) &= p_1(t) + \Delta t[-(\lambda + \mu)p_1(t) + \lambda p_0(t) + \mu p_2(t)], \\
&\dots\dots\dots \\
p_i(t + \Delta t) &= p_i(t) + \Delta t[-(\lambda + \mu)p_i(t) + \lambda p_{i-1}(t) + \mu p_{i+1}(t)], i \geq 1, \\
&\dots\dots\dots \\
p_0(0) &= 1, p_i(0) = 0, i > 0.
\end{aligned}
\tag{3.11}$$

Для розв'язання системи (3.11) оберемо чисельний метод Ейлера розв'язку систем диференціальних рівнянь [24, с. 445]. Позначимо через p_i^k значення імовірності $p_i(t)$ на k -му кроці обчислень. Система (3.11) переписеться у вигляді, готовому для ітераційного обчислення:

$$\begin{aligned}
p_0^{k+1} &= p_0^k + h \cdot (-\lambda p_0^k + \mu p_1^k), k = 0, 1, 2, \dots, \\
p_i^{k+1} &= p_i^k + h \cdot [-(\lambda + \mu)p_i^k + \lambda p_{i-1}^k + \mu p_{i+1}^k], i > 0, k = 0, 1, 2, \dots, \\
p_0^0 &= 1, p_i^0 = 0, i \geq 1,
\end{aligned}
\tag{3.12}$$

де h крок інтегрування.

Для чисельного розв'язання системи диференціальних рівнянь (3.9) за формулами (3.12) потрібно задати конкретні значення параметрів λ, μ і обрати крок інтегрування h . Наявні програмні системи не можуть забезпечити розв'язання системи рівнянь з нескінченною кількістю рівнянь, тому під час використання формул (3.12) обмежмося певною їх кількістю. Підбирати її доведеться експериментально, до того ж розв'язок буде наближеним.

На рисунку 3.6 зображено графіки ймовірностей стану системи М/М/1 для перехідного режиму при $\lambda = 0,6$; $\mu = 1$; $h = 0,06$.

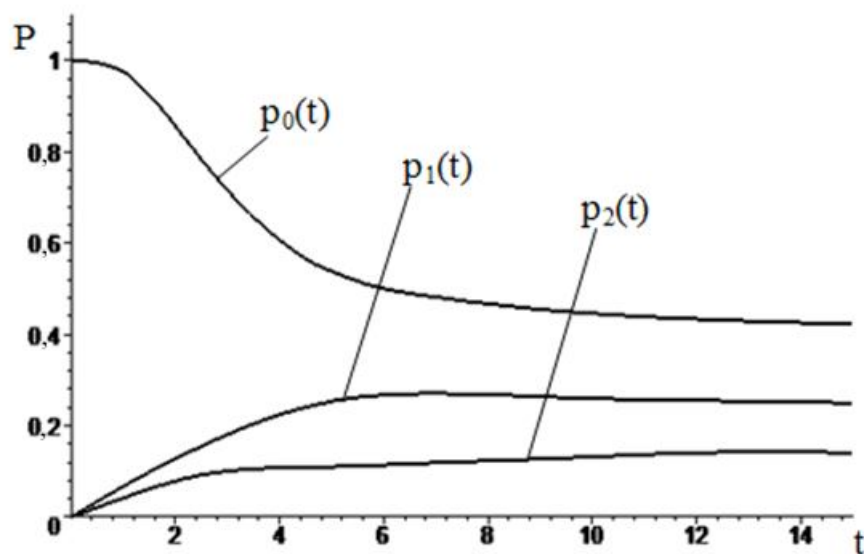


Рисунок 3.6 – Перехідний режим системи М/М/1

Зрозуміло, що з моменту часу $t = 8$ система переходить в стаціонарний режим функціонування.

Тепер дослідимо стаціонарний режим функціонування системи М/М/1. При стаціонарнім режимі $p_i(t) \rightarrow p_i, p'_i = 0$ і система (3.9) набуде такого вигляду:

$$\begin{aligned}
 & -\lambda p_0 + \mu p_1 = 0, \\
 & -(\lambda + \mu)p_1 + \lambda p_0 + \mu p_2 = 0, \\
 & -(\lambda + \mu)p_2 + \lambda p_1 + \mu p_3 = 0, \\
 & \dots\dots\dots \\
 & -(\lambda + \mu)p_i + \lambda p_{i-1} + \mu p_{i+1} = 0, i \geq 1, \\
 & \dots\dots\dots
 \end{aligned} \tag{3.13}$$

Умова нормування

$$\sum_{i=0}^{\infty} p_i = 1. \tag{3.14}$$

Поділимо кожне рівняння системи (3.13) на μ і введемо параметр $\rho = \lambda / \mu$ навантаження системи. Система рівнянь (3.13) переписеться так:

$$\begin{aligned}
 & -\rho p_0 + p_1 = 0, \\
 & -(\rho + 1)p_1 + \rho p_0 + p_2 = 0, \\
 & -(\rho + 1)p_2 + \rho p_1 + p_3 = 0, \\
 & \dots\dots\dots \\
 & -(\rho + 1)p_i + \rho p_{i-1} + p_{i+1} = 0, i \geq 1, \\
 & \dots\dots\dots
 \end{aligned} \tag{3.15}$$

З першого рівняння системи (3.15) випливає, що $p_1 = \rho p_0$. Підставимо цей вираз у друге рівняння: $-(\rho + 1)\rho p_0 + \rho p_0 + p_2 = 0$. Після приведення подібних членів отримаємо: $p_2 = \rho^2 p_0$. Продовжуючи обчислення, отримаємо: $p_i = \rho^i p_0, i \geq 1$. Підставимо отримані вирази для p_i в умову нормування:

$$p_0 + \sum_{i=1}^{\infty} \rho^i p_0 = 1 \rightarrow p_0 \left(1 + \sum_{i=1}^{\infty} \rho^i \right) = 1.$$

Якщо $\rho < 1$, то сума $\sum_{i=1}^{\infty} \rho^i$ є геометричною прогресією, яка збігається до

$$\frac{\rho}{1 - \rho}. \text{ Тоді } 1 + \frac{\rho}{1 - \rho} = \frac{1}{1 - \rho}, \text{ звідки}$$

$$p_0 = 1 - \rho, p_i = (1 - \rho)\rho^i. \tag{3.16}$$

Визначимо базові характеристики функціонування системи. Для цього знадобиться виробляюча функція:

$$P(z) = \sum_{i=0}^{\infty} p_i z^i = (1 - \rho) \sum_{i=0}^{\infty} (\rho z)^i = \frac{(1 - \rho)}{1 - \rho z}, (\rho z < 1). \quad (3.17)$$

Середню кількість запитів у системі обчислимо, як математичне сподівання

$$\begin{aligned} L_s &= \sum_{i=1}^{\infty} i p_i = \sum_{i=1}^{\infty} i \rho^i = \sum_{i=0}^{\infty} \frac{\partial}{\partial z} (\rho z)^i \Big|_{z=1} = \frac{\partial}{\partial z} P(z) \Big|_{z=1} = \\ &= (1 - \rho) \frac{\partial}{\partial z} \frac{1}{1 - \rho z} \Big|_{z=1} = (1 - \rho) \frac{\rho}{(1 - \rho z)^2} \Big|_{z=1}. \end{aligned} \quad (3.18)$$

Остаточно отримаємо:

$$L_s = P'(1) = \frac{\rho}{1 - \rho}. \quad (3.19)$$

Середня кількість запитів у черзі –

$$L_q = \sum_{i=1}^{\infty} (i - 1) p_i = \sum_{i=1}^{\infty} i p_i - \sum_{i=1}^{\infty} p_i = \frac{\rho}{1 - \rho} - (1 - p_0) = \frac{\rho}{1 - \rho} - (1 - 1 + \rho).$$

Остаточно отримаємо:

$$L_q = \frac{\rho^2}{1 - \rho}. \quad (3.20)$$

Визначимо середній час перебування запитів у системі τ_s . Для цього застосуємо формулу Літтла (2.42) – $L_s = \lambda \tau_s$, звідки

$$\tau_s = \frac{\rho}{\lambda(1 - \rho)} = \frac{1}{\mu(1 - \rho)}. \quad (3.21)$$

Аналогічно, застосувавши формулу (2.39), визначимо середній час перебування в черзі: $\tau_q = \frac{L_q}{\lambda} = \frac{\rho^2}{\lambda(1 - \rho)}$. Після перетворення отримаємо:

$$\tau_q = \frac{\rho}{\mu(1 - \rho)}. \quad (3.22)$$

На рисунку 3.7 зображено залежність середнього часу перебування запитів у системі залежно від навантаження системи ρ .

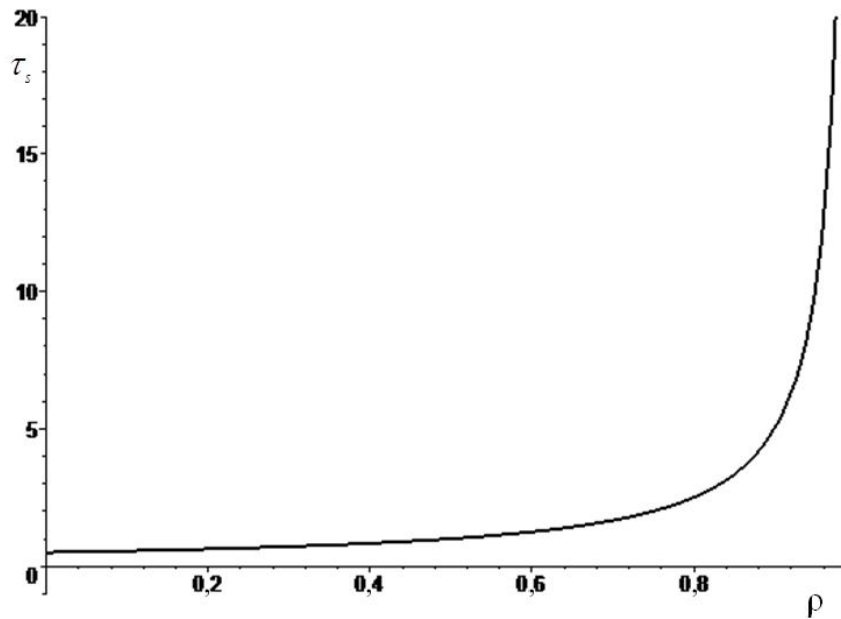


Рисунок 3.7 – Залежність середнього часу перебування запитів у системі залежно від навантаження системи ρ

У разі наближення ρ до одиниці час перебування запитів у системі різко зростає і наближається до нескінченності.

Для декількох СМО з різними дисциплінами обслуговування середній час перебування в системі, згідно з формулою Літгла, буде однаковим. Але для СМО з різними дисциплінами обслуговування дисперсія часу перебування в системі буде різною. Визначимо дисперсію часу перебування запитів у системі М/М/1.

Для обчислення дисперсії використаємо формулу (1.8), яка набуде такого вигляду:

$$D_s = \sum_{i=1}^{\infty} i^2 p_i - (L_s)^2, \quad (3.23)$$

де $\sum_{i=0}^{\infty} i^2 p_i$ другий початковий момент.

Із (3.17) випливає:

$$P'(z) = \sum_{i=0}^{\infty} p_i \frac{\partial}{\partial z} z^i = \sum_{i=0}^{\infty} p_i i z^{i-1}. \quad (3.24)$$

Помножимо обидві частину виразу (3.24) на z :

$$zP'(z) = \sum_{i=0}^{\infty} p_i i z^i. \quad (3.25)$$

Знайдемо похідну по z від (3.25):

$$P'(z) + zP''(z) = \sum_{i=0}^{\infty} p_i i^2 z^{i-1}. \quad (3.26)$$

Звідси

$$\sum_{i=0}^{\infty} i^2 p_i = \sum_{i=0}^{\infty} p_i i^2 z^{i-1} \Big|_{z=1} = P'(1) + P''(1). \quad (3.27)$$

Знайдемо $P''(z)$. Із (3.18) випливає:

$$P''(z) = \rho(1-\rho) \left[\frac{1}{(1-\rho z)^2} \right]' = \frac{\rho(1-\rho)[2(1-\rho z)\rho]}{(1-\rho z)^4}. \quad (3.28)$$

Із (3.19) і (3.28) випливає:

$$P'(1) + P''(1) = \frac{\rho}{1-\rho} + \frac{2\rho^2(1-\rho)^2}{(1-\rho)^2} = \frac{\rho + \rho^2}{(1-\rho)^2}. \quad (3.29)$$

Підставивши (3.29) у (3.27), а знайдений вираз для $\sum_{i=0}^{\infty} i^2 p_i$ і $(L_s)^2$ у (3.23)

остаточно отримаємо:

$$D_s = \frac{\rho}{(1-\rho)^2}. \quad (3.30)$$

З виразу (3.30) випливає, що зі зростанням ρ дисперсія часу перебування запитів у системі М/М/1 також зростає, і, коли навантаження системи наближається до одиниці, прямує до нескінченості.

Функція розподілу часу перебування запитів у черзі описується виразом $W(t) = 1 - \rho e^{-(\mu-\lambda)t}$ [25, с. 133], щільність розподілу $w(t) = \rho(\mu - \lambda)e^{-(\mu-\lambda)t}$. Визначимо середній час перебування запитів у черзі з формулою (1.14):

$$\begin{aligned} \tau_q &= \int_0^{\infty} t \rho(\mu - \lambda) e^{-(\mu-\lambda)t} dt = \rho(\mu - \lambda) \int_0^{\infty} \frac{t d e^{-(\mu-\lambda)t}}{-(\mu - \lambda)} = -\rho \left[t e^{-(\mu-\lambda)t} \Big|_0^{\infty} - \int_0^{\infty} e^{-(\mu-\lambda)t} dt \right] = \\ &= -\rho \left[\frac{e^{-(\mu-\lambda)t} \Big|_0^{\infty}}{-(\mu - \lambda)} \right] = \frac{\rho}{\mu - \lambda}. \end{aligned}$$

Під час інтегрування виразу було використано метод інтегрування частинами [24, с. 340]. Якщо винести μ за дужки, то остаточно отримаємо:

$$\tau_q = \frac{\rho}{\mu(1-\rho)}. \quad (3.31)$$

Вираз (3.31), отриманий за класичною схемою обчислення математичного сподівання, повністю співпадає з виразом (3.22), отриманим за формулою Літтла, що підтверджує її справедливність.

3.3 Багатоканальна система масового обслуговування з обмеженою чергою

Розглянемо пуассонівську систему масового обслуговування з n каналами обслуговування і кількістю міст для очікування m (M/M/n/m в позначеннях Кендала). Позначимо через $\{i\}$, $i = 1, 2, \dots, (n+m-1), (n+m)$ стан системи, коли в ній перебуває i запитів. Якщо $i \geq n$, то n запитів будуть обслуговуватися, а решта перебуває в черзі. Імовірність стану $\{i\}$ позначимо через p_i . Граф станів і переходів для цієї системи зображено на рисунку 3.8.

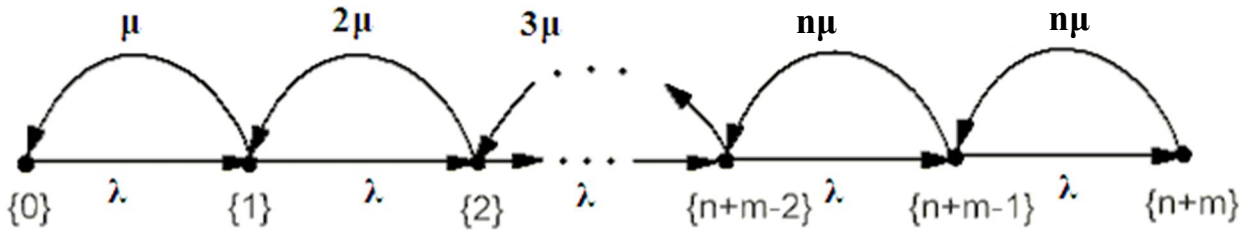


Рисунок 3.8 – Граф станів і переходів для системи M/M/n/m

Використовуючи методику п. 1.2.3, складемо систему лінійних рівнянь для ймовірностей $p_i, i = 1, 2, \dots, n + m$:

$$\begin{aligned}
 & -\lambda p_0 + \mu p_1 = 0, \\
 & -(\lambda + i\mu)p_i + \lambda p_{i-1} + (i+1)\mu p_i = 0, \quad 1 \leq i < n, \\
 & -(\lambda + n\mu)p_i + \lambda p_{i-1} + n\mu p_i = 0, \quad n \leq i < n + m, \\
 & -n\mu p_{n+m} + \lambda p_{n+m-1} = 0.
 \end{aligned} \tag{3.32}$$

Система (3.32) доповнюється умовою нормування

$$p_0 + p_1 + p_2 + \dots + p_{n+m-1} + p_{n+m} = 1. \tag{3.33}$$

Поділимо кожне рівняння на $n\mu$ і введемо позначення $\rho = \lambda / n\mu$ – завантаженість всієї системи. Система (3.32) переписеться так:

$$\begin{aligned}
 & -\rho p_0 + p_1 / n = 0, \\
 & -(\rho + i/n)p_i + \rho p_{i-1} + (i+1)/n \cdot p_i = 0, \quad 1 \leq i < n, \\
 & -(\rho + 1)p_i + \rho p_{i-1} + p_i = 0, \quad n \leq i < n + m, \\
 & -p_{n+m} + \rho p_{n+m-1} = 0.
 \end{aligned} \tag{3.34}$$

З першого рівняння системи (3.34) випливає, що: $p_1 = n\rho p_0$. Підставимо цей вираз у друге рівняння системи (2.34). Отримаємо: $p_2 = (n^2 / 2!) \rho^2 p_0$. Продовживши цю операцію, отримаємо:

$$\begin{aligned}
 p_i &= (n^i / i!) \rho^i p_0, \quad 1 \leq i \leq n, \\
 p_i &= (n^n / n!) \rho^i p_0, \quad n < i \leq n + m.
 \end{aligned} \tag{3.35}$$

Підставивши p_i в умови нормування (3.33), знайдемо вираз для імовірності простоювання системи:

$$p_0 = \left[1 + \sum_{i=1}^{n-1} \frac{n^i \rho^i}{i!} + \sum_{i=n}^{n+m} \frac{n^i \rho^i}{n!} \right]^{-1} = \left[\sum_{i=0}^{n-1} \frac{n^i \rho^i}{i!} + \frac{n^n \rho^n (1 - \rho^{m+1})}{n! (1 - \rho)} \right]^{-1}. \quad (3.36)$$

Отже, усі імовірності станів визначено.

Визначимо базові показники цієї системи.

Імовірність відмови щодо обслуговування –

$$P_{vidm} = p_{m+n} = (n^n / n!) \rho^{n+m} p_0. \quad (3.37)$$

Цей вираз означає, що запит отримає відмову щодо обслуговування, якщо всі лінії й місця очікування будуть зайняті.

Кількість запитів, які отримають відмову щодо обслуговування за час T – N_{vidm} , можна обчислити за виразом:

$$N_{vidm} = \lambda p_{m+n} T = \lambda (n^n / n!) \rho^{n+m} p_0 T. \quad (3.38)$$

Імовірність того, що запит буде обслужено системою, –

$$P_{obsl} = 1 - P_{vidm} = 1 - (n^n / n!) \rho^{n+m} p_0. \quad (3.39)$$

Окрім того, це відносна пропускна здатність системи, яка дорівнює середній частці запитів, які надійшли в систему й обслуговується нею.

Абсолютна пропускна здатність A дорівнює середній кількості запитів, які може обслужити СМО за одиницю часу, тобто це частка вхідного потоку заявок, яка надходить у систему й обслуговується нею:

$$A = \lambda P_{obsl} = \lambda (1 - (n^n / n!) \rho^{n+m} p_0). \quad (3.40)$$

Визначимо середню кількість запитів, що очікують на обслуговування – L_q . Якщо СМО перебуває в стані $\{n+1\}$, то в черзі буде один запит, імовірність цього – p_{n+1} . Якщо СМО буде перебувати в стані $\{n+m\}$, то в черзі буде m запитів, імовірність цього – p_{n+m} . Середня кількість запитів у черзі на обслуговування дорівнює математичному сподіванню кількості запитів у черзі, її можна обчислити за формулою:

$$L_q = \sum_{i=1}^m i \cdot p_{n+i} = \sum_{i=1}^m i \cdot (n^n / n!) \cdot \rho^{n+i} p_0 = (n^n / n!) \rho^{n+1} (1 + 2\rho + \dots + m\rho^{m-1}) p_0. \quad (3.41)$$

Якщо $\rho = 1$, то вираз у дужках перетворюється на арифметичну прогресію $1 + 2 + 3 + \dots + m$. Її сума дорівнює $m(m+1)/2$. Тоді

$$L_q = 0,5(n^n / n!)m(m+1)p_0. \quad (3.42)$$

Нехай $\rho \neq 1$. Спростимо вираз у дужках за виразом (3.41):

$$1 + 2\rho + 3\rho^2 + \dots + m\rho^{m-1} = \frac{d}{d\rho} (\rho + \rho^2 + \rho^3 + \dots + \rho^m).$$

Вираз у дужках – це кінцева геометрична прогресія зі знаменником ρ .
Сума її членів

$$\rho + \rho^2 + \rho^3 + \dots + \rho^m = \frac{\rho(1 - \rho^m)}{1 - \rho}. \quad (3.43)$$

Обчислимо похідну від виразу (3.43):

$$\frac{d}{d\rho} \left(\frac{\rho(1 - \rho^m)}{1 - \rho} \right) = \frac{1 - \rho^m(1 + m - m\rho)}{(1 - \rho)^2}. \quad (3.44)$$

Підставивши (3.44) у (3.41), можна знайти L_q при $\rho \neq 1$. Остаточоно :

$$L_q = \begin{cases} \frac{n^n}{n!} \frac{1 - \rho^m(1 + m - m\rho)}{(1 - \rho)^2} \rho^{n+1} p_0, & \rho \neq 1, \\ 0,5(n^n / n!)m(m + 1)p_0, & \rho = 1. \end{cases} \quad (3.45)$$

На рисунку 3.9 зображено графік залежностей середньої кількості запитів, що очікують на обслуговування, від навантаження всієї СМО ρ при різних значеннях числа каналів n і однаковій кількості місць очікування $m=7$.

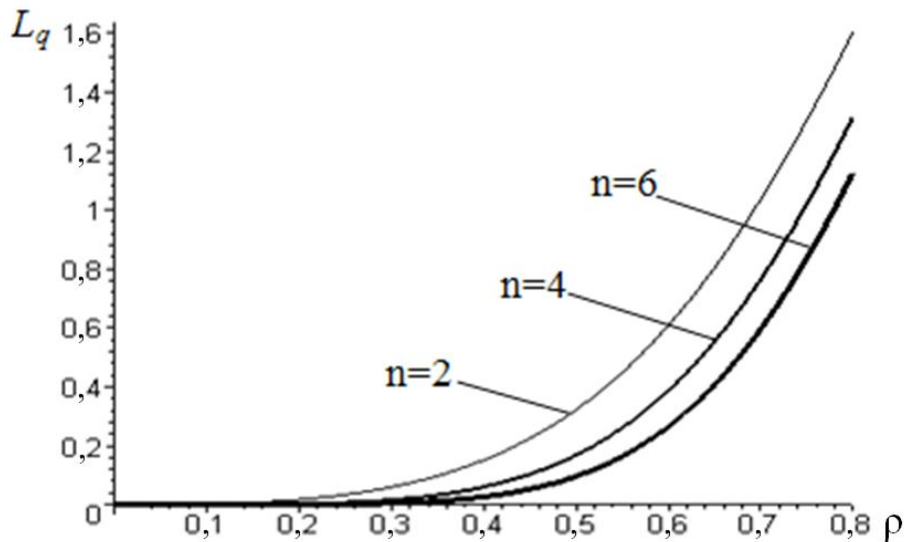


Рисунок 3.9 – Залежність L_q від завантаженості системи ρ при різних n

Зрозуміло, що при $\rho \geq 0,5$ довжина черги різко збільшується; при зростанні n і однаковому значенні ρ L_q теж різко збільшується, а потім збільшення сповільнюється.

На рисунку 3.10 зображено графік залежностей середньої кількості запитів, що очікують на обслуговування, від навантаження всієї СМО ρ при різних значеннях кількості місць очікування каналів n і однаковій кількості каналів $n = 5$.

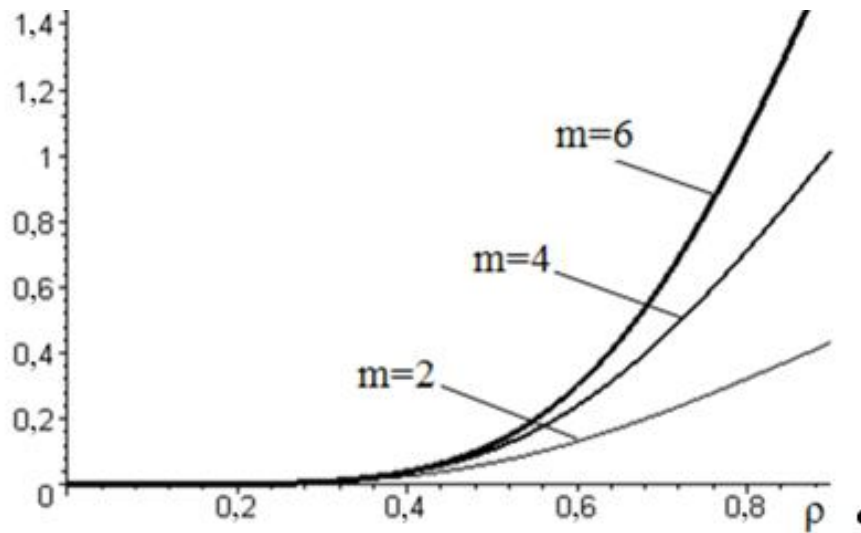


Рисунок 3.10 – Залежність L_q від завантаження системи ρ при різних m

Зрозуміло, що кількість місць очікування істотно впливає на довжину черги в системі. Зі зростанням m L_q також зростає. Це пояснюється тим, що менша кількість запитів отримує відмову щодо обслуговування. Окрім того, як і в попередньому випадку, при $\rho \geq 0,5$ довжина черги різко збільшується.

Визначимо середню кількість зайнятих обслуговуванням каналів – \bar{n} , а відповідно, і середню кількість запитів, що обслуговуються. Кожен зайнятий канал обслуговує в середньому μ запитів за одиницю часу, а СМО загалом – A запитів. Поділивши A на μ , отримаємо \bar{n} :

$$\bar{n} = \frac{A}{\mu} = \frac{\lambda(1 - (n^n/n!) \cdot \rho^{n+m} p_0)}{\mu} = n\rho(1 - (n^n/n!) \rho^{n+m} p_0). \quad (3.46)$$

Звідси впливає важливий показник – коефіцієнт завантаженості каналів системи - z_c :

$$z_c = \bar{n} / n = \rho(1 - (n^n/n!) \rho^{n+m} p_0). \quad (3.47)$$

Середня кількість запитів щодо обслуговування та в черзі (у системі):

$$L_s = \bar{n}_z + L_q = n\rho(1 - (n^n/n!) \cdot \rho^{n+m} p_0) + \frac{n^n}{n!} \frac{1 - \rho^m(1 + m - m\rho)}{(1 - \rho)^2} \rho^{n+1} p_0, \rho \neq 1. \quad (3.48)$$

Середній час очікування запиту в черзі τ_q знайдемо за формулою Літтла (2.39) – $L_q = \lambda \tau_q$:

$$\tau_q = \frac{L_q}{\lambda} = \frac{L_q}{n\mu\rho}. \quad (3.49)$$

Підставивши (3.45) у (4.39), отримаємо:

$$\tau_q = \begin{cases} \frac{n^n}{\lambda n!} \cdot \frac{1 - \rho^m (1 + m - m\rho)}{(1 - \rho)^2} \rho^{n+1} p_0, & \rho \neq 1, \\ 0,5(n^n / \lambda n!)m(m+1)p_0, & \rho = 1. \end{cases} \quad (3.50)$$

або

$$\tau_q = \begin{cases} \frac{n^{n-1}}{\mu n!} \cdot \frac{1 - \rho^m (1 + m - m\rho)}{(1 - \rho)^2} \rho^n p_0, & \rho \neq 1, \\ 0,5(n^{n-1} / \mu n!)m(m+1)p_0, & \rho = 1. \end{cases} \quad (3.51)$$

Середній час перебування запиту в системі τ_s для $\rho \neq 1$ також знайдемо за формулою Літтла (2.42) – $L_s = \lambda \tau_s$:

$$\tau_s = \frac{L_s}{\lambda} = \frac{L_s}{n\mu\rho}. \quad (3.52)$$

Підставивши (3.48) у (3.52), отримаємо:

$$\tau_s = (1 - (n^n / n!) \cdot \rho^{n+m} p_0) / \mu + \frac{n^n}{n!} \frac{1 - \rho^m (1 + m - m\rho)}{\lambda(1 - \rho)^2} \rho^{n+1} p_0 \quad (3.53)$$

або

$$\tau_s = (1 - (n^n / n!) \cdot \rho^{n+m} p_0) / \mu + \frac{n^{n-1}}{n!} \frac{1 - \rho^m (1 + m - m\rho)}{\mu(1 - \rho)^2} \rho^n p_0. \quad (3.54)$$

Якщо СМО М/М/п/п моделює економічну систему, у якій обслуговуються запити, то для її оптимізації можна використати критерій – функцію вартості збитків у системі за час $T - G$:

$$E = (C_{prost} (1 - z_c)n + C_q L_q + \lambda C_{zb} P_{vidm} + C_e z_c n) T, \quad (3.55)$$

де C_{prost} – вартість одиниці простоювання каналу; C_q – вартість збитків від простоювання запиту в черзі за одиницю часу; C_{zb} – вартість збитків від виходу з системи запиту, який не був обслужений; C_e – вартість експлуатації кожного каналу системи за одиницю часу.

Як критерій можна також використовувати прибуток від функціонування СМО – G :

$$G = \lambda C_s P_{obsl} T - (C_{prost} (1 - z_c)n + C_q L_q + \lambda C_{zb} P_{vidm} + C_e z_c n) T, \quad (3.56)$$

де C_s – вартість обслуговування кожного запиту, тобто це валовий прибуток, отриманий під час обслуговування кожного запиту.

Оптимізацію можна проводити або за кількістю каналів обслуговування n , або за інтенсивністю обслуговування μ за умови, що цей параметр можна змінювати. Також можна призначати вартість обслуговування кожного запиту

C_s , щоб прибуток G був не менше запланованого. Оптимізацію доцільно здійснювати чисельними методами [18].

3.4 Одноканальна система масового обслуговування з довільними потоками подій

У нотації Кендала одноканальна система позначається як G/G/1, тобто це система масового обслуговування з довільним вхідним потоком, який описується функцією розподілу $A(t)$ і довільно розподіленою функцією обслуговування – $B(t)$. Детально цю систему досліджено в [25].

Нехай $\bar{\lambda} = \left[\int_0^{\infty} t dA(t) \right]^{-1}$ – середня інтенсивність вхідного потоку, а $\bar{\mu} = \left[\int_0^{\infty} t dB(t) \right]^{-1}$ – середня інтенсивність потоку обслуговування. Якщо $\bar{\mu} > \bar{\lambda}$, то для сталого режиму цієї системи існує функція розподілу часу очікування $W(t)$, для визначення якої можна скласти інтегральне рівняння Ліндлі [25, с. 124].

$$W(t) = \begin{cases} \int_0^t W(t-x) dC(x), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (3.57)$$

де $C(t) = \int_0^{\infty} B(x+t) dA(t)$.

Розв'язок можна знайти за допомогою факторизації перетворення Лапласа – Стілтєса ядра інтегрального рівняння. Відповідне факторизаційне рівняння буде мати такий вигляд:

$$\alpha(-s)\beta(s) - 1 = 0, \quad (3.58)$$

де $\alpha(s)$ і $\beta(s)$ – перетворення Лапласа – Стілтєса від $A(t)$ і $B(t)$ відповідно. Використаємо теорему 4 з [25, с. 140]: якщо перетворення Лапласа – Стілтєса від $B(t)$ – $\beta(s)$ є відношенням двох багаточленів $\beta(s) = P_m / Q_n$, до того ж $n > m$ і $\exp(\lambda t)[1 - A(t)]$ має обмежену варіацію у $(0, \infty)$, то перетворення Лапласа – Стілтєса від $W(t)$ – $\varphi(s)$ буде виглядати так:

$$\varphi(s) = \frac{Q_n(s)}{Q_n(0) \prod_{i=1}^n \left(1 - \frac{s}{q_i} \right)} \quad (3.59)$$

де q_i – нулі функції $1 - \alpha(-s)\beta(s)$ у лівій напівплощині, де $\operatorname{Re} s < 0$. Обравши обернене перетворення Лапласа – Стілтєса від $\varphi(s)$, можна отримати вираз для функції розподілу часу очікування $W(t)$ у черзі.

Приклад 10. Дослідимо систему з гіперекспоненціальним вхідним потоком і ерлангівським розподілом часу обслуговування, тобто

$$A(t) = 1 - \sum_{i=1}^m a_i \exp(-\lambda_i t), \quad B(t) = 1 - \sum_{i=1}^n \frac{(\mu t)^{i-1}}{(i-1)!} e^{-\mu t}. \quad \text{Відповідно} \quad \bar{\lambda} = \left(\sum_{i=1}^m \frac{a_i}{\lambda_i} \right)^{-1},$$

$\bar{\mu} = \frac{\mu}{n}$. Перетворення Лапласа – Стілтєса від $A(t)$ і $B(t)$ має такий вигляд:

$$\alpha(s) = \sum_{i=1}^m \frac{a_i \lambda_i}{s + \lambda_i}, \quad \beta(s) = \left(\frac{\mu}{s + \mu} \right)^n = \frac{\mu^n}{(s + \mu)^n} = \frac{Q_0(s)}{Q_n(s)}.$$

Тоді

$$\varphi(s) = \frac{(\mu + s)^n}{\mu^n \prod_{i=1}^n \left(1 - \frac{s}{q_i} \right)}, \quad (3.60)$$

де q_i – корені рівняння в лівій напівплощині

$$\left(\sum_{i=1}^m \frac{a_i \lambda_i}{\lambda_i - s} \right) \cdot \left(\frac{\mu}{\mu + s} \right)^n - 1. \quad (3.61)$$

Оберемо $s = \mu(z - 1)$, тоді рівняння (3.61) буде таким:

$$\sum_{i=1}^m \frac{a_i \lambda_i}{\lambda_i + \mu(1 - z)} - z^n = 0. \quad (3.62)$$

Рівняння (3.62) має $m + n$ нулів, n нулів цього рівняння містяться в колі $|z| = 1$, беручі до уваги їхню кратність. Для доведення цього факту знайдемо значення модулів обох додатків (3.62) на границі області $|z| = 1 - \varepsilon$, $\varepsilon \rightarrow +0$,

тоді
$$|-z|^n = (1 - \varepsilon)^n. \quad (3.63)$$

Якщо розкласти праву частину (3.63) за формулою Тейлора й відкинути члени найвищого порядку малості щодо першого, то отримаємо: $|-z|^n = 1 - n\varepsilon$

Аналогічно:

$$\left| \sum_{i=1}^m \frac{a_i \lambda_i}{\lambda_i + \mu(1 - z)} \right| = 1 - \mu\varepsilon \sum_{i=1}^m \frac{a_i}{\lambda_i},$$

Оскільки $\bar{\mu} > \bar{\lambda}$, то $\varepsilon n < \varepsilon \mu \sum_{i=1}^m \frac{a_i}{\lambda_i} \Rightarrow n < \mu \sum_{i=1}^m \frac{a_i}{\lambda_i}$, тобто на границі області

$|z|=1-\varepsilon, \varepsilon \rightarrow +0$ $|-z^n| > \left| \sum_{i=1}^m \frac{a_i \lambda_i}{\lambda_i + \mu(1-z)} \right|$. Згідно з теоремою Руше, усередині

області $|z|=1$, $|-z^n|$ і $\sum_{i=1}^m \frac{a_i \lambda_i}{\lambda_i + \mu(1-z)}$ мають однакову кількість нулів $-n$.

Позначимо їх як z_1, z_2, \dots, z_n . Використавши те, що $s = \mu(z-1)$, отримаємо $q_i = \mu(z_i - 1), i = 1, 2, \dots, n$. Остаточно:

$$\varphi(s) = (\mu + s)^n \prod_{i=1}^n \left(\frac{z_i - 1}{\mu(z_i - 1) - s} \right). \quad (3.64)$$

Застосувавши обернене перетворення Лапласа – Стільтєса від (3.64), отримаємо вираз для функції розподілу часу очікування в черзі:

$$W(t) = 1 - \frac{\sum_{i=1}^n \frac{\prod_{e=1}^n (z_e - 1) z_i^n}{\prod_{j=1}^n (z_j - z_i)} e^{-\mu(1-z_i)t}}{\sum_{e=1}^n \frac{\prod_{j=1}^n (z_e - z_j)}{(z_j - z_i)}}. \quad (3.65)$$

Застосовуючи отримані вирази, можна знайти базові показники досліджуваної системи масового обслуговування.

Імовірність простоювання –

$$P_{prost} = p_0 = W(0) = \lim_{s \rightarrow \infty} \varphi(s) = \prod_{i=1}^n (1 - z_i).$$

Середній час очікування у черзі

$$\tau_q = -\varphi'(s)|_{s=0} = \sum_{i=1}^n \left[\frac{1}{\mu(1-z_i)} - \frac{n}{\mu} \right].$$

Середню кількість запитів у черзі визначимо за формулою Літтла (2.39)

$$L_q = \left(\sum_{i=1}^m \frac{a_i}{\lambda_i} \right)^{-1} \sum_{i=1}^n \left[\frac{1}{\mu(1-z_i)} - \frac{n}{\mu} \right].$$

За цими виразами можна оцінити основні характеристики функціонування системи в залежності від її параметрів.

3.5 Система масового обслуговування з кінцевою кількістю джерел запитів

Значна кількість реальних технічних систем, які моделюються за допомогою СМО, має кінцеву кількість джерел запитів. Джерело запитів не

надсилає нового запиту щодо обслуговування доти не буде обслужений попередній запит. Прикладом є система вводу–виводу, пристрої якої періодично надсилають запити для передачі масивів інформації. Імовірнісними моделями таких систем є системи масового обслуговування з кінцевим числом джерел запитів, які ще називаються замкнутими системами.

Дослідимо таку систему за таких передумов:

- система – марківська, кількість джерел запитів – k , інтенсивність ініціювання запитів від одного джерела – λ , джерело не надсилає нового запиту щодо обслуговування допоки не буде обслужено попередній запит ;
- система однолінійна, інтенсивність обслуговування запитів – μ ;
- якщо канал обслуговування зайнятий, то ініційований запит чекає на обслуговування або в черзі або в самому джерелі запитів.

Позначимо через $S_i, i = 0, 1, \dots, k$ стан системи, коли ініційовані i запитів, до того ж один обслуговується, $i - 1$ очікують на обслуговування, а $k - i$ джерел не ініціювали запити на обслуговування. Імовірністю стану S_i буде $p_i(t)$.

На рисунку 3.11 зображено граф станів та переходів такої СМО, коли кількість джерел запитів замкнутої системи дорівнює п'яти.

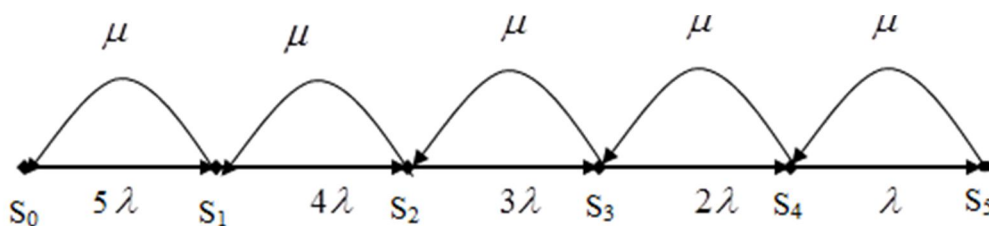


Рисунок 3.11 – Граф станів і переходів замкнутої СМО

Виведемо систему диференціальних рівнянь для ймовірностей станів. Інтенсивність ініціювання запитів від одного джерела – λ , а від j джерел – $j\lambda$. Якщо система перебуває в стані $i, i = 0, 1, \dots, k$, то сумарна інтенсивність надходження запитів – $(k - i)\lambda$.

Використавши методику складання диференціальних рівнянь для ймовірностей станів із підрозділу 1.2.3; отримаємо таку систему лінійних диференціальних рівнянь:

$$\begin{aligned}
 p_0(t) &= -k\lambda p_0(t) + \mu p_1(t), \\
 p_1(t) &= -[(k - 1)\lambda + \mu]p_0(t) + k\lambda p_0(t) + \mu p_2(t), \\
 p_i(t) &= -[(k - i)\lambda + \mu]p_i(t) + (k - i + 1)\lambda p_{i-1}(t) + \mu p_{i+1}(t), 0 < i < k \\
 p_k(t) &= -\mu p_k(t) + \lambda p_{k-1}(t).
 \end{aligned}
 \tag{3.66}$$

Початкові умови – $p_0(0) = 1, p_i(0) = 0, 0 < i \leq k$.

Для розв'язання системи (3.66) оберемо чисельний метод Ейлера розв'язання систем диференціальних рівнянь [24, с. 445]. Позначимо через p_i^l значення імовірності $p_i(t)$ на l -му кроці обчислень. Система (3.66) буде записана у вигляді, готовому для ітераційного обчислення

$$\begin{aligned} p_0^{l+1} &= p_0^l + h \cdot (-k\lambda p_0^l + \mu p_1^l), l = 0, 1, 2, \dots, \\ p_i^{l+1} &= p_i^l + h \cdot [-((k-i)\lambda + \mu)p_i^l + (k-i+1)\lambda p_{i-1}^l + \mu p_{i+1}^l], 0 < i < k, l = 0, 1, 2, \dots, \\ p_k^{l+1} &= p_k^l + h \cdot (-\mu p_k^l + \lambda p_{k-1}^l), l = 0, 1, 2, \dots, \\ p_0^0 &= 1, p_i^0 = 0, i \geq 1, \end{aligned} \quad (3.67)$$

де h – крок інтегрування.

Для чисельного розв'язання системи диференціальних рівнянь (3.66) за формулами (3.67) потрібно обрати конкретні значення параметрів λ, μ і h . Обирати крок інтегрування h доведеться експериментально, до того ж розв'язок буде наближеним.

На рисунку 3.12 зображено графіки ймовірностей стану досліджуваної системи для перехідного режиму при $\lambda = 0,2; \mu = 1; k = 5; h = 0,1$.

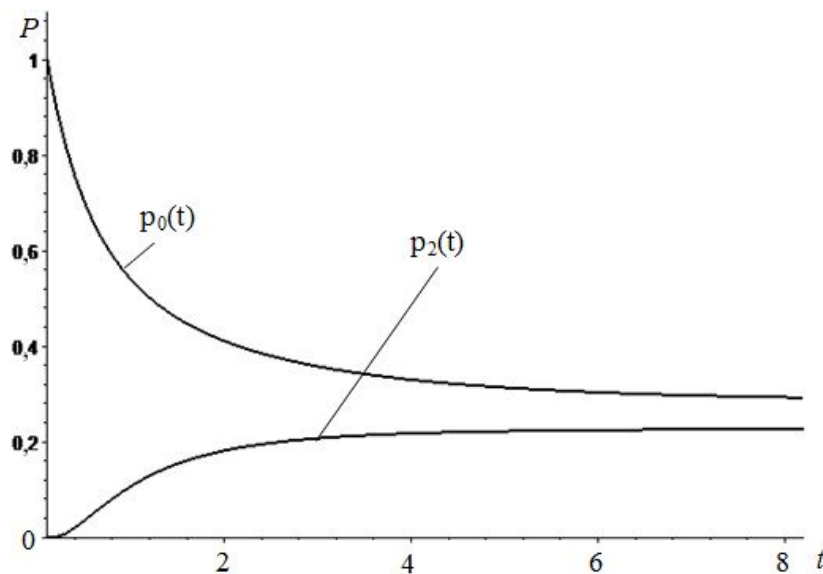


Рисунок 3.12 – Перехідний режим СМО з кінцевим числом джерел запитів

Зрозуміло, що з моменту часу $t = 5$ система переходить в стаціонарний режим функціонування.

Дослідимо стаціонарний режим функціонування СМО з кінцевою кількістю джерел запитів. Для стаціонарного режиму $p_i(t) \rightarrow p_i, p_i' = 0$ система (3.66) набуде такого вигляду:

$$\begin{aligned}
& -k\lambda p_0 + \mu p_1 = 0, \\
& -[-(k-1)\lambda + \mu]p_1 + k\lambda p_0 + \mu p_2 = 0, \\
& -[-(k-2)\lambda + \mu]p_2 + (k-1)\lambda p_1 + \mu p_3 = 0, \\
& -[-(k-i)\lambda + \mu]p_i + (k-i+1)\lambda p_{i-1} + \mu p_{i+1} = 0, \quad 0 < i < k, \\
& -\mu p_k + \lambda p_{k-1} = 0.
\end{aligned} \tag{3.68}$$

Умова нормування –

$$\sum_{i=0}^k p_i = 1. \tag{3.69}$$

Поділимо кожне рівняння системи (3.68) на μ і введемо показник $\rho = \lambda/\mu$ завантаження обслуговуючого каналу одним джерелом запитів. Система (3.68) буде такою:

$$\begin{aligned}
& -k\rho p_0 + p_1 = 0, \\
& -[-(k-1)\rho + 1]p_1 + k\rho p_0 + p_2 = 0, \\
& -[-(k-2)\rho + 1]p_2 + (k-1)\rho p_1 + p_3 = 0, \\
& -[-(k-i)\rho + 1]p_i + (k-i+1)\rho p_{i-1} + p_{i+1} = 0, \quad 0 < i < k, \\
& -p_k + \rho p_{k-1} = 0.
\end{aligned} \tag{3.70}$$

Послідовно розв'язуючи рівняння системи (3.70) відносно p_0 , отримаємо:

$$p_i = \frac{k!}{(k-i)!} \rho^i p_0. \tag{3.71}$$

Використавши умову нормування (3.69), знайдемо вираз для p_0 :

$$p_0 = \left[\sum_{i=0}^k \frac{k!}{(k-i)!} \rho^i \right]^{-1}. \tag{3.72}$$

Звідси середня кількість запитів у системі буде такою: $L_s = \sum_{i=1}^k i \cdot p_i$, а середня кількість джерел запитів, які очікують на обслуговування, – $L_q = \sum_{i=1}^k (i-1) \cdot p_i$.

Після перетворень отримаємо:

$$L_q = \sum_{i=1}^k (i-1) \cdot p_i = \sum_{i=1}^k i p_i - \sum_{i=1}^k p_i = L_s - (1 - p_0). \tag{3.73}$$

Середня кількість джерел, які не надіслали запитів для обслуговування, буде $k - L_s$; частка часу, коли канал обслуговування зайнятий, дорівнює

$1 - p_0$. Числа $k - L_s$ і $1 - p_0$ пропорційні середньому значенню часу перебування запитів у відповідному стані, звідси

$$\frac{1/\lambda}{k - L_s} = \frac{\tau_q}{L_q} = \frac{1/\mu}{1 - p_0} = \frac{\tau_s}{L_s}. \quad (3.74)$$

Використовуючи вираз (3.74), можна подати середню кількість запитів

у системі як функцію p_0 . $\frac{1}{\mu} \frac{\lambda(1 - p_0)}{1} = k - L_s \Rightarrow L_s = k - \frac{1 - p_0}{\frac{\lambda}{\mu}}$. Остаточню:

$$L_s = k - \frac{1 - p_0}{\rho}. \quad (3.75)$$

З виразу (3.74) випливає: $\frac{1/\lambda}{k - L_s} = \frac{\tau_s}{L_s}$, звідси середній час перебування

запиту від джерела в системі – $\tau_s = \frac{L_s}{\lambda(k - L_s)} = \frac{k - \frac{1 - p_0}{\rho}}{\lambda \left(k - k + \frac{1 - p_0}{\rho} \right)} = \frac{k\rho - (1 - p_0)}{\lambda(1 - p_0)}$.

Остаточню:

$$\tau_s = \frac{1}{\lambda} \left(\frac{k\rho}{1 - p_0} - 1 \right). \quad (3.76)$$

З виразу (3.74) випливає, що середній час очікування запитом обслуговування дорівнює $\tau_q = L_q / [\lambda(k - L_s)]$. Використавши (3.73) і (3.75), отримаємо:

$$\tau_q = \frac{L_q}{\lambda(k - L_s)} = \frac{L_s - (1 - p_0)}{\lambda(k - L_s)} = \frac{k - \frac{1 - p_0}{\rho} - (1 - p_0)}{\lambda \left(k - k - \frac{1 - p_0}{\rho} \right)} = \frac{k\rho - (1 + \rho)(1 - p_0)}{\lambda(1 - p_0)}.$$

$$\text{Остаточню: } \tau_q = \frac{1}{\lambda} \left(\frac{k\rho}{1 - p_0} - (1 + \rho) \right).$$

На рисунку 3.13 зображено графік залежностей середнього часу очікування запитом обслуговування τ_q від інтенсивності надходження запитів від джерела при $\mu = 1$ і різних кількостях джерел запитів.

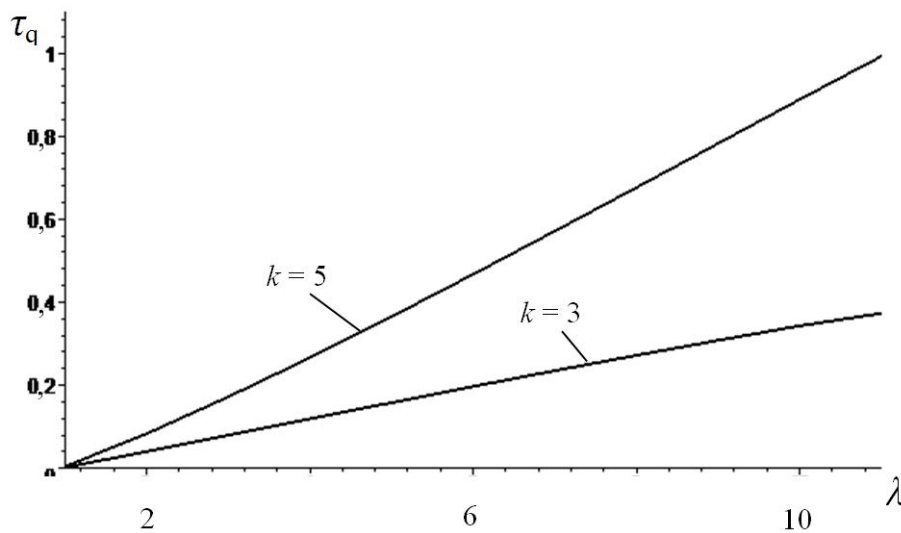


Рисунок 3.13 – Залежності часу очікування обслуговування від параметрів λ і k

Ці залежності, на відміну від розімкнутої системи, є лінійними.

3.6 Система масового обслуговування з абсолютним пріоритетом

У комп'ютерних системах, особливо таких, які працюють у реальному масштабі часу, передбачено переваги щодо обслуговування одних запитів стосовно інших – пріоритети. Сучасні операційні системи мають до 16 рівнів пріоритетів. Розрізняють відносні, абсолютні, динамічні, ситуаційні та змішані пріоритети [17, с. 63]. При відносних пріоритетах не дозволяється переривати обслуговування запитів; тільки після звільнення каналу з черги для обслуговування обирається запит із найвищим пріоритетом. При абсолютних пріоритетах запит, що надійшов нещодавно з певним пріоритетом, припиняє обслуговування запиту з меншим пріоритетом. При динамічних пріоритетах пріоритет окремих запитів у черзі може змінюватися залежно від змінювання деяких показників, наприклад часу очікування в черзі. Загальна теорія СМО з пріоритетами досить складна й виходить за межі курсу вищого навчального закладу.

Як приклад СМО з пріоритетами розглянемо одноканальну пуассонівську систему з абсолютним пріоритетом, до якої надходить два потоки вимог: безпріоритетний і пріоритетний. Пріоритетні запити обслуговуються в порядку надходження. Якщо пріоритетний запит надійшов у систему, коли вона обслуговує безпріоритетний запит, то це обслуговування припиняється, канал починає обслуговувати пріоритетний

запит. Безпріоритетний запит, обслуговування якого було перервано, обслуговується надалі тільки після звільнення каналу від обслуговування пріоритетних запитів. Інтенсивність пріоритетного потоку вимог – λ_1 , інтенсивність безпріоритетного потоку вимог – λ_2 . Інтенсивності обслуговування вхідних потоків, відповідно, μ_1 і μ_2 . Позначимо через S_{ij} стан системи, коли на обслуговування надійшло i пріоритетних і j безпріоритетних вимог. Якщо $j > 0$, то це означає, що на цей момент часу обслуговується пріоритетний запит. Граф станів і переходів досліджуваної системи з абсолютним пріоритетом зображено на рисунку 3.14.

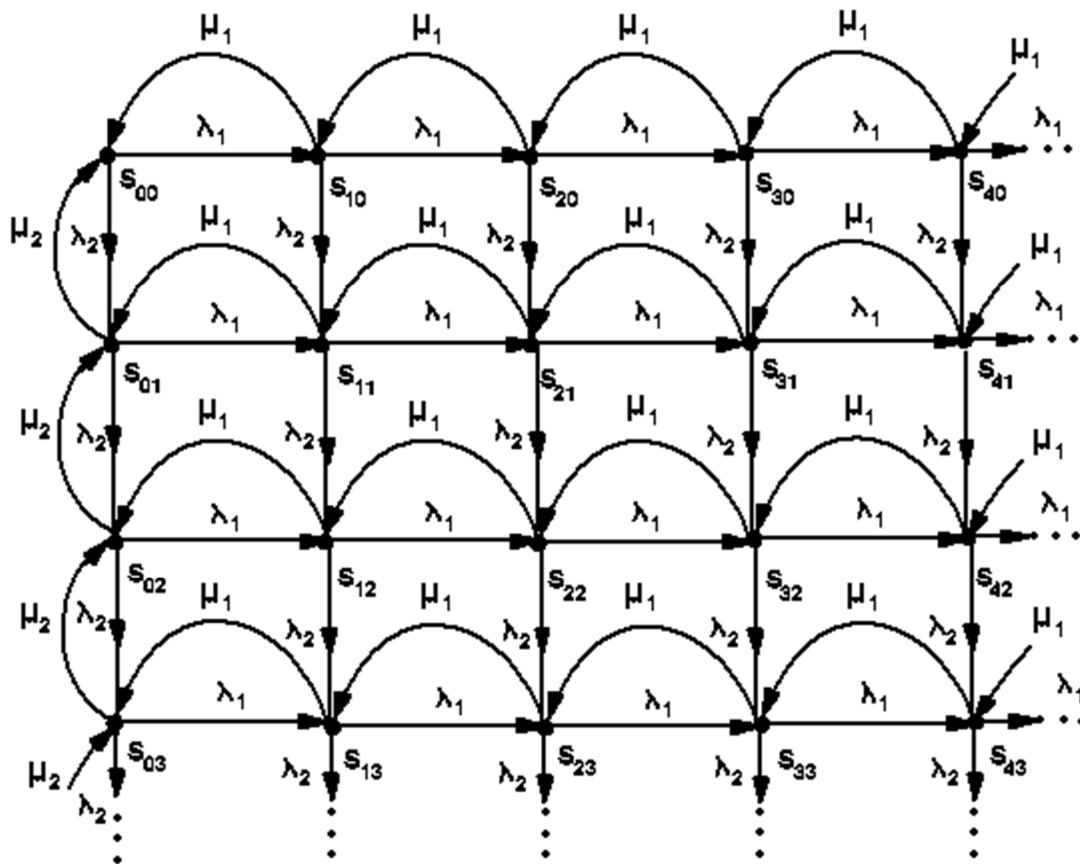


Рисунок 3.14 – Граф станів і переходів для СМО з двома вхідними потоками й абсолютним пріоритетом

Якщо кількість рівнів пріоритету буде більшою або дорівнюватиме двом, то граф станів і переходів набуде об'ємного вигляду й дослідити такі СМО за допомогою апарату ймовірностей стану буде неможливо. Із огляду на це, під час дослідження пріоритетних систем за допомогою аналітичних методів виникають певні труднощі.

Дослідження СМО з двома вхідними потоками й абсолютним пріоритетом продемонструємо на підставі положень з видання [20, с. 193].

Прийmemo, що $\rho_1 = \lambda_1 / \mu_1$, $\rho_2 = \lambda_2 / \mu$ – завантаження каналу обслуговування за кожним потоком запитів. Якщо $\rho = \rho_1 + \rho_2 < 1$, то існує стаціонарний режим функціонування системи. Позначимо через $p_{i,j}$ стаціонарну ймовірність стану S_{ij} . Використавши методикy п. 1.2.3, складемо систему лінійних рівнянь для ймовірностей $p_{i,j}, i = 0, 1, 2, \dots, j = 0, 1, 2, \dots$:

$$\begin{aligned} & -(\lambda_1 + \lambda_2)p_{0,0} + \mu_1 p_{1,0} + \mu_2 p_{0,1} = 0, \\ & -(\lambda_1 + \lambda_2 + \mu_2)p_{0,j} + \mu_1 p_{1,j} + \mu_2 p_{0,j+1} + \lambda_2 p_{0,j-1} = 0, j > 0, \\ & -(\lambda_1 + \lambda_2 + \mu_1)p_{i,0} + \lambda_1 p_{i-1,0} + \mu_1 p_{i+1,0} = 0, i > 0, \\ & -(\lambda_1 + \lambda_2 + \mu_1)p_{i,j} + \lambda_1 p_{i-1,j} + \lambda_2 p_{i,j-1} + \mu_1 p_{i+1,j}, i > 0, j > 0. \end{aligned} \quad (3.77)$$

Умова нормування –

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j} = 1. \quad (3.78)$$

Для розв'язання системи рівнянь (3.77) відповідно до умови нормування (3.78) введемо виробляючу функцію

$$G(u, z) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j} u^i z^j. \quad (3.79)$$

Для СМО з нескінченною кількістю місць для очікування ймовірність простоювання не залежить від дисципліни обслуговування й буде співпадати з імовірністю простоювання для системи М/М/1, тобто $p_0 = p_{0,0} = 1 - \rho = 1 - \rho_1 - \rho_2 = G(0,0)$.

При $u=1, z=1$ (3.79) перетворюється на умову нормування і $G(1,1)=1$. Помноживши кожне рівняння (3.77) на $u^i z^j$ і підсумувавши за всіма значеннями i, j , отримаємо:

$$\begin{aligned} & (\lambda_1 + \lambda_2 + \mu_1)G(u, z) + \sum_{j=0}^{\infty} (\mu_2 - \mu_1)p_{0,j}z^j - \mu_2 p_{0,0} = \\ & = \lambda_1 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i-1,j} u^i z^j + \mu_1 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i+1,j} u^i z^j + \lambda_2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j-1} u^i z^j + \mu_2 \sum_{j=0}^{\infty} p_{0,j+1} z^j. \end{aligned} \quad (3.80)$$

З визначення $G(u, z)$ випливає:

$$G(u, z) = \frac{[\mu_1 z(u-1) - \mu_2 u(z-1)]G(0, z) + \mu_2 u(z-1)(1-\rho)}{\lambda_1 u z(1-u) + \lambda_2 u z(1-z) + \mu_1 z(u-1)}; \quad (3.81)$$

Знаменник виразу є багаточленом другого ступеня відносно u і має корінь u_1 менший за нуль та корінь u_2 більший за нуль.

$$u_1 = \frac{\lambda_1 + \lambda_2(1-z) + \mu_1 - \sqrt{[\lambda_1 + \lambda_2(1-z) + \mu_1]^2 - 4\lambda_1\mu_1}}{2\lambda_1}. \quad (3.82)$$

Оскільки подвійний ряд (3.79) є збіжним при $u \leq 1, z \leq 1$, то u_1 має бути коренем і чисельника, звідки

$$G(0, z) = \frac{\mu_2(z-1)u_1(1-\rho)}{\mu_1 z - u_1[\mu_1 z - \mu_2(z-1)]}. \quad (3.83)$$

Підставивши $G(0, z)$ в (3.81), після перетворень отримаємо:

$$G(u, z) = \frac{(1-\rho)(1-\rho_1 u_1)}{(1-\rho_1 u_1 - \rho_2 z)(1-\rho_1 u_1)}. \quad (3.84)$$

Імовірності станів $p_{i,j}$ можна знайти за допомогою частинних похідних від $G(u, z)$:

$$p_{i,j} = \frac{1}{i!j!} \left[\frac{\partial^{i+j}}{\partial u^i \partial z^j} G(u, z) \right]_{u=z=0}. \quad (3.85)$$

Якщо розглядати тільки пріоритетні запити, то безпріоритетні не впливають на процес їх обслуговування, звідси характеристики обслуговування пріоритетних запитів співпадають із характеристиками СМО виду М/М/1 з параметрами $\lambda = \lambda_1, \mu = \mu_1$. Відповідно $\rho = \rho_1 = \lambda_1 / \mu_1$. Наприклад, середнє число пріоритетних запитів у системі – $L_{s1} = 1/(1-\rho_1)$, а середній час очікування в черзі – $\tau_{q1} = \rho_1 / [\mu_1(1-\rho_1)]$.

Позначимо через $q_i = \sum_{j=0}^{\infty} p_{i,j}$ безумовну ймовірність того, що в системі перебуває i пріоритетних запитів, а через $g_j = \sum_{i=0}^{\infty} p_{i,j}$ безумовну ймовірність того, що в системі перебуває j безпріоритетних запитів. Виробляюча функція для безумовних ймовірностей g_j визначиться, як

$$G(1, z) = \sum_{j=0}^{\infty} g_j z^j = \frac{1-\rho}{(1-\rho_1 u_1 - \rho_2 z)}. \quad (3.86)$$

За виразом (3.86) можна знайти середню кількість без пріоритетних запитів у системі:

$$L_{s2} = \sum_{j=0}^{\infty} j g_j = \left. \frac{dG(1, z)}{dz} \right|_{z=1}. \quad (3.87)$$

Обчислимо похідну від $G(1, z)$ за z :

$$\frac{dG(1, z)}{dz} = (1-\rho) \frac{\rho_1(u_1)'_z + \rho_2}{(1-\rho_1 u_1 - \rho_2 z)^2}. \quad (3.88)$$

Використаємо властивість похідних $y'_x = 1/x'_y$. Тоді вираз (3.88) набуде такого вигляду:

$$\frac{dG(1, z)}{dz} = (1 - \rho) \frac{\rho_1 \frac{1}{(z)'_u} + \rho_2}{(1 - \rho_1 u_1 - \rho_2 z)^2}. \quad (3.89)$$

Оскільки u_1 є коренем знаменника (3.81), то

$$\lambda_1 u_1 z (1 - u_1) + \lambda_2 u_1 z (1 - z) + \mu_1 z (u_1 - 1) = 0. \quad (3.90)$$

Звідси

$$z = \lambda / \lambda_2 - (\lambda_1 / \lambda_2) u_1 + 1 + \mu_1 / \lambda_2 - \mu_1 / (\lambda_2 u_1). \quad (3.91)$$

Знайдемо похідну від z за u_1 :

$$(z)'_{u_1} = \frac{\mu_1}{\lambda_2 u_1^2} - \frac{\lambda_1}{\lambda_2}. \quad (3.92)$$

Вираз (3.89) набуває такого вигляду:

$$\frac{dG(1, z)}{dz} = (1 - \rho) \frac{\rho_1 \frac{\lambda_2 u_1^2}{\mu_1 - \lambda_1 u_1^2} + \rho_2}{(1 - \rho_1 u_1 - \rho_2 z)^2}. \quad (3.93)$$

Зауважимо, що $u_1(1)=1$, звідси

$$L_{s_2} = \sum_{j=0}^{\infty} j g_j = \left. \frac{dG(1, z)}{dz} \right|_{z=1} = (1 - \rho) \left. \frac{\rho_1 \frac{\lambda_2 u_1^2}{\mu_1 - \lambda_1 u_1^2} + \rho_2}{(1 - \rho_1 u_1 - \rho_2 z)^2} \right|_{z=1}.$$

Остаточно:
$$L_{s_2} = \frac{\rho_2}{1 - \rho} \left(1 + \frac{\mu_2}{\mu_1} \cdot \frac{\rho_1}{1 - \rho_1} \right). \quad (3.94)$$

Використавши формулу Літтла (2.42), знайдемо середній час перебування в системі безпріоритетних запитів:

$$\tau_{s_2} = \frac{L_{s_2}}{\lambda_2} = \frac{1}{\mu_2 (1 - \rho)} \left(1 + \frac{\mu_2}{\mu_1} \cdot \frac{\rho_1}{1 - \rho_1} \right). \quad (3.95)$$

На рисунку 3.15 зображено графіки залежностей часу перебування низькопріоритетних запитів у системі від завантаженості ними каналу обслуговування при різних значеннях завантаженості каналу високопріоритетними запитами. Для порівняння подано графік залежностей часу перебування запитів у системі для безпріоритетної системи М/М/1 при тих самих параметрах.

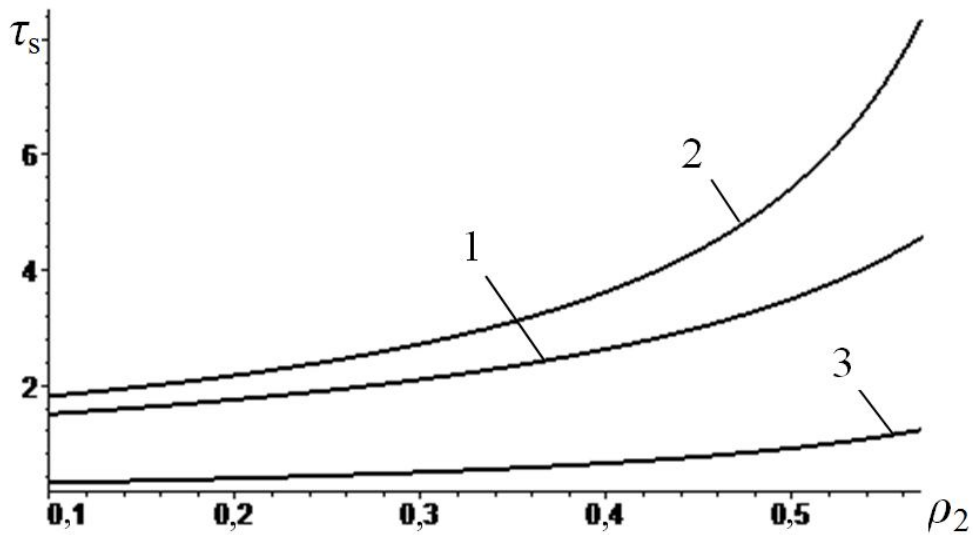


Рисунок 3.15 – Залежність часу знаходження в системі для:

- 1 – для безпріоритетних запитів при $\rho_1 = 0,2$;
- 2 – для безпріоритетних запитів при $\rho_1 = 0,3$;
- 3 – для довільних запитів системи М/М/1 при $\rho = 0,2 + \rho_2$

Зрозуміло, що введення пріоритетного потоку різко збільшує час перебування в системі безпріоритетних запитів порівняно з СМО виду М/М/1.

3.7 Мережі систем масового обслуговування

Сучасний етап розвитку комп'ютерної техніки вирізняється масовим упровадженням комп'ютерних мереж, у яких для розв'язання певного завдання можуть використовуватися різні комп'ютери мережі. Комп'ютерні мережі характеризуються складною топологією, тому досить часто виникає необхідність попередньо змоделювати майбутню мережу. Шляхом моделювання визначають характеристики мережі, необхідне мережне обладнання, оптимальну топологію, можливі резерви для майбутнього розвитку. Крім того, моделювання мережі забезпечує уникнення витрат, унаслідок додаткової реструктуризації мережі в майбутньому. Як аналітичні моделі комп'ютерних мереж використовують мережі систем масового обслуговування.

Загалом мережа СМО складається з декількох окремих СМО, виходи яких пов'язані із входами СМО мережі, як це зображено на рисунку 3.16.

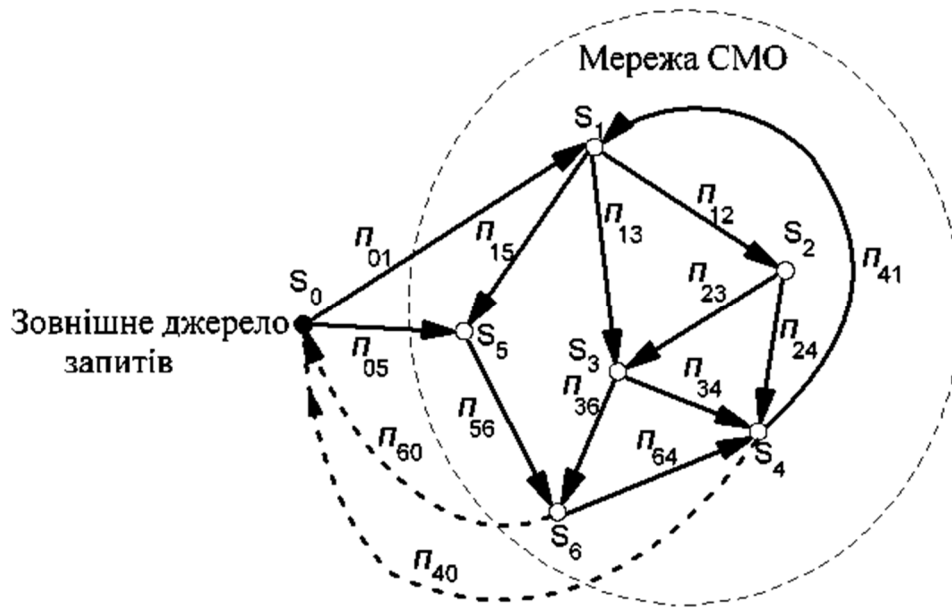


Рисунок 3.16 – Топологія мережі СМО

Мережа СМО складається з N окремих систем масового обслуговування S_i без втрат. Запити на обслуговування з імовірністю Π_{0i} надходять в i -ту СМО мережі ($i = 1, 2, \dots, N$) із зовнішнього джерела S_0 . Запит, який пройшов обслуговування на i -й СМО, з імовірністю Π_{ij} надходить для обслуговування в j -ту СМО або з імовірністю Π_{i0} залишає систему.

Розглянемо довільний запит, який після надходження в мережу послідовно проходить окремі СМО – $S_1, S_3, S_4, S_1, \dots$.

Послідовність $S_1, S_3, S_4, S_1, \dots$ утворює певну реалізацію ланцюга Маркова, яка описує функціонування мережі зі стохастичною матрицею:

$$[\Pi] = \begin{bmatrix} 0 & \Pi_{01} & \Pi_{02} & \dots & \Pi_{0N} \\ \Pi_{10} & \Pi_{11} & \Pi_{12} & \dots & \Pi_{1N} \\ \Pi_{20} & \Pi_{21} & \Pi_{22} & \dots & \Pi_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ \Pi_{N0} & \Pi_{N1} & \Pi_{N2} & \dots & \Pi_{NN} \end{bmatrix}, \quad (3.96)$$

де $0 \leq \Pi_{ij} \leq 1, i = \overline{1, N}, j = \overline{1, N}$ – імовірності переходу з i -ї СМО в j -ту. Матриця $[\Pi]$ називається матрицею передач мережі. Сума ймовірностей цієї матриці по строкам дорівнює одиниці:

$$\sum_{j=0}^N \Pi_{ij} = 1. \quad (3.97)$$

Для мережі за рисунком 3.16

$$[\Pi] = \begin{bmatrix} 0 & \Pi_{01} & 0 & 0 & 0 & \Pi_{05} & 0 \\ 0 & 0 & \Pi_{12} & \Pi_{13} & 0 & \Pi_{15} & 0 \\ 0 & 0 & 0 & \Pi_{23} & \Pi_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \Pi_{34} & 0 & \Pi_{36} \\ \Pi_{40} & \Pi_{41} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Pi_{56} \\ \Pi_{60} & 0 & 0 & 0 & \Pi_{64} & 0 & 0 \end{bmatrix},$$

ДО ТОГО Ж

$$\Pi_{01} + \Pi_{05} = 1, \Pi_{12} + \Pi_{13} + \Pi_{15} = 1, \Pi_{23} + \Pi_{34} = 1, \Pi_{40} + \Pi_{41} = 1, \Pi_{56} = 1, \Pi_{60} + \Pi_{64} = 1.$$

Досліджувати мережі СМО дуже складно. Зазвичай їх досліджують у стаціонарному режимі й при пуассонівських потоках подій.

Нехай λ_i – сумарна інтенсивність потоку запитів до i -ї СМО, тобто середня кількість запитів, які надходять у цю систему за одиницю часу в стаціонарному режимі, а μ_i – середня інтенсивність потоку обслуговування запитів в i -й СМО мережі. Інтенсивність потоку на виході з цієї системи також дорівнює λ_i (система без втрат). Імовірність Π_{ji} того, що запит, який був обслугований на j -й СМО, знаходить у систему i не залежить від попереднього шляху цього запиту й стану мережі. Отже, для стаціонарного режиму на підставі формули повної ймовірності (1.4) можна записати:

$$\lambda_i = \sum_{j=0}^N \lambda_j \Pi_{ji}, i = 0, 1, 2, \dots, N. \quad (3.98)$$

У виразі (3.98) $\lambda_0 = \sum_{j=1}^N \lambda_{0j}$ – сумарна інтенсивність джерела запитів.

Запишемо $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_N$ як матрицю-стовпець $[\Lambda] = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_N \end{bmatrix}$, тоді систему

(3.98) можна записати в матричному вигляді:

$$[\Lambda] = [\Pi]^T \cdot [\Lambda], \quad (3.99)$$

де $[\Pi]^T$ – матриця, транспонована до матриці $[\Pi]$.

З (3.99) знайдемо –

$$([\Pi]^T - I) \cdot [\Lambda] = 0, \quad (3.100)$$

де I – одинична матриця порядку $N+1$.

Вирази (3.98) і (3.100) описують систему $N+1$ лінійних однорідних рівнянь відносно інтенсивностей $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_N$, а вираз $(\|\Pi\|^T - I)$ є системною матрицею цієї системи лінійних рівнянь. Щоб ця система мала нетривіальний розв'язок $\lambda_i \neq 0$, необхідно, щоб визначник її системної матриці дорівнював нулю:

$$\left| \begin{bmatrix} -1 & \Pi_{10} & \Pi_{20} & \dots & \Pi_{N0} \\ \Pi_{01} & \Pi_{11} - 1 & \Pi_{21} & \dots & \Pi_{N1} \\ \Pi_{02} & \Pi_{12} & \Pi_{22} - 1 & \dots & \Pi_{N2} \\ \dots & \dots & \dots & \dots & \dots \\ \Pi_{0N} & \Pi_{1N} & \Pi_{2N} & \dots & \Pi_{NN} - 1 \end{bmatrix} \right| = 0. \quad (3.101)$$

Якщо до кожного елемента останнього рядка визначника (3.101) додати суму відповідних елементів інших стрічок, то цей рядок перетвориться на нуль:

$$\Pi_{Nj} + \Pi_{0j} + \Pi_{1j} + \dots + \Pi_{jj} - 1 + \dots + \Pi_{N-1j} = \Pi_{0j} + \Pi_{1j} + \dots + \Pi_{jj} + \dots + \Pi_{N-1j} + \Pi_{Nj} - 1 = 0.$$

Отже, визначник (3.101) буде дорівнювати нулю, а система (3.98) має безліч розв'язків, хоча всі $\lambda_i, i=1,2,\dots,N$ можна виразити через λ_0 : $\lambda_i = \alpha_i \lambda_0, i=1,2,\dots,N$, розв'язавши систему неоднорідних рівнянь:

$$\begin{aligned} \Pi_{01} \lambda_1 + \Pi_{02} \lambda_2 + \Pi_{03} \lambda_3 + \dots + \Pi_{0N} \lambda_N &= \lambda_0, \\ (\Pi_{11} - 1) \lambda_1 + \Pi_{12} \lambda_2 + \Pi_{13} \lambda_3 + \dots + \Pi_{1N} \lambda_N &= \Pi_{10} \lambda_0, \\ \Pi_{21} \lambda_1 + (\Pi_{22} - 1) \lambda_2 + \Pi_{23} \lambda_3 + \dots + \Pi_{2N} \lambda_N &= \Pi_{20} \lambda_0, \\ \dots & \\ \Pi_{N1} \lambda_1 + \Pi_{N2} \lambda_2 + \Pi_{N3} \lambda_3 + \dots + (\Pi_{NN} - 1) \lambda_N &= \Pi_{N0} \lambda_0. \end{aligned} \quad (3.102)$$

Розглянемо мережу, зображену на рисунку 3.17 [20, с. 122].

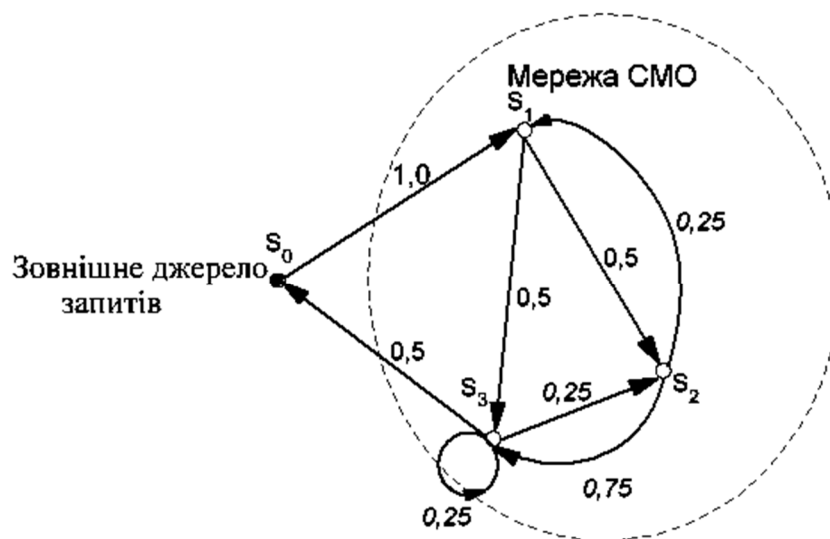


Рисунок 3.17 – Мережа СМО до прикладу.

Вона описується матрицею передач

$$[\Pi] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0,5 & 0,5 \\ 0 & 0,25 & 0 & 0,75 \\ 0,5 & 0 & 0,25 & 0,25 \end{bmatrix}.$$

Їй відповідає така системна матриця

$$[\Pi]^r - I = \begin{bmatrix} -1 & 0 & 0 & 0,5 \\ 1 & -1 & 0,25 & 0 \\ 0 & 0,5 & -1 & 0,25 \\ 0 & 0,5 & 0,75 & -0,75 \end{bmatrix},$$

якій відповідає система рівнянь вигляду (3.102):

$$\begin{aligned} 0,5\lambda_3 &= \lambda_0; \\ -\lambda_1 + 0,25\lambda_2 &= -\lambda_0; \\ 0,5\lambda_1 - \lambda_2 + 0,25\lambda_3 &= 0; \\ 0,5\lambda_1 + 0,75\lambda_2 - 0,75\lambda_3 &= 0. \end{aligned}$$

Запишемо її коефіцієнти при невідомих у матричному вигляді й виконаємо перетворення за методом Гауса [24, с.27]. Римськими цифрами позначимо номери рядків:

$$\begin{array}{ccc|c} 0 & 0 & 0,5 & 1 \\ -1 & 0,25 & 0 & -1 \\ 0,5 & -1 & 0,25 & 0 \\ 0,5 & 0,75 & -0,75 & 0 \end{array} \Rightarrow II = II + (III + IV) \Rightarrow \begin{array}{ccc|c} 0 & 0 & 0,5 & 1 \\ 0 & 0 & -0,5 & -1 \\ 0,5 & -1 & 0,25 & 0 \\ 0,5 & 0,75 & -0,75 & 0 \end{array} \Rightarrow$$

$$\Rightarrow \begin{array}{ccc|c} 0 & 0 & 0,5 & 1 \\ 0,5 & -1 & 0,25 & 0 \\ 0,5 & 0,75 & -0,75 & 0 \end{array} \Rightarrow \left\{ \begin{array}{l} II = II - 0,5I \\ III = III + 0,5I \end{array} \right\} = \begin{array}{ccc|c} 0 & 0 & 0,5 & 1 \\ 0,5 & -1 & 0 & -0,5 \\ 0,5 & 0,75 & 0 & 1,5 \end{array} \Rightarrow$$

$$\Rightarrow II = II - III \Rightarrow \begin{array}{ccc|c} 0 & 0 & 0,5 & 1 \\ 0 & -1,75 & 0 & -2 \\ 0,5 & 0,75 & 0 & 1,5 \end{array}$$

Остаточно:

$$\lambda_1 = \begin{vmatrix} -2 & -1,75 \\ 1,5 & 0,75 \\ 0 & -1,75 \\ 0,5 & 0,75 \end{vmatrix} \lambda_0 = 1,285\lambda_0; \lambda_2 = \frac{-2}{-1,75} \lambda_0 = 1,142\lambda_0; \lambda_3 = \frac{1}{0,5} \lambda_0 = 2\lambda_0$$

За цим прикладом зрозуміло, що система лінійних рівнянь (3.102), яка описує мережу СМО, має єдиний розв'язок відносно $\lambda_0 - \lambda_i = \alpha_i \lambda_0, i = 1, 2, \dots, N$.

Щоб у мережі СМО виник стаціонарний режим, потрібно, щоб кожна СМО, яка входить до мережі, функціонувала в стаціонарному режимі. Оскільки СМО, які входять у мережу, є системами без втрат, то для створення стаціонарного режиму необхідно забезпечити виконання умови, що $\lambda_i < \mu_i$ для кожної СМО. Її можна записати як $\alpha_i \lambda_0 < \mu$ або $\lambda_0 < \frac{\mu_i}{\alpha_i}, i = 1, 2, \dots, N$.

Отже, щоб у мережі СМО виник стаціонарний режим, необхідно забезпечити таку умову:

$$\lambda_0 < \min_{i=1,2,\dots,N} \frac{\mu_i}{\alpha_i}. \quad (3.103)$$

Під час розрахування характеристик мереж СМО доцільно знати час перебування запиту в мережі. Якщо відомо середній час перебування запиту в кожній СМО мережі (τ_{si}), то застосовують таку формулу [20, с. 127]:

$$T_m = \sum_{i=1}^N \alpha_i \tau_{si}. \quad (3.104)$$

Вираз (3.104) означає, що окремий запит може декілька разів обслуговуватися на одній і тій самій СМО мережі.

Контрольні питання

1. Яка система масового обслуговування називається СМО з відмовами?
2. Складіть систему рівнянь щодо СМО з відмовами для перехідного режиму, коли $n=3, \lambda=2, \mu=1, 2$.
3. Складіть систему рівнянь щодо СМО з відмовами для стаціонарного режиму, коли $n=4, \lambda=1,5, \mu=1,3$.
4. Обчисліть ймовірність відмови щодо обслуговування для СМО з відмовами, коли $n=3, \lambda=2, \mu=1,2$.
5. Що означає аббревіатура М/М/1?

6. У яких діапазонах для завантаження доцільно досліджувати СМО виду $M/M/1$?
7. Виведіть економічний критерій для СМО виду $M/M/1$.
8. Що означає аббревіатура $M/M/n/m$?
9. Зобразіть граф станів і переходів для системи $M/M/2/3$.
10. Виведіть систему лінійних рівнянь для ймовірностей станів системи виду $M/M/2/3$.
11. Назвіть базові показники для СМО виду $M/M/n/m$.
12. Виведіть функцію вартості збитків для СМО виду $M/M/n/m$.
13. Виведіть функцію вартості прибутків для СМО виду $M/M/n/m$.
14. Що означає аббревіатура $G/G/1$?
15. За допомогою якого методу можна досліджувати СМО виду $G/G/1$?
16. Назвіть базові показники для СМО виду $G/G/1$.
17. У чому полягає особливість систем масового обслуговування з кінцевою кількістю джерел запитів?
18. Назвіть базові показники для СМО з кінцевою кількістю джерел запитів.
19. У чому полягає особливість систем масового обслуговування з пріоритетами?
20. Назвіть базові види систем масового обслуговування з пріоритетами.
21. Які виникають труднощі під час дослідження систем масового обслуговування з пріоритетами?
22. Зобразіть граф станів і переходів для СМО з абсолютним пріоритетом, двома вхідними потоками й двома місцями для очікування.
23. Якими ознаками характеризуються мережі систем масового обслуговування?
24. За яких умов мережа систем масового обслуговування функціонує в стаціонарному режимі?

4 ЗАСТОСУВАННЯ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ ПІД ЧАС РОЗРОБЛЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ

Матеріал цього розділу базується на досвіді автора у розробці моделей масового обслуговування при проектуванні комп'ютерних систем різного призначення [7, 31, 32].

4.1 Дослідження й вибір параметрів модуля прийому числа-імпульсної інформації

У сучасних автоматичних і автоматизованих системах управління для організації раціонального сполучення управляючого обчислювального комплексу з об'єктом управління зазвичай використовують спеціальний процесор вводу-виводу. Однією з вимог, щодо процесора вводу-виводу є мінімум витрат машинного часу центрального процесора (ЦП) на обмін інформацією. У разі, якщо до системи управління входять датчики активної, число-імпульсної інформації, де кожен імпульс містить інформацію про певне збільшення вимірюваної величини за відповідним каналом зв'язку, до складу процесора вводу-виводу необхідно включати спеціальний модуль прийому число-імпульсної інформації (далі – МПЧІ). Цей модуль, що включає лічильники імпульсів за кожним каналом, дає змогу попередньо обробити інформацію, що надходить про стан об'єкта управління, а отже, зменшити навантаження на ЦП. Це особливо важливо при високих частотах надходження число-імпульсних сигналів.

Знімати інформацію з МПЧІ можна в двох режимах.

1. За командою ЦП, який зчитує дані з одного або декількох лічильників МПЧІ; до того ж після зчитування інформації лічильники скидаються.

2. За командою МПЧІ в разі заповнення лічильників на певну величину або їх переповнення; до того ж у процесор надходить сигнал переривання, за яким необхідно прийняти й обробити чергову порцію інформації. Переривають та скидають інформацію з лічильників за допомогою процесора одразу ж після надходження сигналу переривання. У разі, якщо сигнал переривання надходить після заповнення лічильника на певну величину, процесор на деякий час може припинити зняття інформації, зокрема якщо виконання поточної програми в цей момент не можна перервати.

Процес зняття інформації за командою центрального процесора зазвичай відбувається в циклі управління; процес зняття інформації за командою МПЧІ, тобто за запитами від лічильників, здійснюється

асинхронно щодо циклу управління, а тому бажано, щоб їх було як можна менше. Позначимо через F сумарну частоту переривань, що надходять від лічильників МПЧП. Тоді F можна обрати критерієм ефективності функціонування модуля прийому число-імпульсної інформації.

Дослідження функціонування МПЧП базується на такому:

а) сигнал переривання надходить в ЦП при заповненні i -го лічильника ($i = \overline{1, n}$) на q_i імпульсів; розрядність i -го лічильника – m_i , а його ємність – $W_i = 2^{m_i} - 1$ імпульсів. Імовірність того, що процесор одразу зніме інформацію з i -го лічильника в разі надходження від нього сигналу переривання a_i ;

б) потік імпульсів, що надходять по кожному каналу, випадковий і описується експоненціальною функцією розподілу з параметром f_i , де f_i – середня частота надходження імпульсів по i -му каналу;

в) процес зняття інформації за запитами процесора в циклі управління (до надходження сигналу переривання) випадковий, описується експоненціальною функцією розподілу з параметром μ_i , де μ_i – середня частота опитування i -го лічильника число-імпульсної інформації в циклі управління. Зняття інформації з лічильника після видачі ним сигналу переривання із затримкою на випадкове час ξ здійснюється з імовірністю $1 - a_i$ (ця ймовірність мала). Випадковий час ξ , що має середнє значення τ_i , підпорядкований експоненціальному розподілу з параметром $\nu_i = 1/\tau_i$.

Досліджуємо спочатку процес функціонування одного i -го лічильника прийому число-імпульсної інформації. Граф випадкового процесу станів і переходів для нього зображений на рисунку 4.1, де під записом $\{j\}$ ($j = \overline{0, w_i}$) розуміють стан лічильника, коли в ньому міститься j імпульсів, стаціонарна ймовірність цього – p_j ($j = \overline{0, W_i}$).

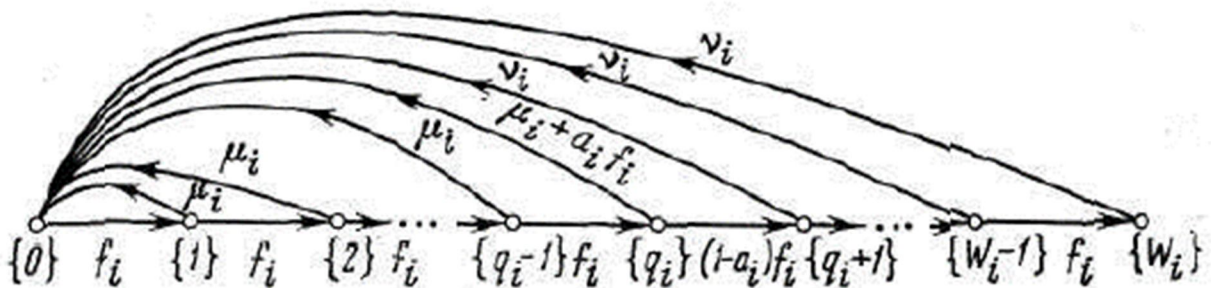


Рисунок 4.1 – Граф станів і переходів для i -го лічильника

Система стаціонарних рівнянь, що відповідають цьому графу, має такий вигляд:

$$\begin{aligned}
& -f_i p_0 + \mu_i \sum_{j=1}^{q_i-1} p_j + (\mu_i + a_i f_i) p_{q_i} + v_i \sum_{j=q_i+1}^{W_i} p_j = 0, \\
& -(f_i + \mu_i) p_j + f_i p_{j-1} = 0, \quad 0 < j \leq q_i, \\
& -(f_i + v_i) p_{q_i+1} + (1 - a_i) f_i p_{q_i} = 0, \\
& -(f_i + v_i) p_j + f_i p_{j-1} = 0, \quad q_i + 1 < j < W_i, \\
& -v_i p_{W_i} + f_i p_{W_i-1} = 0.
\end{aligned} \tag{4.1}$$

Умова нормування –

$$p_0 + \sum_{j=1}^{W_i} p_j = 1. \tag{4.2}$$

За другим рівнянням системи (4.1) послідовно знаходимо:

$$p_1 = \left(\frac{f_i}{f_i + \mu_i} \right) p_0, \quad p_j = \left(\frac{f_i}{f_i + \mu_i} \right)^j p_0, \quad j \leq q_i.$$

Підставивши p_{q_i} в третє рівняння системи (4.1), отримаємо:

$$p_{q_i+1} = \left(\frac{f_i}{f_i + \mu_i} \right)^{q_i} \frac{(1 - a_i)}{f_i + \mu_i} p_0.$$

За четвертим і п'ятим рівнянням системи (4.1) послідовно знаходимо:

$$p_j = \frac{(1 - a_i) f_i^j}{(f_i + \mu_i)^{q_i} (f_i + v_i)^{j - q_i}} p_0, \quad 1 + q_i < j < W_i,$$

$$p_{W_i} = \frac{(1 - a_i) f_i^{W_i}}{v_i (f_i + \mu_i)^{q_i} (f_i + v_i)^{W_i - q_i - 1}} p_0.$$

Підставивши ці ймовірності в умову нормування (4.2), після відповідних перетворень знаходимо ймовірність p_0 :

$$p_0 = \left\{ 1 + \frac{f_i}{\mu_i} \left[1 - \left(\frac{f_i}{f_i + \mu_i} \right)^{q_i} \right] + \frac{(1 - a_i) f_i^{q_i+1}}{(f_i + \mu_i)^{q_i} v_i} \left[1 - \left(\frac{f_i}{f_i + \mu_i} \right)^{W_i - q_i - 1} \right] + \frac{(1 - a_i) f_i^{W_i}}{v_i (f_i + \mu_i)^{q_i} (f_i + v_i)^{W_i - q_i - 1}} \right\}^{-1}.$$

Ймовірність втрати інформації p_{W_i} за i -м лічильником:

$$p_{W_i} = \frac{(1 - a_i) f_i^{W_i}}{v_i (f_i + \mu_i)^{q_i} (f_i + v_i)^{W_i - q_i - 1}}.$$

За останнім рівнянням можна обрати розрядність лічильника й частоту його опитування так, щоб ймовірність втрати інформації була в заданих межах.

Важливою характеристикою функціонування i -го лічильника є середній час T_i між видачею ним сигналу переривання в ЦП. Для визначення T_i дещо

змінимо граф (рис. 4.1), увівши новий стан $\{r_i\}$ і позначивши інтенсивність переходу зі стану $\{r_i\}$ в стан $\{0\}$ через η_i (рис. 4.2).

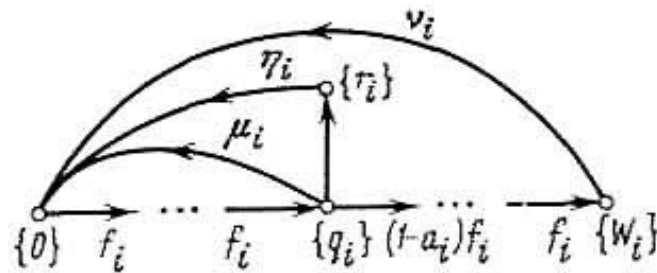


Рисунок 4.2 – Змінений граф станів системи

Граф на рисунку 4.2 буде еквівалентним до графу на рисунку 4.1, якщо $\eta_i = \infty$.

Стан $\{r_i\}$ може виникнути тільки по сигналу переривання лічильника. Позначимо через T_i^* середній час між перериваннями, беручи до уваги введений стан $\{r_i\}$. При $\eta_i = \infty$ маємо: $T_i^* \rightarrow T_i$. Час T_i буде складатися з середнього часу L_i – переміщення випадкового процесу по станах $\{0\}, \{1\}, \dots, \{q_i\}$, із середнього часу N_i – переміщення випадкового процесу по станах $\{q_i + 1\}, \{q_i + 2\}, \dots, \{W_i\}$ і середнього часу R_i – перебування випадкового процесу в стані $\{r_i\}$, тобто $T_i^* = L_i^* + N_i^* + R_i^*$. Для визначення L_i^* застосовуємо методику, викладену в [26, с. 74]. Для цього перетворимо граф, зображений на рисунку 4.2, у граф, зображений на рисунку 4.3, тобто виокремимо стани $\{r_i\}$ і $\{q_i + 1\}$ і зробимо їх замкнутими.

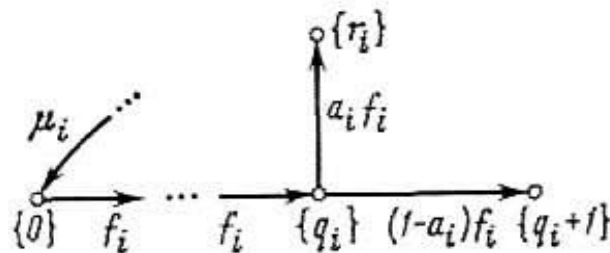


Рисунок 4.3 – Граф станів системи із замкнутими станами $\{r_i\}$ і $\{q_i + 1\}$

Межею станів $\{r_i\}$ і $\{q_i + 1\}$ є стан $\{0\}$. Початкові умови – $p_0(0) = 1, p_i(0) = 0, i \neq 0$.

Відповідно до [26, с. 74], щільність Імовірності $U(t)$ перебування в підмножині $[\{0\}, \{1\}, \dots, \{q_i\}]$ дорівнює $\frac{d}{dt} p_{r_i}(t) + \frac{d}{dt} p_{q_i+1}(t)$, а

$$\frac{d}{dt} p_{r_i}(t) + \frac{d}{dt} p_{q_i+1}(t) = f_i p_{q_i}(t), \text{ тобто } U(t) = f_i p_{q_i}(t).$$

Для визначення $p_{q_i}(t)$ необхідно скласти систему диференціальних рівнянь:

$$\begin{aligned} \frac{d}{dt} p_0(t) &= -f_i p_0(t) + \mu_i \sum_{i=1}^{q_i} p_i(t), \\ \frac{d}{dt} p_j(t) &= -(f_i + \mu_i) p_j(t) + f_i p_{j-1}(t), \quad j = 1, 2, \dots, q_i, \end{aligned} \quad (4.3)$$

із початковими умовами $p_0(0) = 1, p_i(0) = 0, i \neq 0$.

Знайдемо розв'язок системи (4.3) за перетвореннями Лапласа:

$$p_{q_i}(t) = \frac{(\mu_i + s) \left(\frac{f_i}{f_i + \mu_i + s} \right)^{q_i}}{s(f_i + \mu_i + s) + f_i \mu_i \left(\frac{f_i}{f_i + \mu_i + s} \right)^{q_i}}. \quad (4.4)$$

Застосувавши відомі відношення, отримаємо:

$$L_i^* = \int_0^{\infty} t U(t) dt = - \frac{d}{ds} U(s) \Big|_{s=0} = - f_i \frac{d}{ds} p_{q_i}(s) \Big|_{s=0} = \frac{(f_i + \mu_i)^{q_i+1} - f_i^{q_i+1}}{\mu_i f_i^{q_i+1}}. \quad (4.5)$$

Виконавши аналогічні дії, отримаємо:

$$N_i^* = \frac{1 - a_i}{v_i}, \quad R_i^* = \frac{a_i}{\eta_i}. \quad (4.6)$$

Таким чином,

$$T_i = \lim_{\eta_i \rightarrow 0} T_i^* = \frac{v_i \left[\left(1 + \frac{\mu_i}{\lambda_i} \right)^{q_i+1} - 1 \right] + \mu_i (1 - a_i)}{\mu_i v_i}. \quad (4.7)$$

Звідси середня частота надходження сигналів переривання від i -го лічильника –

$$F_i = \frac{\mu_i v_i}{v_i \left[\left(1 + \frac{\mu_i}{\lambda_i} \right)^{q_i+1} - 1 \right] + \mu_i (1 - a_i)}. \quad (4.8)$$

Сумарна частота надходження сигналів переривання від усього модуля прийому число-імпульсної інформації –

$$F = \sum_{i=1}^n \frac{\mu_i v_i}{v_i \left[\left(1 + \frac{\mu_i}{\lambda_i} \right)^{q_i+1} - 1 \right] + \mu_i (1 - a_i)}. \quad (4.9)$$

Використовуючи вираз (4.9), можна так вибрати частоти опитування лічильників прийому число-імпульсної інформації, щоб при одних і тих

самих витратах часу ЦП на обмін вплив на його роботу з боку МПЧП шляхом переривань був мінімальним. Для цього необхідно мінімізувати функцію F за змінними $\mu_i, i = \overline{1, n}$ за умови, що $\sum_{i=1}^n \mu_i = const$. Це завдання можна вирішити за допомогою відомих методів оптимізації, наприклад методу найшвидшого градієнтного спуску [27, с. 25].

У тих випадках, коли сигнал переривання надається після переповнення лічильників МПЧП, вираз (4.11) набуде такого вигляду:

$$F = \sum_{i=1}^n \frac{\mu_i}{\left(1 + \frac{\mu_i}{\lambda_i}\right)^{2^{m_i}} - 1}. \quad (4.10)$$

У цьому разі забезпечується вирішення вже двох завдань оптимізації:

1. За постійних розрядностей окремих лічильників обрати такі частоти опитування μ_i при $\sum_{i=1}^n \mu_i = const$, щоб сумарна частота переривань F була мінімальною.

2. За постійних частот опитування окремих лічильників обрати такі розрядності окремих лічильників m_i (тобто число тригерів та інших електронних схем), щоб при загальній незмінній кількості електронних схем ($\sum m_i = const$) сумарна частота переривань F була мінімальною.

Приклад. Нехай $n = 3$, $f_1 = 1\,000$ Гц, $f_2 = 5\,000$ Гц, $f_3 = 8\,000$ Гц. Розрядності окремих лічильників – $m_1 = m_2 = m_3 = 8$, $\sum \mu_i = 300$ Гц. Переривання здійснюється після переповнення окремих лічильників МПЧП. Необхідно оптимально (щодо мінімуму F) обрати частоти опитування окремих лічильників.

Використавши чисельний метод найшвидшого спуску, реалізований на ПЕОМ, отримаємо: $F_{\min} = 1,92$ Гц при $\mu_1 = 50,03$ Гц, $\mu_2 = 96,25$ Гц, $\mu_3 = 153,76$ Гц. Варто відзначити, що при $\mu_1 = \mu_2 = \mu_3 = 100$ Гц $F = 4,971$ Гц, тобто оптимальний вибір частот опитування уможливив зменшення сумарної частоти переривань більш ніж в два рази порівняно з рівномірним розподілом частот опитування.

4.2 Імовірнісне моделювання та оцінювання якості функціонування інформаційно-управляючих систем

Інформаційно-управляючі системи (далі – ІУС), які належать до систем реального часу, посідають важливе місце в сучасному виробництві й суспільному житті. Вони складні щодо реалізації, постійно вдосконалюються, за їх допомогою розв’язують усе більш складні завдання. Розроблення ІУС – досить складний процес, який примушує від проєктувальників приймати до уваги багато факторів. Для забезпечення високої якості ІУС на етапі проєктування потрібно використовувати моделювання, яке забезпечить перевірку базових проєктних рішень, уможливить виявлення вузьких місць у системі. Найважливішим складником ІУС є програмне забезпечення. Воно забезпечує розв’язання в реальному масштабі часу широкого кола завдань управління окремими підсистемами ІУС, які характеризуються частотою, тривалістю роботи, пріоритетом, обсягом даних, що вводяться і виводяться. Запуск тих чи інших завдань здійснюється системою диспетчеризації ІУС за командами, які періодично готує планувальник з урахуванням функціонування ІУС. Для дослідження процесів диспетчеризації в ІУС потрібно використовувати відповідні моделі, що забезпечують проведення моделювання обробки процесів системою залежно від кількості часткових завдань, їх частоти та з врахуванням заданих алгоритмів. Відомо декілька підходів до моделювання ІУС, але на стадії ескізного проєктування доцільно використовувати аналітичні моделі, які враховують стохастичність функціонування ІУС і дають змогу, нехай і відносно, оцінити важливі характеристики системи, що розробляється [28, с. 11].

Простою формою планування завдань, що вирішуються в реальному масштабі часу, є синхронна форма, коли інтервали часу між послідовними запусками однієї і тієї самої завдання кратні одному якому-небудь кванту часу. Гнучкість і ефективність такої системи планування значно підвищується внаслідок введення пріоритетів, а також можливості вирішувати фонові задачі, коли обчислювальний комплекс ІУС звільняється від розв’язання завдань реального масштабу часу. Критерієм ефективності функціонування системи диспетчеризації можуть слугувати різні характеристики затримки часу запуску окремих завдань щодо запланованих моментів: середній час, імовірність того, що час затримки перевищить якусь фіксовану величину тощо. Відомості про ці характеристики можна використовувати під час синтезу системи планування програмного забезпечення ІУС.

Нехай весь інтервал часу за допомогою таймера розбивається на окремі інтервали Δ . Для кожного інтервалу планується виконання певної кількості високопріоритетних і низькопріоритетних завдань відповідно до алгоритму роботи ІУС. Час, вільний в інтервалі від розв'язання завдань реального часу, використовується для розв'язання фонових завдань (наприклад тестовий контроль обладнання). Переривати високопріоритетні задачі заборонено, тоді як низькопріоритетні задачі реального часу – дозволяється.

Перервані низькопріоритетні задачі завершують обслуговування після обробки високопріоритетних задач. Якщо низькопріоритетна задача перервана в i -му інтервалі, то її продовження додається до завдань в $(i + 1)$ -му інтервалі часу. Виконувати кожену задачу планується з певним періодом T_i . Можна виокремити головний цикл обчислень T , що є найменшим спільним кратним $T_i, i=1,2,\dots,N$, де N – загальна кількість поодиноких задач, що вирішують за допомогою програмного забезпечення ІУС. Загалом, час виконання кожної завдання може бути випадковим, у зв'язку з чим важливо знати час тимчасових затримок під час їхнього розв'язання. Особливо це важливо щодо низькопріоритетних завдань, початок виконання яких може зміщуватися високопріоритетними задачами.

Для розгляду процесу розв'язання високопріоритетних задач припустимо таке. Нехай ξ_{ij} – час виконання завдання i -го пріоритету в j -му інтервалі часу; $\Phi_j = \sum_{i \in I_j} \xi_{ij}, i=1,2,\dots,n, n+1, n+2$, є часом розв'язання високопріоритетних задач у j -му інтервалі часу, де I_j – заплановані завдання на j -й інтервал, а $n = T_h / \Delta$, T_h – базовий цикл виконання високопріоритетних задач. Нехай φ_j – середній час розв'язання пріоритетних задач в j -му інтервалі. Сумарний потік обслуговування в кожному інтервалі формується як сума декількох потоків обслуговування щодо кожної задачі і його можна вважати найпростішим, із параметром $\mu_j = 1 / \varphi_j$. Загальну функцію розподілу часу обслуговування запишемо як суперпозицію потоків обслуговування за кожним інтервалом – $B(t) = 1 - \frac{1}{m} \sum_{j=1}^m e^{-\mu_j t}$, тобто вона описується гіперекспоненціальним розподілом. Вхідний потік для такої системи є регулярним, але його можна апроксимувати ерлангівським в тому разі, якщо коефіцієнт варіації наближається до нуля. Функція розподілу такого потоку,

загалом, виглядає так: $A(t) = 1 - \sum_{i=1}^k \frac{(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t}$. Таким чином, уся система

зводиться до однолінійної системи масового обслуговування (СМО) виду Е/Н/1. Дослідимо її.

Нехай $\bar{\lambda} = \frac{\lambda}{k}$ і $\bar{\mu} = \left[\sum_{i=1}^m \frac{1}{m\mu_i} \right]^{-1}$ – середні інтенсивності потоків,

описуваних відповідно до розподілів $A(t)$ і $B(t)$, а $\alpha(s)$ і $\beta(s)$ – перетворення

Лапласа – Стілтєса цих розподілів, тобто $\alpha(s) = \left(\frac{\lambda}{\lambda + s} \right)^k$ і $\beta(s) = \frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + s}$

Якщо $\bar{\mu} > \bar{\lambda}$, то для сталого режиму цієї системи існує функція розподілу часу очікування $F(t)$, для знаходження якої можна скласти інтегральне рівняння Ліндлі [25, с. 124]:

$$F(t) = \begin{cases} \int_{-\infty}^t F(t-x) dC(x), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (4.11)$$

де $C(t) = \int_0^{\infty} B(x+t) dA(t)$.

Розв'язок можна знайти за допомогою факторизації перетворення Лапласа – Стілтєса ядра інтегрального рівняння. Відповідне факторизаційне рівняння буде мати такий вигляд:

$$\gamma(s) = \frac{K_+(s)}{K_-(s)} = \alpha(-s)\beta(s) - 1 = \frac{1}{m} \left(\frac{\lambda}{\lambda - s} \right)^k \sum_{i=1}^m \frac{\mu_i}{\mu_i + s} - 1 = 0. \quad (4.12)$$

Зауважимо, що функція

$$\beta(s) = \frac{\frac{1}{m} \sum_{i=1}^m \mu_i \prod_{j=1}^m (\mu_j + s)}{\prod_{i=1}^m (\mu_i + s)} = \frac{P_{m-1}(s)}{Q_m(s)} \quad (4.13)$$

є відношенням двох багаточленів, до того ж ступінь чисельника менший за ступінь знаменника. Отже, за [25, с. 140] можна записати, що перетворенням Лапласа – Стілтєса від $F(t)$ є

$$\varphi(s) = \frac{Q_m(s)}{Q_m(0) \prod_{i=1}^m \left(1 - \frac{s}{q_i}\right)}, \quad (4.14)$$

де $q_i (i=1,2,\dots,m)$ – корені рівняння

$$\gamma(s) = 0 \Rightarrow \frac{1}{m} \left(\frac{\lambda}{\lambda - s} \right)^k \sum_{i=1}^m \frac{\mu_i}{\mu_i + s} - 1 = 0, \quad (4.15)$$

які містяться в лівій півплощині $\operatorname{Re} s < 0$. Використовуючи (4.13), отримаємо у лівій півплощині $\operatorname{Re} s < 0$.

$$\varphi(s) = \frac{Q_m(s)}{\prod_{i=1}^m \mu_i \left(1 - \frac{s}{q_i}\right)}. \quad (4.16)$$

Нехай $s = \lambda(1 - z)$, тоді рівняння (4.15) буде таким:

$$\frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + \lambda(1 - z)} - z^k = 0. \quad (4.17)$$

Рівняння (4.17) має $m + k$ нулів. k нулів цього рівняння містяться в межах $|z|=1$, беручи до уваги їх кратність. Щоб це довести знайдемо значення модулів обох додатків (4.19) на границі області $|z|=1 - \varepsilon, \varepsilon \rightarrow +0$. Тоді

$$|-z|^k = (1 - \varepsilon)^k. \quad (4.18)$$

Якщо розкласти праву частину (4.19) за формулою Тейлора й відкинути члени вищого порядку малості порівняно з першим, то отримаємо

$$|-z|^k = 1 - k\varepsilon. \text{ Аналогічно } \left| \frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + \lambda(1 - z)} \right| = 1 - \lambda\varepsilon \sum_{i=1}^m \frac{1}{m\mu_i} = 1 - \varepsilon k \frac{\bar{\lambda}}{\bar{\mu}}.$$

Оскільки $\bar{\mu} > \bar{\lambda}$, то $\varepsilon k > \varepsilon \frac{k\bar{\lambda}}{\bar{\mu}} \Rightarrow k > \lambda \sum_{i=1}^m \frac{1}{m\mu_i}$, тобто на границі області

$$|z|=1 - \varepsilon, \varepsilon \rightarrow +0 \quad |-z^k| > \left| \frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + \lambda(1 - z)} \right|. \text{ Тоді, згідно з теоремою Руше,}$$

всередині області $|z| < 1 - \varepsilon$ і $\left(\frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + \lambda(1 - z)} - z^k \right)$ мають однакову кількість нулів – k . Решта m нулів рівняння (4.17) містяться поза областю $|z|=1$.

Позначимо їх як $z_i, i=1,2,\dots,m$. Функція $\gamma_1(s) = \frac{1}{m} \sum_{i=1}^m \frac{\mu_i}{\mu_i + \lambda(1 - z)} - z^k \in$

мероморфною з полюсами у точках $1 + \frac{\mu_i}{\lambda}$, $i = 1, 2, \dots, m$, між якими міститься, принаймні по одному нулю. Оскільки всього полюсів m і в області $|z| > 1 - \frac{\mu_m}{\lambda}$ нулі відсутні, то

$$1 < z_1 < 1 + \frac{\mu_1}{\lambda} < z_2 < \dots < 1 + \frac{\mu_{k-1}}{\lambda} < z_k < 1 + \frac{\mu_k}{\lambda} < \dots < z_m < 1 + \frac{\mu_m}{\lambda}. \quad (4.19)$$

Виразимо q_i через z_i : $q_i = \lambda(1 - z_i)$, $i = 1, 2, \dots, m$. Оскільки $|z_i| > 1$, то q_i міститься у лівій півплощині $\text{Re } s < 0$.

Остаточно отримаємо:

$$\varphi(s) = \lambda^m \prod_{i=1}^m \frac{(\mu_i + s)(1 - z_i)}{\mu_i + [\lambda(1 - z_i) - s]}. \quad (4.20)$$

Використавши обернене перетворення Лапласа – Стілтєса з (4.20), отримаємо вираз для функції розподілу часу затримки щодо обслуговування високопріоритетних задач:

$$F(t) = 1 - \sum_{i=1}^m \frac{\prod_{i=1}^m (1 - z_i)(\mu_i + \lambda - \lambda z_i)}{(z_i - 1) \prod_{l=1}^m \mu_l \sum_{r=1}^m \frac{\prod_{k=1}^m (z_i - z_k)}{(z_i - z_r)}} e^{-\lambda(z_i-1)t}. \quad (4.21)$$

Середній час затримки щодо обслуговування високопріоритетних задач обчислюють так:

$$\bar{t}_d = -\varphi'(s)|_{s=0} = \sum_{i=1}^m \left[\frac{1}{\lambda(z_i - 1)} - \frac{1}{\mu_i} \right].$$

Визначимо дисперсію часу затримки щодо обслуговування високопріоритетних задач:

$$D = \sigma_d^2 = m^2 - \bar{t}_d^2, \quad (4.22)$$

де σ_d – середнєквдратичне відхилення часу затримки, m_2 – другий початковий момент часу затримки. Обчисливши m_2 і підставивши (4.22) \bar{t}_d^2 і m_2 , після перетворення отримаємо:

$$D = \sigma_d^2 = \sum_{i=1}^m \left[\frac{1}{\lambda^2(z_i - 1)^2} - \frac{1}{\mu_i^2} \right].$$

Імовірність того, що час розв'язання високопріоритетних задач ξ перевищить інтервал часу Δ , визначається законом розподілу:

$$P\{\xi > \Delta\} = 1 - B(\Delta).$$

Умови реального часу більше впливають на виконання низькопріоритетних задач, для яких допускається переривання обслуговування. Нехай i – пріоритетний рівень аналізованої низькопріоритетної задачі, а $T_i = N_i \Delta$ – її період виконання. Позначимо через u_j час, необхідний для розв’язання цієї задачі в j -му інтервалі, а через ω_j – сумарний час розв’язання усіх високопріоритетних задач у цьому інтервалі, включаючи й час для закінчення обслуговування задач з $(j - 1)$ -го інтервалу. Тоді час, що залишиться в j -му періоді T_i для виконання розглянутої задачі, визначається виразом $v_j^i = \max\{0, T_i - \omega_j\}$, до того ж v_j^i є випадковою величиною. Вона може бути як більше, так і менше часу, необхідного для розв’язання задачі u_j . Якщо $u_j > v_j^i$, то розглянута низькопріоритетна задача буде очікувати на звільнення процесора від вирішення високопріоритетних задач. Кожному інтервалу v_j^i відповідає інтервал часу u_j . Таким чином, взаємозв’язок між цими випадковими величинами подібний до процесу обслуговування запитів в однолінійній СМО типу G/G/1, до того ж послідовність величин v_j^i описує вхідний потік запитів, а u_j – час обслуговування. Для визначення зв’язку між часом затримки Z під час виконання низькопріоритетної задачі й часом очікування η в еквівалентній однолінійній СМО використаємо таке: якщо задачу починають розв’язувати в запланованому інтервалі часу Δ , то $\eta = z$, в іншому разі $\eta = z - \Delta \cdot \text{int}(z/\Delta)$, де $\text{int}(x)$ – ціла частина виразу x . Отже, $P\{z > t\} = P^*\{\eta > t^*\} = 1 - F(t - \Delta \cdot \text{int}(t/\Delta)) = 1 - W(t - \Delta \cdot \text{int}(t/\Delta))$, де $F(t)$ – закон розподілу часу очікування для еквівалентної СМО, а $W(t)$ – функція розподілу часу очікування.

Апроксимуємо вхідний потік і потік обслуговування розподілами, що належать до сім’ї загальних розподілів Ерланга. Перетворення Лапласа – Стілтєса від їх функцій розподілу виглядають так:

$$f(s) = \prod_{i=1}^k \sum_{j=1}^{n_i} \frac{a_{ij} \lambda_{ij}}{s + \lambda_{ij}}, \quad \sum_{j=1}^{n_i} a_{ij} = 1, \quad (4.23)$$

вони становлять відношення двох поліномів. Експоненціальний, гіперекспоненціальний і ерлангівський розподіли є окремими випадками розподілу (4.23). Усякий розподіл тривалості інтервалів між сусідніми подіями можна апроксимувати з будь-якою точністю за загальним розподілом Ерланга [33, с. 25], необхідно тільки

обрати відповідні параметри $a_{ij}, \lambda_{ij}, n_i, k$. Бажано, щоб величини n_i і k були малими.

Розглянемо однолінійну систему масового обслуговування з очікуванням, яка моделює процес обслуговування низькопріоритетних задач, вхідний потік заявок та час обслуговування і для якої задані узагальнені розподіли Ерланга з перетвореннями Лапласа-Стілтєса від функцій розподілу відповідно:

$$\alpha(s) = \prod_{i=1}^k \sum_{g=1}^{l_i} \frac{a_{ig} \lambda_{ig}}{\lambda_{ig} + s}, \quad \sum_{g=1}^{l_i} a_{ig} = 1, \quad i = 1, 2, \dots, k, \quad (4.24)$$

$$\beta(s) = \prod_{j=1}^n \sum_{q=1}^{c_j} \frac{b_{jq} \mu_{jq}}{\mu_{jq} + s}, \quad \sum_{q=1}^{c_j} b_{jq} = 1, \quad j = 1, 2, \dots, n. \quad (4.25)$$

Вираз для середньої інтенсивності запитів на обслуговування низькопріоритетних завдань $\bar{\lambda}$ обчислюють так:

$$\bar{\lambda} = -\alpha'(0)^{-1} = \left[\sum_{i=1}^k \sum_{g=1}^{l_i} \frac{a_{ig}}{\lambda_{ig}} \right]^{-1}, \quad (4.26)$$

а вираз для середньої інтенсивності процесу обслуговування низькопріоритетних задач $\bar{\mu}$ буде таким:

$$\bar{\mu} = -\beta'(0)^{-1} = \left[\sum_{i=1}^n \sum_{q=1}^{c_i} \frac{b_{iq}}{\mu_{iq}} \right]^{-1}. \quad (4.27)$$

Якщо $\bar{\mu} > \bar{\lambda}$, то для сталого режиму системи існує функція розподілу часу очікування $W(t)$, для визначення якої знову використаємо метод інтегрального рівняння Ліндлі (4.11). Відповідне факторизаційне рівняння виглядає так:

$$\gamma(s) = \frac{K_+(s)}{K_-(s)} = \alpha(-s)\beta(s) - 1 = 0$$

або

$$\gamma(s) = \prod_{i=1}^k \sum_{g=1}^{l_i} \frac{a_{ig} \lambda_{ig}}{\lambda_{ig} + s} \cdot \prod_{j=1}^n \sum_{q=1}^{c_j} \frac{b_{jq} \mu_{jq}}{\mu_{jq} + s} - 1 = 0.$$

Зауважимо, що функція

$$\beta(s) = \prod_{j=1}^n \frac{b_{jq} \mu_{jq} \prod_{l=1}^{c_j} (\mu_{jl} + s)}{\prod_{q=1}^{c_j} (\mu_{jq} + s)} = \frac{P(s)}{Q(s)} \quad (4.28)$$

є відношенням двох багаточленів, до того ж ступінь чисельника менший за ступінь знаменника. Тоді можна записати [25, с. 140]

$$\varphi(s) = -\frac{Q(s)}{Q(0) \prod_{i=1}^f \left(1 - \frac{s}{\eta_i}\right)},$$

де $\varphi(s)$ - перетворення Лапласа - Стілтєса від функції розподілу часу очікування $W(t)$; $\eta_i (i=1, 2, \dots, f)$ - корні рівняння $\gamma(s) = 0$,

що лежать в лівій півплощині $\text{Re } s < 0$, і $f = \sum_{j=1}^n c_j$.

Використавши (4.28), отримаємо

$$\varphi(s) = \frac{\prod_{j=1}^n \prod_{q=1}^{c_j} (\mu_{jq} + s) \prod_{i=1}^f \eta_i}{\prod_{j=1}^n \prod_{q=1}^{c_j} \mu_{jq} \prod_{i=1}^f (\eta_i - s)}. \quad (4.29)$$

Застосувавши зворотнє перетворення Лапласа - Стілтєса від (4.29), знайдемо $W(t)$. Обравши $z_i = -\eta_i$, остаточно отримаємо:

$$W(t) = 1 - \sum_{i=1}^f \frac{\prod_{j=1}^n \prod_{q=1}^{c_j} (\mu_{jq} - z_i) \prod_{i=1}^f z_i}{\prod_{j=1}^n \prod_{q=1}^{c_j} \mu_{jq} \sum_{l=1}^f \frac{\prod_{r=1}^f (z_r - z_i)}{z_l - z_i}} \cdot e^{-z_i t}.$$

Середній час очікування запуску на обслуговування низько-пріоритетних задач у черги -

$$t_q = -\varphi'(0) = \sum_{i=1}^f \frac{1}{z_i} - \sum_{j=1}^n \sum_{q=1}^{c_j} \frac{1}{\mu_{jq}}.$$

Середня кількість запитів щодо розв'язання низькопріоритетних задач у черзі -

$$L_q = \bar{t}_q \cdot \bar{\lambda} = \frac{\sum_{i=1}^f \frac{1}{z_i} - \sum_{j=1}^n \sum_{q=1}^{c_j} \frac{1}{\mu_{jq}}}{\sum_{i=1}^k \sum_{g=1}^{c_j} \frac{a_{ig}}{\lambda_{ig}}},$$

а загальна середня кількість запитів щодо розв'язання низькопріоритетних завдань у системі –

$$L_s = \bar{t}_q \cdot \bar{\mu} = \frac{\sum_{i=1}^f \frac{1}{z_i} - \sum_{j=1}^n \sum_{q=1}^{c_j} \frac{1}{\mu_{jq}}}{\sum_{i=1}^k \sum_{g=1}^{c_j} \frac{b_{ig}}{\lambda_{ig}}}.$$

На підставі результатів проведеного дослідження, можна отримати вичерпні характеристики щодо виконання низькопріоритетних завдань у реальному масштабі часу.

Таким чином, аналіз функціонування процесу розв'язання завдань ІУС в реальному масштабі часу на базі досліджених СМО виду G/G/1 дає змогу спланувати порядок їх функціонування та оцінити важливіші характеристики функціонування всієї інформаційно-управляючої системи.

4.3 Імовірнісний аналіз і вибір параметрів процесора вводу-виводу управляючих комп'ютерів

Використання в системах управління різними об'єктами комп'ютерів передбачає необхідність розв'язання завдань раціонального обміну інформацією комп'ютера з об'єктом управління. Як уже зазначалося, «виведення і, особливо, введення інформації – найвразливіший складник сучасних комп'ютерів» [29, с. 64]. Найбільше це стосується управляючих комп'ютерів нестационарних об'єктів унаслідок жорстких обмежень за габаритами, вагою, споживаною потужністю. Прогрес у мікроелектроніці, створення мікропроцесорів уможливили реалізацію пристрою вводу-виводу інформації управляючого комп'ютера у вигляді автономного блока - процесора вводу-виводу, що здійснює обмін інформацією паралельно до роботи центрального процесора (далі ЦП).

Проектування процесора вводу-виводу для управляючих комп'ютерів нестационарних об'єктів має певні особливості порівняно з проектуванням систем обміну інформацією для універсальних комп'ютерів. Це, насамперед, створення алгоритмів і програм обміну інформацією, з урахуванням

специфіки зовнішніх абонентів системи управління (далі СУ), вибір раціональної структури й параметрів процесора вводу-виводу.

Одним з базових режимів роботи процесора вводу-виводу є обмін масивами слів між оперативним запам'ятовуючим пристроєм (далі ОЗП) ЦП і абонентами СУ. Оскільки ЦП зазвичай обмінюється з різними абонентами масивами слів випадкового обсягу, то час обробки одного запиту процесором вводу-виводу є випадковим. Час між надходженням окремих запитів для обміну теж випадковий. Таким чином, процесор вводу-виводу працює в режимі системи масового обслуговування.

Для спрощення аналізу будемо вважати, що процес передачі інформації відбувається в одному напрямі - від ЦП до абонентів СУ і джерелом запитів на обмін є центральний процесор.

Процес передачі одного слова з ОЗП ЦП абоненту СУ відбувається в два етапи: спочатку це слово передається з ОЗП ЦП на один із регістрів процесора вводу-виводу в паралельному коді, а потім - в інтерфейс, із розбивкою на байти або біти інформації залежно від типу лінії зв'язку.

Таким чином, якщо одне слово передається з ОЗП ЦП у процесор вводу-виводу за час τ , то це саме слово передається з процесора вводу-виводу абоненту СУ за час $k\tau$, де k може змінюватися від 1 до m - розрядності паралельного слова.

З урахуванням зазначеного вище для спрощення аналізу роботи процесора вводу-виводу формально вважається, що процес передачі масиву слів обсягом N з ОЗП ЦП абоненту СУ складається з двох етапів: передача цього масиву з ОЗП ЦП у процесор вводу-виводу за час $N\tau$ і передача масиву слів із процесора вводу-виводу абоненту СУ за час $kN\tau$.

Вважатимемо, що потік запитів на обмін, який надходить із центрального процесора в процесор вводу-виводу, підпорядковується експоненціальному розподілу з параметром λ ; розмір масивів слів, якими обмінюються ЦП і абонент СУ, також підпорядковується експоненціальному розподілу із середнім значенням m . Середня інтенсивність обробки запитів під час прийому інформації в процесор вводу-виводу з ОЗП ЦП - $\mu = 1/m\tau$, а під час видачі - $\mu_{out} = \mu/k$. Розглянемо щодо цих припущень низку схем організації процесора вводу-виводу.

Схема 1. Центральний процесор ініціює обмін масивом слів, процесор вводу-виводу послідовно, слово за словом, передає цей масив абоненту СУ. До закінчення обміну ЦП не може ініціювати нової вимоги на обмін.

Граф станів і переходів зазначеної схеми модуля обміну зображений на рисунку 4.4, а, де [0] - стан, коли модуль простоює, [10] - стан, коли процесор вводу-виводу приймає масив слів з ОПУ ЦП, а [01] - стан, коли процесор вводу-виводу передає черговий масив слів абоненту СУ.

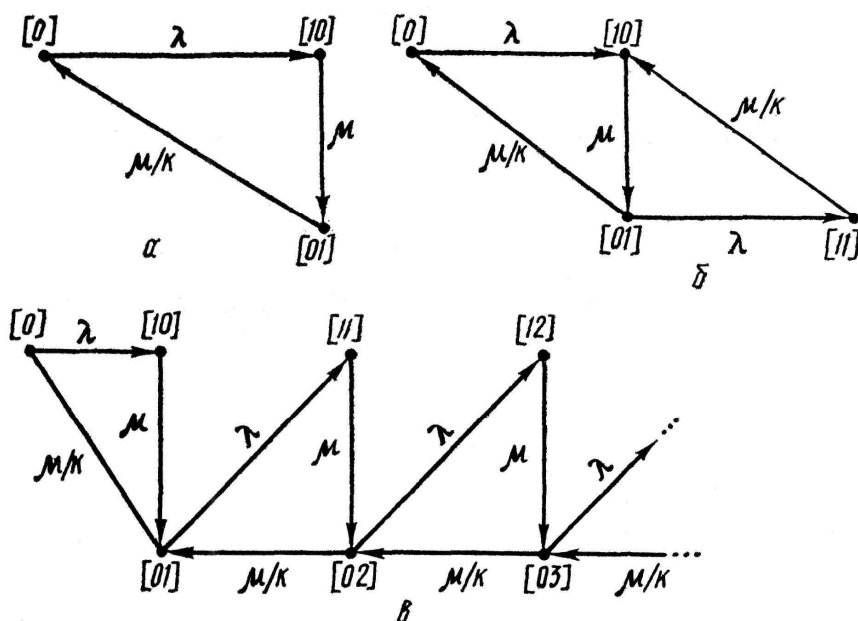


Рисунок 4.4 – Графи станів і переходів для схем побудови процесора вводу-виводу

Визначимо характеристики функціонування процесора вводу-виводу, побудованого за цією схемою, склавши для цього систему рівнянь для ймовірностей станів за цим графом станів і переходів:

$$\begin{aligned}
 -\lambda p_0 + \mu/k p_{01} &= 0, \\
 -\mu p_{10} + \lambda p_0 &= 0, \\
 p_0 + p_{01} + p_{10} &= 1.
 \end{aligned}
 \tag{4.30}$$

Розв'язавши систему, одержимо таке:

$$\text{– ймовірність простоювання процесора вводу-виводу – } p_0 = \frac{k}{k + (1+k)\rho},$$

де $\rho = \frac{\lambda k}{\mu}$ – завантаження процесора вводу-виводу процесом видачі інформації;

$$\text{– пропускна здатність } q_1 = \frac{1 - p_0}{\frac{1}{\mu} + \frac{k}{\mu}} = \frac{\lambda k}{k + (1+k)\rho};$$

$$\text{– ймовірність блокування ОЗП ЦП - } R_1 = 1 - p_0 = \frac{(1+k)\rho}{k + (1+k)\rho}.$$

Схема 2. У процесор вводу-виводу введено буферний запам'ятовуючий пристрій (далі БЗП), із розмірністю максимально можливого масиву слів. Процес передачі масиву слів змінений порівняно зі схемою 1, а саме: спочатку масив, що передається, записується в БЗП, а потім передається абоненту СУ. Граф станів і переходів для цієї схеми той самий, що й для схеми 2. Особливістю цієї схеми є зменшення ймовірності блокування ОЗУ ЦП, а саме:

$$R_2 = p_{01} = \frac{\rho}{k + (1+k)\rho}.$$

Час блокування схеми 2 організації модуля обміну порівняно зі схемою 1 зменшується тим більше, чим більшим є час передачі одного слова з процесора вводу-виводу абоненту СУ порівняно з часом передачі з ОЗП ЦП у процесор вводу-виводу.

Схема 3. Модуль має БЗП, що вміщує масив інформації щодо одного запиту. Процесор може генерувати новий запит на обмін, коли модуль обміну здійснює передачу чергового масиву слів абоненту СУ. Граф станів і переходів для такого варіанту процесора вводу-виводу зображений на рисунку 4.4, б: [11] - стан, коли процесор вводу-виводу здійснює передачу масиву слів абоненту, а процесор видає новий запит на обмін.

Система рівнянь для ймовірностей станів за цим графом є такою:

$$\begin{aligned} -\lambda p_0 + \mu/k p_{01} &= 0, \\ -\mu p_{10} + \lambda p_0 + \mu/k p_{11} &= 0, \\ -(\lambda + \mu/k)p_{01} + \mu p_{10} &= 0, \\ p_0 + p_{01} + p_{10} + p_{11} &= 1. \end{aligned} \tag{4.31}$$

Її розв'язок виглядає так:

$$\begin{aligned} p_0 &= \frac{k}{k + (1+\rho)(1+k)\rho}, p_{01} = \frac{k\rho}{k + (1+\rho)(1+k)\rho}, p_{10} = \frac{\rho(1+\rho)}{k[k + (1+\rho)(1+k)\rho]}, \\ p_{11} &= \frac{k\rho^2}{k + (1+\rho)(1+k)\rho}, \rho = \frac{\lambda k}{\mu}. \end{aligned}$$

Базові характеристики функціонування процесора вводу-виводу будуть такими:

$$- \text{імовірність блокування ОЗП ЦП } R_3 = p_{10} + p_{11} = \frac{\rho[1 + \rho(1+k)]}{k + (1+\rho)(1+k)\rho},$$

$$\text{– пропускна здатність } q_3 = \frac{1 - p_0}{\frac{1}{\mu} + \frac{k}{\mu}} = \frac{\lambda k(1 + \rho)}{k + (1 + \rho)(1 + k)\rho}.$$

Порівняємо за пропускною здатністю схему 3 організації процесора вводу-виводу зі схемою 2: $\Delta q_{3,2} = q_3 - q_2 = \frac{\lambda \rho k^2}{[k + (1 + k)\rho][k + (1 + \rho)(1 + k)\rho]}$.

Графік залежності $\delta_{3,2} = \Delta q_{3,2} / q_2 \cdot 100\%$ у разі фіксованого λ подано на рисунку 4.5, а.

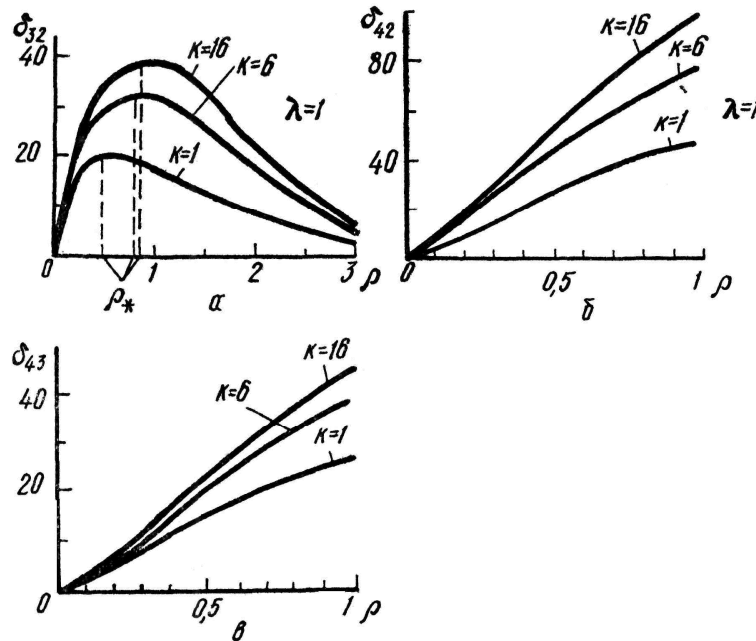


Рисунок 4.5 – Відносний відсотковий виграш схем побудови процесора вводу-виводу

Як зрозуміло з графіка, функція $\delta_{3,2}$ має максимум в точці ρ_* , яку можна визначити з рівняння $\delta'_{3,2} = 0$.

Таким чином, у разі фіксованого значенні λ схема 3 організації процесора вводу-виводу в околиці ρ_* за пропускною здатністю набагато перевищує схеми 1 і 2 організації обміну. Водночас в області малих і великих значень ρ схема 3 організації процесора вводу-виводу фактично не має переваг порівняно зі схемами 1 і 2.

Схема 4. Особливістю цієї схеми організації процесора вводу-виводу є алгоритм прийому масивів інформації і видачі їх абонентам СУ. Масив інформації спочатку приймається з ОЗП ЦП в БЗП процесора вводу-виводу, після чого відбувається видача цього масиву з БЗУ процесора вводу-виводу абоненту СУ з відповідною розбивкою слів на байти або біти інформації.

Якщо в момент передачі чергового масиву інформації з БЗП процесора вводу-виводу абоненту СУ надійшов новий запит на обмін, то процес видачі інформації припиняється і модуль обміну починає приймати в БЗП новий масив інформації. Після прийому продовжується процес видачі.

Граф станів і переходів для такої схеми організації процесора вводу-виводу зображений на рисунку 4.4, в: $[1n]$ - стан, коли в БЗУ є n масивів слів, а модуль здійснює прийом нового масиву з ОЗУ ЦП; $[0n]$ - стан, коли в БЗП процесора вводу-виводу є n масивів слів і модуль здійснює видачу чергового масиву абоненту СУ.

Передбачається, що обсяг БЗП досить великий і втрат інформації не відбувається. При цих припущеннях система рівнянь, що описує роботу такого варіанта процесора вводу-виводу в стаціонарному режимі, має такий вигляд:

$$\begin{aligned}
 -\lambda p_0 + \mu/k p_{01} &= 0, \\
 -\mu p_{10} + \lambda p_0 &= 0, \\
 -(\lambda + \mu/k)p_{01} + \mu p_{10} + \mu/k p_{02} &= 0, \\
 -\mu p_{1n} + \lambda p_{0n} &= 0, \quad n > 0, \\
 -(\lambda + \mu/k)p_{0,n+1} + \mu p_{1n} + \mu/k p_{0,n+2} &= 0, \quad n > 0.
 \end{aligned} \tag{4.32}$$

Умовою нормування є $p_0 + \sum_{n=1}^{\infty} p_{0n} + \sum_{n=1}^{\infty} p_{1n} = 1$, умовою стаціонарного режиму функціонування є виконання умови $\rho = k\lambda/\mu < 1$, де ρ – завантаження усього процесора вводу-виводу.

Розв'язавши систему рівнянь (4.32), отримаємо:

$$p_0 = \frac{(1-\rho)k}{k+\rho}, \quad p_{0n} = \frac{1-\rho}{k+\rho} k\rho^n, \quad p_{1n} = \frac{1-\rho}{k+\rho} \rho^n. \tag{4.33}$$

$$\text{Середня кількість запитів у процесорі вводу-виводу} - L_s = \frac{(1+k)\rho}{(k+\rho)(1-\rho)},$$

пропускна здатність – $q_4 = \frac{\lambda k}{k+\rho}$. Перевага пропускної здатності процесора

вводу-виводу, побудованого за схемою 4, порівняно зі схемами 1, 2 і 3 визначається за такими виразами

$$\Delta q_{4,2} = q_4 - q_2 = \frac{\lambda \rho k^2}{(k+\rho)[k+(1+k)\rho]}, \tag{4.34}$$

$$\Delta q_{4,3} = q_4 - q_3 = \frac{\lambda \rho^2 k^2}{(k+\rho)[k+(1+\rho)(1+k)\rho]}. \tag{4.35}$$

Графіки залежностей $\delta_{4,2} = \frac{\Delta q_{4,2}}{q_2} 100\%$ і $\delta_{4,3} = \frac{\Delta q_{4,3}}{q_3} 100\%$ щодо

завантаженості процесора вводу-виводу ρ подані на рисунках 4.5, б і 4.5, в. Із графіків зрозуміло, що схема 4 організації процесора вводу-виводу в області високого завантаження має суттєві переваги за пропускнуою здатністю істотно переважає інші схеми організації процесора вводу.

Контрольні питання

1. Які функції виконує процесор вводу-виводу?
2. Що таке число-імпульсна інформація і як вона використовується?
3. Як обробляється число-імпульсна інформація?
4. Як відбувається зняття інформації в модулі прийому число-імпульсної інформації?
5. Обґрунтуйте використання ймовірнісної моделі для аналізу процесу функціонування модуля прийому число імпульсної інформації.
6. Побудуйте граф станів і переходів для одного лічильника модуля прийому число-імпульсної інформації.
7. Який критерій використовують для оптимізації процесу функціонування модуля прийому число-імпульсної інформації?
8. Який метод доцільно використовувати для оптимізації критерію функціонування модуля прийому число імпульсної інформації? Відповідь обґрунтуйте.
9. Яку модель використовують для оцінки якості функціонування інформаційно-управляючих систем?
10. Який критерій використовують для оцінки ефективності функціонування системи диспетчеризації?
11. Які закони розподілу використовують під час моделювання процесу функціонування системи диспетчеризації?
12. Який метод використовують для аналізу моделі функціонування системи диспетчеризації?
13. Яку функцію виконує процесор вводу-виводу управляючих комп'ютерів?
14. У яких режимах функціонує процесор вводу-виводу управляючих комп'ютерів?
15. Обґрунтуйте ймовірнісність процесу функціонування процесору вводу-виводу.

16. За яким критерієм порівнюють схеми побудови процесора вводу-виводу?

17. Які ймовірнісні закони розподілу використовуються в моделях функціонування процесора вводу-виводу?

ПІСЛЯМОВА

Значна частина процесів, що відбуваються в комп'ютерних системах, є ймовірнісними, більшість їх пов'язана з обслуговуванням певних запитів. Особливо це стосується функціонування операційних систем.

Процеси масового обслуговування формуються в найрізноманітніших галузях техніки, науки й у повсякденному житті. Прикладом може служити поліклініка, де спостерігаються черги на прийом до спеціалістів, яких можна вважати обслуговуваними об'єктами. Автомобілі, що приїжджають на автозаправочні станції, змушені вишикуватися в чергу, очікуючи обслуговування. Деколи така черга становить на сотні метрів.

Математична дисципліна, що вивчає моделі реальних систем масового обслуговування, отримала назву «Теорія систем масового обслуговування». Завдання цієї дисципліни встановити залежність результуючих показників роботи системи масового обслуговування (імовірності того, що вимога буде обслужена, математичного сподівання кількості обслужених вимог тощо) від вхідних показників (кількості каналів у системі, параметрів вхідного потоку вимог тощо). Теорія систем масового обслуговування є новим напрямом у теорії ймовірностей, що сформувався як самостійна наукова дисципліна, внаслідок специфіки застосовуваного математичного апарату і важливості розв'язуваних практичних задач.

Розпочав розв'язувати практичні задачі масового обслуговування співробітник Копенгагенської телефонної компанії данський математик А. К. Ерланг у період 1908-1922 років. У 1909 році з'явилася його робота «Теорія ймовірностей і телефонні переговори» та інші публікації, у яких було сформульовано перші прикладні задачі теорії масового обслуговування. Ці завдання були пов'язані з необхідністю впорядкувати роботу телефонної мережі і розробити методи оцінки залежності якості обслуговування споживачів від кількості використовуваних пристроїв. Протягом наступних років теорія систем масового обслуговування інтенсивно розвивалася працями як зарубіжних, так і вітчизняних учених. Було досліджено багато моделей, що мають велике практичне значення. До таких моделей належать багатолінійні й багатофазні, із кінцевим числом джерел запитів, із урахуванням надійності каналу обслуговування. Однак системи масового обслуговування з пріоритетами досліджувалися недостатньо, хоч вони дуже потрібні для дослідження процесів, що відбуваються в операційних системах. Наприклад, в операційній системі Windows передбачено до двадцяти рівнів пріоритетів.

Аналіз систем масового обслуговування спрощується, якщо в системі відбуваються марківські процеси; тоді систему можна описати звичайними диференціальними рівняннями, а граничні ймовірності - лінійними алгебраїчними рівняннями. У разі використання немарківських систем більшість результатів стосується однолінійних моделей, що зазвичай мають тільки теоретичне значення, тому під час реалізації складних і дороговартісних проектів, пов'язаних із масовим обслуговуванням, використовують імітаційне моделювання, хоча розробка імітаційної моделі може коштувати дуже дорого і займе багато часу [30, с. 26].

У зв'язку з важливістю теорії систем масового обслуговування для розуміння процесів, що відбуваються в комп'ютерних системах, у план підготовки бакалаврів, які навчаються за напрямом «Комп'ютерні науки», включена дисципліна «Теорія систем масового обслуговування». Щоб оволодіти цією дисципліною, студенти повинні отримати знання з дисциплін «Вища математика», «Теорія ймовірностей і математична статистика», «Чисельний аналіз», «Програмування».

Мета дисципліни «Теорія систем масового обслуговування» - надати студентам теоретичні відомості з основ дисципліни, сформувати практичні навички щодо побудови моделей масового обслуговування та їх дослідження на підставі залучення сучасних інформаційних технологій. Дисципліна «Теорія систем масового обслуговування» забезпечує подальше підвищення рівня фундаментальної математичної та комп'ютерної підготовки студентів, а також розвиток абстрактного та логічного мислення студентів.

Теорія систем масового обслуговування використовується в багатьох сферах науки, техніки та економіки. Одночасно з технікою передачі, розподілу та обробки інформації моделі масового обслуговування використовуються під час організації виробництва й логістики. Системи масового обслуговування служать чудовим матеріалом для наукових досліджень, їх можна використовувати в дипломному проектуванні для обґрунтування прийнятих проектних рішень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Цикритзис Д. Операционные системы / Д. Цикритзис, Ф. Бернстайн. – М. : Мир, 1977. – 336 с.
2. Овчаров Л. А. Прикладные задачи теории массового обслуживания / Л. А. Овчаров. – М. : Машиностроение, 1969. – 324 с.
3. Клейнрок Л. Вычислительные системы с очередями / Л. Клейнрок. – М. : Мир, 1979. – 600 с.
4. Хинчин А. Я. Работы по математической теории массового обслуживания / А. Я. Хинчин. – М. : ГИФМЛ, 1963. – 235 с.
5. Гнеденко Б. В. Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. – М. : Наука, 1966. – 431 с.
6. Литвинов А. Л. Общий анализ пуассоновских моделей обслуживания с двойной обработкой запросов на одном приборе / А. Л. Литвинов // Кибернетика: науч. журнал. – Київ : Наукова думка, 1983. – № 2. – С. 82 – 91.
7. Литвинов А. Л. Вероятностный анализ функционирования и выбор параметров блока приема число–импульсной информации управляющего вычислительного комплекса / А. Л. Литвинов // Автоматика и телемеханика: науч. журнал. – М. : Наука, 1983. – №7. С. 170 – 174.
8. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высш. шк., 2002. – 479 с.
9. Колмогоров А. Н. Основные понятия теории вероятностей / А. Н. Колмогоров. – М. : Наука, 2013. – 120 с.
10. Донченко В.С. Теорія ймовірностей та математична статистика для соціальних наук: навчальний посібник / В. С. Донченко, М. В.-С. Сидоров. – Київ : ВПС Київський університет, 2015. – 400 с.
11. Economic Definition [Электронный ресурс]. – Режим доступа: http://economic-definition.com/Physics/Process_Process_eto.html.
12. Гнеденко Б. В. Курс теории вероятностей / Б. В. Гнеденко. – М. : Наука, 1965. – 400 с.
13. Карлин С. Основы теории случайных процессов / С. Карлин. – М. : Мир, 1971. – 536 с.
14. Вентцель Е. С. Теория случайных процессов и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров. – М. : Наука, 1991. – 384 с.
15. Камке Э. Справочник по обыкновенным дифференциальным уравнениям / Э. Камке. – СПб. : Лань, 2003. – 576 с.

16. Новиков О. А. Прикладные вопросы теории массового обслуживания / О. А. Новиков, С. И. Петухов. – М. : Советское радио, 1969. – 400 с.
17. Джейсуол Н. Очереди с приоритетами / Н. Джейсуол. – М. : Мир, 1973. – 279 с.
18. Литвинов А. Л. Вероятностный анализ буферной памяти вычислительного комплекса с групповым адаптивным выбором информации / А. Л. Литвинов // Автоматизированные системы управления: тематический сборник. – Харьков : ХАИ, 1984. – Вып. 5. – С. 82–87.
19. Клейнрок Л. Теория массового обслуживания / Л. Клейнрок. – М. : Машиностроение, 1979. – 432 с.
20. Кофман А. Массовое обслуживание. Теория и приложения / А. Кофман, Р. Крюон. – М. : Мир, 1965. – 303 с.
21. Литвинов А. Л. Вероятностная модель буферной памяти управляющего вычислительного комплекса с групповым выбором / А. Л. Литвинов, В. А. Попов // Известия АН СССР. Техническая кибернетика : науч. журнал. – М. : АН СССР, 1978. – № 1. – С. 91–95.
22. Риордан Дж. Вероятностные системы обслуживания / Дж. Риордан. – М. : Связь, 1966. – 184 с.
23. Саати Т. Л. Элементы теории массового обслуживания и ее приложения / Т. Л. Саати. – М. : Советское радио, 1965. – 520 с.
24. Дубовик В. П. Вища математика / В. П. Дубовик, І. І. Юрик. – Київ : Ігнатекс-Україна, 2013. – 648 с.
25. Климов Г. П. Стохастические системы обслуживания / Г. П. Климов – М. : Наука, 1966. – 343 с.
26. Тараканов К. В. Аналитические методы исследования систем / К. В. Тараканов, Л. А. Овчаров, А. Н. Тырышкин. – М. : Советское радио, 1974. – 240 с.
27. Алексеева Е. В. Численные методы оптимизации : учеб. пособие / Е. В. Алексеева, О. А. Кутненко, А. В. Плясунов. – Новосибирск : Новосиб. ун-т, 2008. – 128 с.
28. Болтенков В. И. Конфигурирование и настройка автоматизированных информационных систем : учебное пособие / В. И. Болтенков, А. Л. Литвинов, Н. В. Лычева. – Белгород : Изд-во БелГУ, 2005. – 248 с.
29. Глушков В. М. Введение в АСУ / В. М. Глушков. – Киев : Техніка, 1974. – 320 с.

30. Шеннон Р. Имитационное моделирование систем – искусство и наука / Р. Шеннон. – М. : Мир, 1978. – 418 с.

31. Литвинов А. Л. Розробка та дослідження ймовірнісних моделей оцінювання якості інформаційно-управляючих систем / А. Л. Литвинов // Комунальне господарство міст. : Серія: Технічні науки та архітектура. – Харків : ХНУМГ, 2016. – вип. 126, – С. 22 – 26.

32. Литвинов А. Л. Оценка вариантов построения и выбор параметров канала обмена управляющих ЦВМ / А. Л. Литвинов // Управляющие системы и машины : науч. журнал. – Киев : Наукова думка, 1978. – № 5, С. 81 – 84.

33. Кокс Д. Р. Теория восстановления / Д. Р. Кокс, В. Л. Смит. – М. : Советское радио, 1967. – 299 с.

ДОДАТОК А

Таблиця значень функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-z^2/2) dz$

X	$\Phi(X)$	X	$\Phi(X)$	X	$\Phi(X)$	X	$\Phi(X)$	X	$\Phi(X)$	X	$\Phi(X)$	X	$\Phi(X)$
0,00	0,0000	0,39	0,1517	0,78	0,2823	1,17	0,3790	1,56	0,4406	1,95	0,4744	2,38	0,4913
0,01	0,0040	0,40	0,1554	0,79	0,2852	1,18	0,3810	1,57	0,4418	1,96	0,4750	2,40	0,4918
0,02	0,0080	0,41	0,1591	0,80	0,2881	1,19	0,3830	1,58	0,4429	1,97	0,4756	2,42	0,4922
0,03	0,0120	0,42	0,1628	0,81	0,2910	1,20	0,3849	1,59	0,4441	1,98	0,4761	2,44	0,4927
0,04	0,0160	0,43	0,1664	0,82	0,2939	1,21	0,3869	1,60	0,4452	1,99	0,4767	2,46	0,4931
0,05	0,0199	0,44	0,1700	0,83	0,2967	1,22	0,3883	1,61	0,4463	2,00	0,4772	2,48	0,4934
0,06	0,0239	0,45	0,1736	0,84	0,2995	1,23	0,3907	1,62	0,4474	2,01	0,4772	2,50	0,4938
0,07	0,0279	0,46	0,1772	0,85	0,3023	1,24	0,3925	1,63	0,4484	2,02	0,4783	2,52	0,4941
0,08	0,0319	0,47	0,1808	0,86	0,3051	1,25	0,3944	1,64	0,4495	2,03	0,4788	2,54	0,4945
0,09	0,0359	0,48	0,1844	0,87	0,3078	1,26	0,3962	1,65	0,4505	2,04	0,4793	2,56	0,4948
0,10	0,0398	0,49	0,1879	0,88	0,3106	1,27	0,3980	1,66	0,4515	2,05	0,4798	2,58	0,4951
0,11	0,0438	0,50	0,1915	0,89	0,3133	1,28	0,3997	1,67	0,4525	2,06	0,4803	2,60	0,4953
0,12	0,0478	0,51	0,1950	0,90	0,3159	1,29	0,4015	1,68	0,4535	2,07	0,4807	2,62	0,4956
0,13	0,0517	0,52	0,1985	0,91	0,3186	1,30	0,4032	1,69	0,4545	2,08	0,4812	2,64	0,4959
0,14	0,0557	0,53	0,2019	0,92	0,3212	1,31	0,4049	1,70	0,4554	2,09	0,4817	2,66	0,4961
0,15	0,0596	0,54	0,2054	0,93	0,3238	1,32	0,4066	1,71	0,4554	2,10	0,4821	2,68	0,4963
0,16	0,0636	0,55	0,2088	0,94	0,3264	1,33	0,4082	1,72	0,4573	2,11	0,4826	2,70	0,4965
0,17	0,0675	0,56	0,2123	0,95	0,3289	1,34	0,4099	1,73	0,4582	2,12	0,4830	2,72	0,4967
0,18	0,0714	0,57	0,2157	0,96	0,3315	1,35	0,4115	1,74	0,4591	2,13	0,4834	2,74	0,4969
0,19	0,0753	0,58	0,2190	0,97	0,3340	1,36	0,4131	1,75	0,4599	2,14	0,4838	2,76	0,4971
0,20	0,0793	0,59	0,2224	0,98	0,3365	1,37	0,4147	1,76	0,4608	2,15	0,4842	2,78	0,4973
0,21	0,0832	0,60	0,2257	0,99	0,3389	1,38	0,4162	1,77	0,4616	2,16	0,4846	2,80	0,4974
0,22	0,0871	0,61	0,2291	1,00	0,3413	1,39	0,4177	1,78	0,4525	2,17	0,4850	2,82	0,4976
0,23	0,0910	0,62	0,2324	1,01	0,3438	1,40	0,4192	1,79	0,4633	2,18	0,4854	2,84	0,4977
0,24	0,0948	0,63	0,2357	1,02	0,3401	1,41	0,4207	1,80	0,4641	2,19	0,4857	2,86	0,4979
0,25	0,0987	0,64	0,2389	1,03	0,3485	1,42	0,4222	1,81	0,4649	2,20	0,4861	2,88	0,4980
0,26	0,1026	0,65	0,2422	1,04	0,3508	1,43	0,4236	1,82	0,4656	2,21	0,4864	2,90	0,4981
0,27	0,1064	0,66	0,2454	1,05	0,3531	1,44	0,4251	1,83	0,4664	2,22	0,4868	2,92	0,4982
0,28	0,1103	0,67	0,2486	1,06	0,3554	1,45	0,4265	1,84	0,4671	2,23	0,4871	2,94	0,4984
0,29	0,1141	0,68	0,2517	1,07	0,3577	1,46	0,4279	1,85	0,4678	2,24	0,4875	2,96	0,4985
0,30	0,1179	0,69	0,2549	1,08	0,3599	1,47	0,4292	1,86	0,4686	2,25	0,4878	2,98	0,4985
0,31	0,1217	0,70	0,2580	1,09	0,3621	1,48	0,4305	1,87	0,4693	2,26	0,4881	3,00	0,49865
0,32	0,1255	0,71	0,2611	1,10	0,3643	1,49	0,4319	1,88	0,4699	2,27	0,4884	3,20	0,49931
0,33	0,1293	0,72	0,2642	1,11	0,3665	1,50	0,4332	1,89	0,4706	2,28	0,4887	3,40	0,49966
0,34	0,1331	0,73	0,2673	1,12	0,3686	1,51	0,4345	1,90	0,4713	2,29	0,4890	3,60	0,49984
0,35	0,1368	0,74	0,2703	1,13	0,3708	1,52	0,4357	1,91	0,4719	2,30	0,4893	3,80	0,49992
0,36	0,1406	0,75	0,2734	1,14	0,3729	1,53	0,4370	1,92	0,4726	2,32	0,4898	4,00	0,49996
0,37	0,1443	0,76	0,2764	1,15	0,3749	1,54	0,4382	1,93	0,4732	2,34	0,4904	4,50	0,49999
0,38	0,1480	0,77	0,2794	1,16	0,3770	1,55	0,4394	1,94	0,4738	2,36	0,4909	5,00	0,49999

ДОДАТОК Б

Критичні точки розподілу

Кількість степенів свободи	Рівень значимості					
	0,01	0,025	0,05	0,95	0,97	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	9,5	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,5
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,4	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Навчальне видання

ЛИТВИНОВ Анатолій Леонідович

ТЕОРІЯ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ

НАВЧАЛЬНИЙ ПОСІБНИК

Відповідальний за випуск *М. М. Булаєнко*

Редактор *О. А. Норик*

Комп'ютерне верстання *А. Л. Литвинов*

Дизайн обкладинки *Т. А. Лазуренко*

Підп. до друку 17.05.2018. Формат 60x84/16.

Друк на ризографі. Ум. друк. арк. 5.

Тираж 60 пр. Зам. №

Видавець і виготовлювач:

Харківський національний університет міського господарства імені О. М. Бекетова,
вул. Маршала Бажанова, 17, Харків, 61002.

Електронна адреса: rektorat@kname.edu.ua.

Свідоцтво суб'єкта видавничої справи:

ДК № 5328 від 11.04.2017.