МАТЕМАТИЧНІ МЕТОДИ МОДЕЛЮВАННЯ ТА ОПТИМІЗАЦІЇ ПРОЦЕСІВ

Тема 3. Поняття, задачі та класифікація систем масового обслуговування

План лекції:

Вступ.
1. Поняття системи масового обслуговування. Приклади СМО.
2. Предмет і задачі теорії масового обслуговування.
3. Класифікація СМО.
4. Одноканальна СМО з відмовами.
5. Багатоканальна СМО з відмовами.
Заключення.

Вступ.

В **теорії масового обслуговування (ТМО)** розглядаються процеси обробки інформації в складних технічних системах. Основи цієї теорії були закладені в працях датського математика, співпрацівника Копенгагенської телефонної компанії А.К. Ерланга (формулювання принципу статистичної рівноваги) і отримали подальшого розвитку в роботах багатьох вітчизняних та зарубіжних вчених, таких як Т. Енгсет, Г.О. Делл, Е. Молін, О. Колмогоров, А. Хінчин, К. Пальм, Г. Башарін, А. Маркевич, Б. Лівшиц, Ю. Корнишев та ін.

Основними поняттями системи масового обслуговування (СМО) є **заявки (вимоги)** та **сервери**, які також називають **обслуговуючими приладами (каналами)**. Заявки створюють вхідний потік на вході СМО, а система обслуговує ці заявки, використовуючи на кожну деякий час. Якщо кількість каналів замала для обслуговування усіх заявок, що надійшли на якийсь момент, то виникає конфлікт, вирішення якого полягає в тому, що частина заявок <u>відкидається</u> або <u>розміщується в чергу</u>. Тому в англомовній літературі звичайно використовується термін Queuing Theory (теорія черг).

Отже, **теорія масового обслуговування досліджує математичну модель процесу обслуговування**, яка містить такі основні компоненти:

- потік заявок (викликів, повідомлень), що надходить в систему;
- час обслуговування викликів;
- організація обслуговування (система і дисципліна);
- характеристики якості обслуговування вхідних заявок.

Поняття потоку заявок досить широке. Під ним розуміють потік подій, який впливає на систему та вимагає від системи витрачати певний час на їх обслуговування.

Система обслуговування характеризується кількістю каналів та структурою побудови (повно- або неповнодоступна, одно- або багатоланкова, одно- або багатофазна і т.д.) і набором структурних параметрів. Під дисципліною обслуговування розуміють спосіб обслуговування (без втрат повідомлень, з явними втратами, з очікуванням, з повторенням або комбінований), порядок обслуговування (за чергою, у випадковому порядку або за пріоритетом), режим пошуку виходів комутаційної системи (вільний, груповий або індивідуальний), спосіб зайняття вільних каналів (послідовний або випадковий), а також іншу інформацію, що характеризує взаємодію потоку повідомлень з системою обслуговування.

До характеристик якості обслуговування вхідного потоку заявок відносять: імовірність явної або умовної втрати повідомлень через відсутність вільних каналів обслуговування або шляхів встановлення з'єднання, середній час затримки початку обслуговування вхідного повідомлення, імовірність втрати первинного або повторного виклику, інтенсивність обслугованого навантаження, пропускна здатність системи обслуговування та ін.

Домінуюче положення в ТМО займають **задачі аналізу** – визначення характеристик якості обслуговування залежно від параметрів і властивостей вхідного потоку заявок, а також від параметрів і структури системи обслуговування. Іншою важливою задачею є **задача синтезу** – визначення параметрів або структури системи, що здатна забезпечити нормовані показники якості для певного вхідного потоку заявок.

З розвитком елементної бази вузлів мереж зв'язку, вдосконаленням принципів та систем управління в телекомунікаціях та впровадженням нових послуг задачі, що вирішує ТМО, суттєво розширюються. Пошук найбільш економічних структур комутаційних схем, розробка принципів їх побудови пов'язуються з дотриманням умов не тільки щодо пропускної здатності та часу передачі повідомлення через комутаційне поле, але і надійності функціонування, гнучкості розвитку, модульності побудови.

Для складних систем, що можна звести до СМО, рішення задач аналізу і синтезу передбачає визначення необхідного рівня децентралізації управління (вибір оптимальної структури управління) залежно від конкретних характеристик системи (ємність, призначення та ін.), оптимального розподілу функцій та ресурсів системи управління між окремими ЕОМ та мікроЕОМ при децентралізованому та розподіленому управлінні, а також пошук оптимальних алгоритмів обробки викликів і процедур взаємодії між окремими елементами системи управління з метою підвищення її ефективності.

Математичний апарат ТМО базується на теорії імовірності, комбінаториці та математичній статистиці. Методи останньої застосовуються здебільшого для обробки даних, які отримуються при вимірюванні параметрів потоків повідомлень та показників якості обслуговування в реальних системах, а також при моделюванні таких систем на ЕОМ. Для рішення конкретних задач використовуються також інші розділи математики – лінійна алгебра, диференційне та інтегральне обчислення, теорія графів, системний аналіз.

Основним інструментом дослідження в ТМО є **метод рівнянь імовірностей станів,** оснований на принципі статистичної рівноваги. Для системи обслуговування вводиться поняття ***стану***. В найпростішому випадку стан системи характеризується однією випадковою змінною, наприклад числом зайнятих ліній або викликів, що находяться в системі (обслуговуються або чекають в черзі). При надходженні наступного виклику або закінченні обслуговування система змінює свій стан. Інтенсивності переходу з одного стану в інший звичайно відомі на основі властивостей потоків викликів і звільнень. Це дозволяє побудувати розмічений граф станів і скласти систему рівнянь, які зв'язують між собою вірогідності сусідніх станів. Систему можна вирішувати аналітично або чисельно.

Найбільш універсальним, придатним для рішення задач практично будь якої складності, є **метод статистичного моделювання**. Математична модель процесу обслуговування при цьому реалізується в вигляді програми для ЕОМ. Моделювання дозволяє отримати чисельні характеристики якості обслуговування при конкретних параметрах потоку, СМО та заданій дисципліні обслуговування. Результати моделювання використовують для перевірки гіпотез і припущень, уточнення емпіричних коефіцієнтів. При моделюванні отримують приблизну оцінку характеристик якості обслуговування.

Таким чином, вивчивши основні методи ТМО, ви зможете розрахувати характеристики якості обслуговування в інформаційних системах, управляти основними параметрами якості обслуговування реальних систем та вимірювати їх, а також запропонувати оптимальні з точки зору якості обслуговування технічні рішення при проектуванні нових мереж і систем.

1. Поняття системи масового обслуговування. Приклади СМО.

Термін “система масового обслуговування (СМО)” об’єднує досить широке коло складних систем на виробництві, транспорті, сфері обслуговування тощо.

Функціональна схема СМО (рис. 1) відбиває загальні принципи обслуговування вхідного потоку викликів ***v***-канальною СМО. Виклики з деякого потоку надходять на обслуговування в СМО, займаючи вільні канали. При зайнятості всіх каналів можливе очікування викликів у черзі. Після завершення обслуговування виклики звільняють канали і лишають систему.

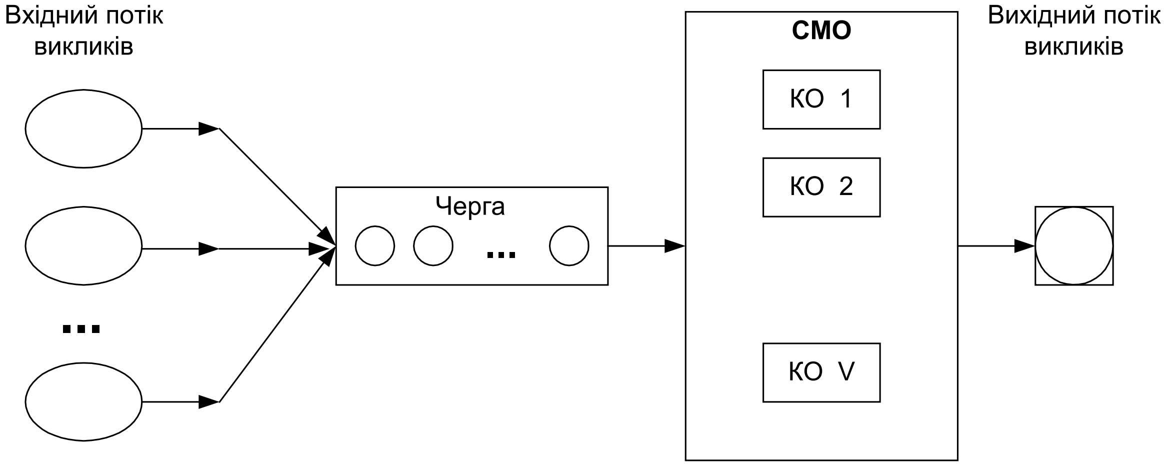

Рис. 1. Функціональна схема СМО (КО – канал обслуговування)

Під таку схему підпадають наступні приклади з різних галузей економіки: обслуговування покупців в супермаркеті, заправка машин на АЗС, надання злітно-посадочної смуги літакам в аеропорту, контроль якості виробів на виході збирального конвеєру, ремонт та відновлення виробів та ін.

Зазвичай СМО також широко розповсюджені і в телекомунікаційних мережах: локальні (районні) та міжміські телефонні станції, комутатори та маршрутизатори обчислювальних мереж, взагалі, – будь-які комутаційні системи функціонують як СМО. Функціонування СМО в телефонних мережах має свої особливості, які і розглядає підрозділ теорії масового обслуговування – “***теорія телетрафіка***”.

Теорія телетрафіка розглядає як основу всіх процесів в телекомунікаційних системах передачу та обробку ***повідомлень (messege)***, під якими будемо розуміти одновимірне представлення інформації з відокремленим початком і кінцем. Появу в системі повідомлення будемо ототожнювати з ***вимогою (arrival)*** на його передачу або обробку. Обробка або передача кожного повідомлення займає деякий кінцевий час – ***час обслуговування (holding time)***.

Частину системи, що приймає участь в процесі передачі або обробки повідомлення так, що одночасно з ним ніяке інше повідомлення не може оброблятися цією частиною, назвемо ***каналом обслуговування (server)***. Таким чином, <u>маючи рівно один КО</u>, система в кожний момент часу здатна обслугувати не більше ніж одну вимогу. Якщо на таку систему, зайняту обслуговуванням, протягом інтервалу часу обслуговування надійде ще одна вимога, то вона не зможе отримати обслуговування.

Це найпростіший випадок ресурсного конфлікту – вимоги, що надходять одна за одною, не можуть бути обслужені негайно при надходженні або, як кажуть, в реальному масштабі часу через те, що канал обслуговування не встигає обслужити вимоги за час між їх надходженнями. Конфлікт не виникне, якщо система буде мати не один, а декілька каналів обслуговування, ввімкнених так, щоб вхідні вимоги розподілялися б для обслуговування на будь-який вільний з них в даний момент.

Проте, якщо час обробки не є безкінечно малим у порівнянні з інтервалом між надходженням вимог, то і в системі з декількома каналами може виникнути ресурсний конфлікт – вимога при надходженні не зможе отримати негайного обслуговування, оскільки усі канали можуть бути зайнятими на даний момент. В такому випадку система може просто проігнорувати вимогу (відмовити у обслуговуванні). Вона буде відкинута, а система, як кажуть, буде вважатися ***заблокованою***. Імовірність такої події є важливою характеристикою системи. Її прийнято називати ***імовірністю блокування (імовірністю відмови обслуговування) (blocking probability)***.

Для уникнення таких ситуацій, в системі може бути передбачений спеціальний буфер пам’яті, до якого будуть поміщатися вимоги, які не зможуть бути обслужені негайно при надходженні внаслідок зайнятості усіх каналів. Таким чином в системі організовується ***черга (qeuеue) вимог*** і така СМО розглядається як *система з чергами* (***queuing system***). В черзі може опинитися не одна, а декілька вимог, якщо кількість вимог за певний інтервал часу перевищить кількість звільнившихся за цей час каналів обслуговування. При виконанні певних умов черга не буде безкінечно зростати і усі вимоги рано чи пізно будуть обслуговані, однак час їх перебування в черзі буде різним і може розглядатися як випадкова величина. Розподіл цієї випадкової величини також є суттєвою характеристикою системи обслуговування. Часто для оцінки якості використовують тільки її середнє значення – ***середній час очікування обслуговування (average waiting time).***

Тому, недостатність ресурсів в обслуговуючій системі може призвести або до втрат вимог (заявок), або до затримки їх обслуговування.

На систему впливають кілька потоків подій. Наприклад, потік заявок на обслуговування, потік обслужених заявок, тощо.

Потік викликів (повідомлень, занять, звільнень) – множина послідовних моментів надходження викликів (повідомлень, занять, звільнень). Потік викликів називають ***детермінованим***, якщо відома послідовність моментів надходження викликів та ***випадковим*** у протилежному випадку.

Найчастіше зустрічаються випадкові потоки викликів. В подальшому будемо розглядати саме цей випадок.

2. Предмет і задачі теорії масового обслуговування.

Предметом вивчення ТМО є процеси в системах масового обслуговування, що виникають при надходженні й обслуговуванні потоків викликів, та їх кількісні характеристики.

Математична модель системи масового обслуговування (СМО) містить наступні основні елементи:

- потік викликів, що надходять, (вхідний потік викликів),
- систему обслуговування, що має певне число каналів, визначену дисципліну обслуговування та забезпечує необхідні характеристики якості.

Так наприклад, для телекомунікаційних систем поняття ***вхідного потоку*** викликів включає інформацію про модель потоку викликів (вимог на з'єднання), закон розподілу тривалості обслуговування (передачі) повідомлення, множину адрес джерел і приймачів повідомлень, а також тип каналу, який потрібен для передачі повідомлень, та спосіб передачі – аналоговий чи дискретний. ***Система обслуговування*** характеризується структурою побудови (кількість каналів, фазність, доступність) та набором структурних параметрів. Під ***дисципліною обслуговування*** викликів розуміють: спосіб обслуговування (з наявними втратами, очікуванням, повтором чи комбінований), порядок обслуговування (за чергою, у випадковому порядку чи з пріоритетом), а також іншу інформацію, яка характеризує взаємодію потоку повідомлень з системою обслуговування. До ***характеристик якості*** обслуговування відносять імовірність явної або умовної втрати повідомлення, середній час затримки повідомлення, середня довжина черги, імовірність втрати первинного або повторного виклику, інтенсивність обслугованого навантаження і т.п.

При дослідженні СМО можуть вирішуватися такі задачі:

- ***задачі аналізу СМО*** – визначення, які характеристики якості забезпечить СМО із заданою структурою і параметрами для відомого вхідного потоку викликів;
- ***задачі параметричного синтезу*** – визначення, які параметри системи обслуговування з фіксованою структурою треба встановити, щоб забезпечити для відомого вхідного потоку викликів задану якість обслуговування;
- ***задачі структурного синтезу*** – визначення, яку структуру системи обслуговування з фіксованими параметрами треба встановити, щоб забезпечити для відомого вхідного потоку викликів задану якість обслуговування;
- ***задачі оптимізації СМО*** – визначення, яку структуру і параметри системи треба встановити, щоб забезпечити мінімальну вартість її функціонування при підтримці заданої якості обслуговування.

В телекомунікаційних системах заявка асоціюється, перш за все, зі спробою абонента отримати доступ до ресурсів мережі для передачі або прийому повідомлень. Наприклад, знімаючи слухавку телефонного апарату, абонент телефонної мережі породжує сигнал, який і є заявкою на обслуговування його цією мережею. Якщо після зняття слухавки абонент не чує сигналу, то мережа відмовляє йому в обслуговуванні, заявка відкидається. Наявність гудка означає, що заявка прийнята й абонент отримає обслуговування.

Абонент телефонної мережі створює заявки й іншого типу – набір номера абонента, що викликається. Ця заявка також може бути виконана, якщо ресурси мережі дозволять установити з’єднання між усіма телефонними станціями, які забезпечують передачу мовного сигналу між телефонами цих абонентів. Однак, це відбувається не завжди. Заявка на встановлення з’єднання може бути не задоволена. Імовірність такої події для абонентів телефонної мережі розглядається як характеристика якості обслуговування цієї мережі. Тому розрахунок імовірності відмови в обслуговуванні є прикладом рішення задачі аналізу теорії телетрафіка.

Розглянемо тепер користувачів комп’ютерної мережі Internet. В цьому випадку якість мережі асоціюється з іншою характеристикою: часом доступу до того чи іншого ресурсу. Причиною уповільнення роботи є черги пакетів, що утворюються в маршрутизаторах. В комп’ютерних (пакетних) мережах заявка на передачу інформації від одного вузла до іншого навіть у випадку нестачі мережного ресурсу, як правило, не відкидається, а ставиться в чергу на очікування звільнення необхідного ресурсу. Тому характеристикою якості обслуговування в цьому випадку вважають час очікування в черзі на обслуговування. Задача розрахунку середнього часу очікування також вирішується в рамках рішення задачі аналізу.

Прикладом задачі параметричного синтезу може бути визначення необхідної швидкодії маршрутизатора для забезпечення нормованого часу затримки обробки пакетів заданої довжини.

При структурному синтезі, наприклад машинної АТС, визначається кількість комутаційних приладів на кожному кроці пошуку для забезпечення нормованої імовірності втрати виклику.

Характеристики СМО.

В залежності від типу СМО можуть застосовуватись різні показники їх ефективності. Наприклад, для СМО з відмовами найважливішою характеристикою є абсолютна пропускна спроможність – середнє число заявок, що може обслужити система за одиницю часу.

Разом з цим, часто розглядається відносна пропускна спроможність – відношення середньої долі заявок, що обслуговується системою за одиницю часу, до середнього числа заявок, що надійшли за цей час.

Інші характеристики для СМО з відмовами:

- середнє число зайнятих каналів;
- середній відносний час простою системи або окремого каналу;
- імовірність відмови в обслуговуванні заявки, тощо.

Для СМО з необмеженим очікуванням важливими показниками є:

- середнє число заявок у черзі;
- середнє число заявок в системі (в черзі та на обслуговуванні);
- середній час очікування в черзі;
- середній час перебування заявки в системі, тощо.

Для СМО з обмеженим очікуванням важливими є всі перелічені вище показники ефективності обслуговування.

Для аналізу СМО необхідно знати:
- число каналів ν;
- інтенсивність потоку заявок λ;
- продуктивність кожного каналу (середнє число заявок μ, що можуть бути обслужені за одиницю часу;
- умови організації черги.

В подальшому всі характеристики ефективності будуть виражатись через вищезазначені показники.

3. Класифікація СМО. Символіка Кендала – Башаріна

Оскільки модель СМО поєднує в собі сукупність багатьох елементів, класифікацію СМО можна здійснювати за різними ознаками. Для забезпечення універсального підходу в позначенні різних типів СМО та взаєморозуміння інженерів та науковців різних країн, які досліджують ці системи, була розроблена наступна класифікація СМО, відома під назвою "Символіка Кендала – Башаріна" (за прізвищами вчених, що її запропонували).

Символіка складається з шести позицій, що розділяються слешами:

1/2/3/4/5/6

Перша позиція – тип потоку, що надходить:
M – найпростіший потік
M_t – пуассонівський потік із змінним параметром (залежить від часу)
M_r – пуассонівський потік з умовним параметром
M_i – примітивний потік
D – детермінований (невипадковий) потік (Determinate)
E^n – потік Ерланга n-ого порядку
Ge – довільний потік (General)

Друга позиція – закон розподілу часу обслуговування виклику
M – експоненціальний
D – детермінований
G – довільний

Третя позиція – структура СМО
V – число каналів

G – неповнодоступні канали обслуговування (тобто існує алгоритм, що визначає, які канали доступні яким заявкам). Якщо не вказано, то усі канали обслуговування доступні усім викликам.

LS – багатофазна система (Link System), якщо не вказано, то це – однофазна система, де заявка проходить тільки одну фазу обслуговування в деякому каналі.

Четверта позиція – дисципліна або спосіб обслуговування

LL – без втрат (Loss Less)

L – з втратами (Loss)

W – з очікуванням (чергою) (Wait)

R – з повторенням (Reatempt)

WL – з умовними втратами (комбінований).

П'ята позиція – тип черги

I – індивідуальна, якщо не вказано – загальна черга до усіх каналів обслуговування

SP – рівно імовірна (Same Probability)

FF – демократична (FIFO)

LF – стекова (LIFO)

PR – з пріоритетом (Priority)

1) PRR – відносний (Relative) – заявка чекає звільнення каналу

2) PRA – абсолютний (Absolute) – заявка перериває обслуговування і займає канал

Шоста позиція – спосіб заняття каналу

S – послідовне (Sequential)

R – випадкове (Random)

Приклад 1.

M/M/10/L//R – означає 10-канальну систему з втратами, на вхід якої надходить найпростіший потік викликів, час обслуговування розподілено за експоненціальним законом, канали займаються випадково.

Приклад 2.

M/G/5/W/FF/ – означає 5-канальну систему з очікуванням, на вхід якої надходить найпростіший потік викликів, час обслуговування розподілено за довільним законом, черга демократична.

Приклад 3.

M_i/M/20/LL//S – означає 20-канальну систему без втрат, на вхід якої надходить примітивний потік викликів, час обслуговування розподілено за експоненціальним законом, канали займаються послідовно.

Приклад 4.

M/M/5/R– означає 5-канальну систему з повторенням, яка обслуговує найпростіший потік викликів, час обслуговування розподілено за експоненціальним законом.

4. Одноканальна СМО з відмовами.

Розглянемо найпростішу задачу теорії масового обслуговування – задачу про функціонування одноканальної СМО з відмовами.

Нехай на СМО поступає найпростіший потік заявок з інтенсивністю λ(t). В загальному випадку вона залежить від часу:

$$\lambda = \lambda(t) \tag{1}$$

Якщо заявка поступила в момент часу, коли канал зайнятий, то вона отримує відмову в обслуговуванні та покидає систему.

Будемо вважати, що обслуговування заявки продовжується в середньому Тоб. Цей час розподілений за показниковим законом з параметром μ:

$$f(t) = \mu e^{-\mu t} \tag{2}$$

Отже, потік обслуговування – найпростіший, його інтенсивність μ.

Необхідно знайти:

1) абсолютну пропускну спроможність СМО (A);

2) відносну пропускну спроможність СМО (q).

Представимо, що система може знаходитись у двох станах:

S0 – канал вільний;

S1 – канал зайнятий.

Граф станів та переходів системи представлено на рис. 2.

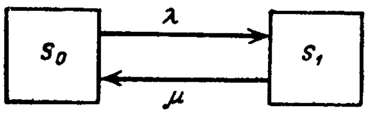

Рис. 2

Із стануS0 в стан S1 систему переводить потік заявок з інтенсивністю λ, а із S1 в S0 – потік обслуговування з інтенсивністю μ.

Імовірності станів позначимо $p_0(t)$ та $p_1(t)$:

$$p_0(t) + p_1(t) = 1. \tag{3}$$

Рівняння Колмогорова:

$$\left.\begin{aligned} \frac{dp_0}{dt} &= -\lambda p_0 + \mu p_1, \\ \frac{dp_1}{dt} &= -\mu p_1 + \lambda p_0. \end{aligned}\right\} \tag{4}$$

Початкові умови: $p_0(0)=1$; $p_1(0)=0$.

Якщо виразити із першого рівняння (4) $p_1(t)$ та підставити до (3), то отримаємо:

$$\frac{dp_0(t)}{dt} = -(\mu+\lambda)\cdot p_0(t) + \mu . \tag{5}$$

Рішенням цього рівняння для λ=const:

$$p_0(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\cdot e^{-(\lambda+\mu)t} . \tag{6}$$

Графіки $p_0(t)$ та $p_1(t)$ представлені на рис. 3.

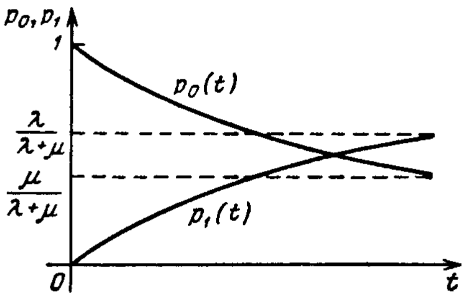

Рис. 3

Граничні імовірності (при $t \to \infty$):

$$p_0 = \frac{\mu}{\lambda + \mu}; \qquad p_1 = \frac{\lambda}{\lambda + \mu}.$$

Тепер можна зробити висновок, що для одноканальної СМО з відмовами імовірність p_0 дорівнює відносній пропускній здатності q.

Це пояснюється тим, що імовірність p_0 – це ймовірність того, що заявка, що надійшла до системи буде обслужена. Або з іншої сторони, p_0 – це відношення числа заявок, що надійшли, до числа заявок, що були обслужені, тобто q:

$$q = \frac{\mu}{\lambda + \mu}. \tag{7}$$

Тоді абсолютна пропускна здатність:

$$A = \lambda \cdot q = \frac{\lambda\mu}{\lambda + \mu}. \tag{8}$$

Ймовірність відмови в обслуговуванні:

$$P_{ВІДМ} = 1 - p_0 = \frac{\lambda}{\lambda + \mu} \tag{9}$$

Таким чином, на основі рішення системи рівнянь Колмогорова знайдені всі показники одноканальної СМО з відмовами.

5. Багатоканальна СМО з відмовами.

Розглянемо n-канальну СМО з відмовами. Будемо нумерувати стани системи по числу зайнятих каналів. Стани системи:

S_0 - всі канали вільні,

S_1 – зайнятий рівно один канал, інші вільні,

. . . .

S_k - зайняті рівно k каналів, інші вільні,

. . . .

S_n – зайняті всі n каналів.

Граф станів СМО представлений на рис. 4. Розмітимо граф, тобто проставимо біля стрілок інтенсивності відповідних потоків подій.

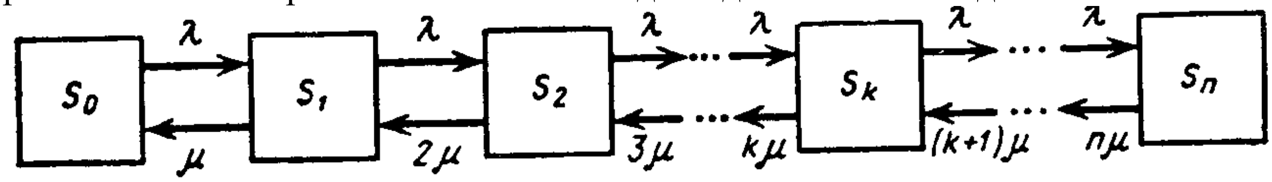

Рис. 4

По стрілкам з л і в а н а п р а в о систему переводить один і той же потік – потік заявок з інтенсивністю λ. Якщо система знаходиться в стані S_k (зайнято k каналів) і прийшла нова заявка, система переходить (перескакує) в стан S_{k+1}.

Визначимо інтенсивності потоків подій, які переводять систему по стрілкам
с п р а в а н а л і в о.

Нехай система знаходиться в стані S_1 (зайнятий один канал). Тоді, як тільки закінчиться обслуговування заявки, яка займає цей канал, система перейде в S_0; значить, потік подій, який переводить систему по стрілці $S_1 \rightarrow S_0$, має інтенсивність μ. Очевидно, якщо обслуговуванням зайнято два канали, а не один, потік обслуговувань, який переводить систему по стрілці $S_2 \rightarrow S_1$, буде вдвічі інтенсивнішим (2μ); якщо зайнято k каналів – в k разів інтенсивнішим (kμ). Проставимо відповідні інтенсивності біля стрілок, які ведуть справа наліво.

З рис. 1 видно, що процес, який протікає в СМО, являє собою окремий випадок п р о ц е с у з а г и б е л і т а р о з м н о ж е н н я, розглянутого в лекції 3.

Користуючись загальними правилами, можна скласти рівняння Колмогорова для імовірностей станів:

$$\left.\begin{aligned}
&\frac{dp_0}{dt} = -\lambda p_0 + \mu p_1,\\
&\frac{dp_1}{dt} = -(\lambda + \mu)\, p_1 + \lambda\, p_0 + 2\mu p_2,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_k}{dt} = -(\lambda + k\mu)\, p_k + \lambda p_{k-1} + (k+1)\,\mu p_{k+1},\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_n}{dt} = -n\mu p_n + \lambda p_{n-1}.
\end{aligned}\right\} \tag{1.1}$$

Рівняння (1.1) називаються р і в н я н н я м и   Е р л а н г а.

Природними початковими умовами для їх рішення є:

$$p_0(0) = 1;\ p_1(0) = p_2(0) = \cdots = p_n(0) = 0$$

(в початковий момент система вільна).

Інтегрування системи рівнянь (1.1) в аналітичному вигляді доволі складне; на практиці такі системи диференційних рівнянь зазвичай вирішуються чисельно на ПЕОМ. Таке рішення дає нам всі імовірності станів

$$p_0(t), \qquad p_1(t), \dots, p_k(t), \dots, p_n(t)$$

як функції часу.

Звісно, нас більше за все будуть цікавити г р а н и ч н і   і м о в і р н о с т і с т а н і в $p_0(t),\ p_1(t), \dots, p_k(t), \dots, p_n(t)$, що характеризують усталений режим роботи СМО (при $t \to \infty$). Для знаходження граничних імовірностей скористаємось вже готовим рішенням задачі, отриманим для схеми загибелі та розмноження (лекція № 3). Згідно цього рішення

$$\left.\begin{aligned}
&p_k = \frac{\lambda^k}{\mu\cdot 2\mu \dots k\mu}\, p_0 = \frac{(\lambda/\mu)^k}{k!}\, p_0, \quad (k = 1, 2, \dots, n);\\
&p_0 = \frac{1}{1 + \frac{\lambda/\mu}{1!} + \frac{(\lambda/\mu)^2}{2!} + \dots \frac{(\lambda/\mu)^n}{n!}}.
\end{aligned}\right\} \tag{1.2}$$

В цих формулах інтенсивність потоку заявок λ та інтенсивність потоку обслуговувань (для одного каналу) μ не фігурують окремо, а входять тільки своїм відношенням λ/μ. Позначимо це відношення

$$\lambda/\mu = \rho$$

і будемо називати величину ρ «приведеною інтенсивністю» потоку заявок. Фізичний зміст її такий: <u>величина ρ являє собою середню кількість заявок, які приходять в СМО за середній час обслуговування однієї заявки</u>.

З урахуванням цього позначення, формули (1.2) приймуть вигляд:

$$\left.\begin{aligned}
&p_k = \frac{\rho^k}{k!}\, p_0, \quad (k = 1, 2, \dots, n);\\
&p_0 = \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!}} = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!}\right]^{-1}.
\end{aligned}\right\} \tag{1.3}$$

Формули (1.3) називаються ф о р м у л а м и   Е р л а н г а. Вони виражають граничні імовірності всіх станів системи в залежності від параметрів λ, μ та n

(λ - інтенсивність потоку заявок, μ - інтенсивність потоку обслуговувань, n – кількість каналів СМО).

Знаючи всі імовірності станів

$$p_0, p_1, \dots, p_k, \dots, p_n,$$

можна знайти **характеристики ефективності СМО**:

відносну пропускну здатність q (відношення середньої долі заявок, що обслуговується системою за одиницю часу, до середнього числа заявок, що надійшли за цей час);

абсолютну пропускну здатність А (середнє число заявок, що може обслужити система за одиницю часу);

імовірність відмови $P_{\text{відм}}$ (імовірність того, що заявка надійде до системи та не буде обслуженою в наслідок зайнятості всіх каналів).

Дійсно, заявка отримує відмову, якщо проходить в момент, коли всі n каналів зайняті. Імовірність цього дорівнює

$$P_{\text{відм}} = p_n = \frac{\rho^n}{n!} p_0. \quad (1.4)$$

Імовірність того, що заявка буде прийнята на обслуговування (вона ж відносна пропускна здатність q) доповнює $P_{\text{відм}}$ до одиниці:

$$q = 1 - p_n. \quad (1.5)$$

Абсолютна пропускна здатність:

$$A = \lambda q = \lambda(1 - p_n). \quad (1.6)$$

Однією із важливих характеристик СМО з відмовами є **середня кількість зайнятих каналів** (в даному випадку вона співпадає із середньою кількістю заявок, які знаходяться в системі). Позначимо цю середню кількість $\bar{k}$.

Величину $\bar{k}$ можна обчислити безпосередньо через імовірності $p_0, p_1, \dots, p_n,$ за формулою:

$$\bar{k} = 0 \cdot p_0 + 1 \cdot p_1 + \dots + n \cdot p_n \quad (1.7)$$

як математичне очікування дискретної випадкової величини, приймаючої значення 0,1,…,n з імовірностями $p_0, p_1, \dots, p_n$. Однак значно простіше виразити середню кількість зайнятих каналів через абсолютно пропускну здатність А, яку ми обчислили. Дійсно, А являється *середньою кількістю заявок, які обслуговуються за одиницю часу.* Один зайнятий канал обслуговує в середньому за одиницю часу μ заявок. Тоді середня кількість зайнятих каналів буде визначатись діленням А на μ:

$$\bar{k} = \frac{A}{\mu} = \frac{\lambda(1 - p_n)}{\mu},$$

або, переходячи до позначення $\lambda/\mu = \rho$,

$$\bar{k} = \rho(1 - p_n). \quad (1.8)$$

Приклад. Трьохканальна СМО з відмовами представляє собою телефонну лінію. Виклики поступають з інтенсивністю λ=0,8 (викликів за хвилину). Середня тривалість розмови дорівнює 1,5 хвилини ($\mu = 0{,}667$). Вважається, що телефонна лінія складається з 3 ліній зв'язку. Зайти імовірності станів, абсолютну та відносну пропускну здатності, імовірність відмови та середню кількість зайнятих каналів.

Розв'язання. Наведена інтенсивність потоку заявок:

$$\rho = \frac{\lambda}{\mu} = \frac{0{,}8}{0{,}667} = 1{,}2.$$

За формулами Ерланга (1.3) отримуємо:

$$p_1 = \frac{\rho}{1!} p_0 = 1{,}2 p_0,$$

$$p_2 = \frac{\rho^2}{2!} p_0 = 0{,}72 p_0,$$

$$p_3 = \frac{\rho^3}{3!} p_0 = 0{,}288 p_0,$$

$$p_0 = \frac{1}{1 + 1{,}2 + 0{,}72 + 0{,}288} \approx 0{,}312;$$

$$p_1 \approx 1{,}2 \cdot 0{,}312 \approx 0{,}374; \quad p_2 \approx 0{,}72 \cdot 0{,}312 \approx 0{,}224;$$

$$p_3 \approx 0{,}288 \cdot 0{,}312 \approx 0{,}090.$$

Обчислюємо імовірність відмови:

$$P_{\text{відм}} = p_3 = 0{,}090.$$

Відносна та абсолютна пропускні здатності:

$$q = 1 - p_3 = 0{,}910; \qquad A = \lambda q = 0{,}8 \cdot 0{,}910 = 0{,}728.$$

Середня кількість зайнятих каналів:

$$\bar{k} = p(1 - p_3) = 1{,}2 \cdot 0{,}91 \approx 1{,}09,$$

Тобто при усталеному режимі роботи СМО в середньому буде зайнято один канал з трьох – решта два будуть простоювати. За рахунок цього досягається достатньо високий рівень ефективності обслуговування – близько 91% всіх викликів, які надійшли, буде обслужено.

Заключення.

Широкий клас технічних та організаційних систем можна представити у вигляді математичної моделі – системи масового обслуговування. При дослідженні систем масового обслуговування використовують граф станів та переходів. Дослідження математичної моделі проводиться з метою вирішення задачі аналізу (визначення показників якості обслуговування для конкретної системи) або задачі синтезу (визначення структури або параметрів системи, що має задовольняти поставленим вимогам до системи з точки зору якості обслуговування). Вирішення задач аналізу та синтезу дозволить розробити практичні рекомендації для удосконалення інформаційних систем.

Завідувач кафедри вищої математики,
математичного моделювання та фізики
кандидат фізико-математичних наук, доцент І.В. Замрій