

**О. Я. Матов**Інститут проблем реєстрації інформації НАН України  
вул. М. Шпака, 2, 03113 Київ, Україна

## Туманні обчислення та їхнє математичне моделювання

*Туманні обчислення доповнюють хмарні обчислення, територіально наближують вузли обробки та зберігання даних до користувачів. За рахунок скорочення трафіка це дозволяє уникнути безлічі проблем у традиційних хмарних інфраструктурах, які можуть виникнути в разі необхідності переміщення зеттабайтних обсягів даних. Одночасно скорочується час доставки рішень задач користувачів. Запропоновано аналітичні моделі туманних обчислень для вирахування характеристик з використанням багатьох потоків і багатьох пріоритетів заявок на рішення задач, різних дисциплін обслуговування та їхніх комбінацій з урахуванням відмов і різних дисциплін дообслуговування та накопичення в чергах на час відновлення.*

**Ключові слова:** *еталонна архітектура, інфраструктура туманних обчислень, моделі вузлів, характеристики обслуговування, відмови, пріоритетні дисципліни обслуговування.*

### Вступ

Розробник цієї відносно нової технології туманних обчислень — Міжнародний консорціум OpenFog Consortium. Інфраструктура туманних обчислень, за визначенням OpenFog Consortium, є горизонтальною системною архітектурою, яка розподіляє обчислювальні потужності, засоби зберігання, мережеві функції, засоби управління в усьому континуумі — від «речей» до хмар — з метою наблизити ці ресурси до кінцевих користувачів і прискорити процеси прийняття рішень. Еталонна архітектура OpenFog Reference Architecture (OpenFog RA) туманних обчислень націлена на певний клас бізнес-задач, для яких хмарні структури, або інтелектуальні граничні пристрої, самі по собі недостатньо ефективні. Вона доповнює традиційну модель хмарних обчислень, забезпечуючи виконання властивих їм функцій на різних рівнях мережевої топології зі збереженням таких технологічних переваг як віртуалізація, контейнеризація, оркестровка, керованість, ефективність. Коротко розглянуто вісім основних технологічних принципів (Pillars) архітектури OpenFog RA, що характеризують приналежність систем до класу OpenFog:

безпека, масштабованість, відкритість, автономність, програмованість, працездатність і ремонтпридатність, адаптивність, ієрархічність.

Розробка математичних моделей туманних обчислень є важливим напрямком для виявлення та покращення їхніх характеристик. Стохастичний характер головних чинників і необхідність кількісної оцінки масових процесів на основі теорії імовірності обумовлює використання теорії масового обслуговування. Розглянуті моделі функціонування вузлів (*Fog Nodes*) з адаптацією до відмов характеризуються наступними параметрами: вхідним потоком заявок і потоком відмов вузлів; дисципліною обслуговування заявок; дисципліною поновлення обслуговування заявок після відновлення вузла, що відмовив; дисципліною прийому заявок до черги під час цього відновлення. Сполучення однієї із дисциплін обслуговування з однією із дисциплін поновлення обслуговування після відновлення вузла, що відмовив, і поводженням заявки, обслуговування якої було перервано відмовленням, задає умови самостійних моделей. Таких моделей розглянуто шість. У моделях використано довільний закон розподілу випадкових величин обслуговування та відновлення вузла, що надає додаткові можливості при дослідженні конкретних туманних обчислень. Розглянуті моделі придатні для аналізу не тільки туманних але й граничних (периферійних, крайніх) вузлів, а також *логічно ізольованих виконавчих середовищ різних користувачів* туманних і граничних вузлів. Моделі базуються на роботах [3–5].

### Концепція туманних обчислень

Концепція *туманних обчислень* (Fog Computing) була сформульована близько десяти років тому компанією Cisco Systems і передбачала розширення хмарних обчислень до кордонів мережі. Туманні обчислення доповнювали централізовану хмарну модель. Основний сенс туманних обчислень — наближення функціональних вузлів до користувачів. Наближення обробки та зберігання даних до кінцевих споживачів дозволяє вирішувати багато завдань, що виникають при експоненційному зростанні числа підключених до мережі пристроїв.

У 2015 році був створений міжнародний консорціум OpenFog Consortium. Серед засновників OpenFog Consortium — ARM Holdings, Cisco, Dell, Intel, Microsoft і Принстонський університет. Зараз до його складу входять близько 60 організацій.

У лютому 2017 року консорціум запропонував еталонну архітектуру туманних обчислень: OpenFog Reference Architecture [1]. А в червні 2018-го асоціація IEEE Standards Association прийняла цю архітектуру за офіційний стандарт.

Новий стандарт IEEE 1934 регламентує застосування еталонної архітектури як універсальної технологічної платформи для підтримки додатків, що вимагають обробки величезних масивів даних, зокрема задач Інтернету речей (Internet of Things, IoT), промислового Інтернету (Industrial Internet of Things, IIoT), штучного інтелекту, мереж 5G і ряду інших сучасних технологій [1].

На думку учасників консорціуму, цифрові інновації сучасного світу — IoT, штучний інтелект, віртуальна реальність, тактильний Інтернет, мережі 5G — здатні докорінно змінити виробничі процеси, бізнес і життя людей. Однак переміщення зеттабайтних обсягів даних, що генеруються підключеними підприємствами,

будівлями, лікарнями, автомобілями, може привести до виникнення безлічі проблем у традиційних хмарних інфраструктурах.

Передача в хмари величезних масивів даних, їхня обробка, формування керуючих впливів і їхня доставка за розумний час вимагають дуже високої продуктивності від хмарних ресурсів і найширшої смуги пропускання від мережевої інфраструктури. Однак побудова, підтримка і розвиток подібних систем пов'язані з гігантськими витратами.

## Архітектура туманних обчислень

Еталонна архітектура туманних обчислень доповнює традиційну модель хмарних обчислень, забезпечуючи виконання властивих їм функцій на різних рівнях мережевої топології зі збереженням таких технологічних переваг як *віртуалізація, контейнеризація, оркестровка, керованість, ефективність*. Така архітектура усуває обмеження централізованих хмарних рішень, надаючи необхідні для конкретних завдань ресурси та канали зв'язку [1].

Інфраструктура туманних обчислень, за визначенням OpenFog Consortium, є *горизонтальною системною архітектурою*, яка розподіляє обчислювальні потужності, засоби зберігання, мережеві функції, засоби управління в усьому континуумі — від «речей» до хмар — з метою наблизити ці ресурси до кінцевих користувачів і прискорити процеси прийняття рішень.

Еталонна архітектура OpenFog Reference Architecture (OpenFog RA) туманних обчислень націлена на певний клас бізнес-задач, для яких хмарні структури або інтелектуальні граничні пристрої самі по собі недостатньо ефективні. Вона доповнює традиційну модель хмарних обчислень, забезпечуючи виконання властивих їм функцій на різних рівнях мережевої топології зі збереженням таких технологічних переваг як віртуалізація, контейнеризація, оркестровка, керованість, ефективність.

Архітектура OpenFog RA базується на восьми основних технологічних принципах (Pillars), що характеризують приналежність систем до класу OpenFog: безпека, масштабованість, відкритість, автономність, програмованість, працездатність і ремонтпридатність, адаптивність, ієрархічність [1, 2].

Архітектура OpenFog RA дозволяє створювати обчислювальні середовища, оснащені широким спектром *засобів безпеки*, які використовуються в усіх їхніх компонентах — від «речей» і пристроїв IoT до туманних структур і хмарних середовищ. До захисту інфраструктури пред'являється певний набір вимог, яких необхідно дотримуватися при розробці рішень на основі еталонної архітектури. Перш за все, це *конфіденційність, анонімність, цілісність, довіра, атестація, перевірка та вимірювання*. Відповідність перерахованим вимогам гарантує, що рішення OpenFog будуть розгортатися в безпечному обчислювальному середовищі, що забезпечує захист вузлів, мережевих компонентів, процесів управління і оркестровки.

*Вузли (Fog Nodes)* — це фізичні та логічні компоненти, які виконують обчислювальні функції у туманних мережах. Певною мірою вузли служать аналогами серверів хмарних структур. Вузли, розміщені на кордоні мережі, здійснюють управління доступом і шифрування даних. Вони повинні забезпечувати їхню конте-

кстуальну цілісність і ізоляцію, а також агрегувати конфіденційні дані, перш ніж вони будуть спрямовані на наступні рівні.

Хоча вузли і є уніфікованими компонентами туманних мереж, їхні архітектурні елементи (включаючи центральні та графічні процесори, а також програмовані логічні матриці і мережеві модулі) варіюються залежно від місця вузлів у туманній ієрархії і виконуваних ними функцій.

Вузли здатні формувати пористі структури, які здійснюють балансування навантаження, підвищують відмовостійкість і знижують обсяг хмарного трафіка.

У більш складних структурах, що підтримують сервісні моделі FaaS (Fog as a Service), ланцюжки довіри формуються від вузла до вузла та поширюються на хмарні компоненти. Так як вузли можуть динамічно створюватися та розформуватися, використовувані для цього програмні і апаратні ресурси повинні бути довіреними і атестованими.

*Масштабованість* забезпечується динамічною відповідністю технологічних можливостей туманних середовищ бізнес-вимогам з урахуванням таких факторів як робочі навантаження, продуктивність, вартісні показники. Масштабованість передбачає внесення змін в окремі вузли мережі шляхом додавання апаратного та програмного забезпечення, збільшення числа вузлів на особливо завантажених рівнях або на сусідніх з ними, або зменшення їхньої кількості в разі необхідності, а також додавання систем зберігання даних і засобів аналітики. Це дозволяє нарощувати продуктивність у туманних структурах, змінювати розміри мереж при збільшенні числа додатків, «речей» або кінцевих користувачів, розширювати функціональність засобів забезпечення надійності та безпеки. Крім того, для адекватної підтримки додатками, інфраструктурою управління і оркестровки величезної кількості підключених до мережі «речей» і об'єктів здійснюється модифікація апаратних конфігурацій компонентів вузлів, а також програмного забезпечення.

*Відкритість* є фундаментальним принципом, який сприяє формуванню масштабною екосистеми туманних обчислень, що забезпечує розгортання та підтримку платформ і додатків Інтернету речей. Вона перешкоджає обмеженню числа постачальників і переважання пропріетарних «фірмових» рішень, яке може привести до подорожчання продуктів, зниження їхньої якості і уповільнення інноваційного розвитку систем. Завдяки відкритості, вузли можна розміщувати в будь-яких сегментах туманних мереж, розширювати мережі шляхом додавання вузлів, використовувати програмно-конфігуровані вузли, які динамічно формуються.

Відкритість забезпечує інтеоперабельність, підтримує побудову компонованої інфраструктури, надання завантаженим додаткам можливість використовувати вільні ресурси, дозволяє реалізувати принцип прозорості місця розташування (Location Transparency). Прозорість гарантує вузлам їхнє розміщення на будь-якому ієрархічному рівні туманної мережі, а «речам» IoT — оптимізацію мережевих підключень і вибір найбільш зручних маршрутів доступу до обчислювальних та інших ресурсів.

*Автономність* визначається можливістю виконання вузлами туманних мереж запропонованих їм функцій в умовах відмов або відсутності підтримуючих їх зовнішніх сервісів. В архітектурі OpenFog RA цей принцип поширюється на всі рівні мережевий ієрархії. Наприклад, у разі операційної автономності централізоване прийняття рішень у хмарі не є єдиною можливою опцією. Завдяки автономній

реалізації вузлів подібні функції виконуються локальними вузлами на кордоні мережі на основі оброблюваних ними даних.

Серед інших областей, де автономність необхідна, — визначення та реєстрація підключень до мережі об'єктів, оркестровка, управління, забезпечення безпеки.

Автономність — один із найважливіших факторів підвищення економічної ефективності туманних структур, оскільки завдяки їй витрати на передачу в хмару великих обсягів необроблених даних значно знижуються.

*Програмованість*, яка повинна підтримуватись і апаратними компонентами, забезпечує високий рівень адаптованості програм, розгорнутих у туманних середовищах. Це дозволяє повністю автоматизувати повторну постановку завдань, які повинні виконуватись вузлами туманної мережі або кластерами, що складаються з декількох вузлів. Такий результат досягається завдяки застосуванню інтерфейсів програмування функцій, які описуються універсальними інтерфейсами обчислювальних або прискорювальних компонентів.

Програмованість забезпечує формування адаптивної інфраструктури, що відповідає вимогам різних сценаріїв розгортання додатків IoT, а також дозволяє оптимізувати використання наявних ресурсів, для чого, серед іншого, застосовується контейнеризація. У разі реалізації у туманних структурах сервісних моделей програмованість дозволяє логічно ізолювати виконавчі середовища різних користувачів. Вона автоматизує оновлення засобів безпеки і дозволяє швидше реагувати на виникаючі загрози.

*Працездатність і ремонтпридатність* туманних інфраструктур є невід'ємним атрибутом сучасної системної архітектури, сприяючи зменшенню часу простою, зниженню ризиків порушення безперервності бізнес-процесів і спрощенню технічної підтримки.

Технології, що необхідні для їхньої реалізації і представляють собою найважливіші компоненти архітектури OpenFog RA, охоплюють апаратні системи, програмне забезпечення та операції. Вони особливо потрібні при створенні туманних інфраструктур, які працюють в умовах промислового виробництва, несприятливого природного середовища або розгорнутих на віддалених майданчиках рішень, доступ до яких утруднений.

За допомогою архітектури OpenFog RA можливо наблизити обробку даних до джерел їхньої генерації, щоб приймати оперативні рішення відразу після того, як дані будуть перетворені в осмислений контекст. Цю властивість архітектури її розробники назвали *адаптивністю* (Agility). Адаптивність дозволяє приймати стратегічні рішення на різних рівнях ієрархії туманних структур, швидко впроваджувати інновації і здійснювати масштабування в рамках загальної інфраструктури.

Адаптивність дозволяють розробникам IoT-рішень оптимізувати розміщення додатків, які є компонентами системи прийняття рішень. Дані сенсорів і інших пристроїв генеруються у вигляді різних за обсягом масивів, які надходять в різний (не визначене заздалегідь) час і при цьому створюють досить значний мережевий трафік. Витягти з них корисну інформацію і сформуванати корисний контекст, в тому числі, рекомендації для бізнесу, в більшості випадків стає можливим тільки після агрегації, зіставлення і аналізу отриманих даних.

Такі дії можуть виконуватися на хмарному рівні, але це істотно збільшує як затримки, пов'язані з передачею великих обсягів даних на значні відстані, так і витрати на оплату каналів зв'язку.

OpenFog RA доповнює традиційні хмарні архітектури в тому числі підтримкою *ієрархічності*. Кожен рівень ієрархії забезпечує підтримку певного набору функцій, необхідних для роботи IoT-систем. Обчислювальні ресурси еталонної архітектури можуть бути представлені у вигляді компонентів логічної ієрархічної інфраструктури, відповідної вимогам розгортання комплексних «наскрізних» систем IoT. Залежно від їхнього масштабу та вирішуваних завдань ієрархія може включати в себе мережу пов'язаних між собою інтелектуальних систем, розташованих на різних фізичних або логічних рівнях архітектури. У той же час із завданнями IoT може справлятися одна фізична система, як це передбачається принципом автономності.

*Моніторинг та управління* здійснюються мікроконтролерами: вони відповідають за перевірку стану процесів, що відбуваються, формування сигналів тривоги, запуск додатків для залучення уваги персоналу або автоматичне коректування ситуації, коли спостерігаються істотні відхилення параметрів від заданих значень.

На рівні операційної підтримки виконується аналіз потоків даних телеметрії з поданням отриманих результатів. Відповідні аналітичні можливості невеликі та фокусуються, в основному, на операційних аспектах фізичного середовища, за який відповідає система IoT. На цьому рівні порівнюються дані, накопичені за невеликі періоди часу, і результати онлайн-аналізу нового трафіка.

На рівні підтримки бізнесу аналізується сукупний обсяг даних платформи IoT, що надходять з усіх її численних систем, а крім того, зберігається інформація про події, що відбулися (відповідно до правил, установлених в організації). Обробка на цьому рівні петабайтних масивів даних допомагає витягати корисні для організації відомості, оптимізувати бізнес-планування, порівнювати та підвищувати операційну ефективність процесів, використовуючи, крім іншого, методи машинного навчання.

Туманні обчислення часто називають *граничними* [2]. Розробники архітектури OpenFog Reference Architecture вважають це твердження помилковим, оскільки туманні структури працюють спільно з хмарними, а граничні (периферійні, крайні — Edge Computing) не взаємодіють з ними. На відміну від ієрархічних туманних інфраструктур, дія яких поширюється на багаторівневі розподілені комплексні рішення, що включають безліч підключених пристроїв, «речей» і хмар, граничні системи охоплюють невелику кількість рівнів. Вони передбачають розміщення виконання додатків з інтенсивною обробкою даних (наприклад, шлюзів, програмованих автоматизованих контролерів і т.п.) на кордоні мережі. Однак останні можуть значно поступатися за функціональністю вузлів туманних структур. Деякі експерти розглядають граничні обчислення як окремий випадок туманних.

Прикладом становлення одного з сегментів ринку можуть служити *безпілотні автономні автомобілі*, за управління якими відповідає безліч взаємопов'язаних вузлів. Ці вузли повинні взаємодіяти з вузлами інших машин, вузлами дорожньої інфраструктури, системами управління рухом і хмарними додатками вищого рівня. Таким чином, в результаті створюються розподілені туманні системи, які будуть охоплювати досить значні за площею території [2].

Архітектура OpenFog RA відкриває можливість побудови інфраструктури, що підтримує послуги FaaS, які допомагають спростити та прискорити розгортання туманних рішень. Як стверджують розробники, до складу FaaS увійдуть не тільки такі добре відомі сервіси як Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), але й багато інших послуг, що створені з урахуванням специфічних вимог хмарних структур [1].

### Математична постановка моделей

Розробка математичних моделей туманних обчислень чи інформаційних систем, що створені з використанням туманів, є важливим напрямком для виявлення та покращення їхніх характеристик [3–6]. Туманні обчислення (ТО), як і хмарні обчислення, є об'єктами з високим рівнем невизначеності процесу функціонування, головними чинниками якої є [3, 5]:

- вірогідність потоку запитів на обчислювальні ресурси (ОР);
- наявність необхідних ОР і випадковість часу їхнього використання клієнтами;
- випадковість відмов інфраструктури ТО та часу їхнього усунення;

Стохастичний характер головних чинників і необхідність кількісної оцінки масових процесів на основі теорії імовірності обумовлює використання теорії масового обслуговування.

Пропонуються аналітичні моделі для вирахування характеристик з використанням різних дисциплін обслуговування та їхніх комбінацій, з урахуванням відмов і різних дисциплін дообслуговування та накопичення в чергах на час відновлення обслуговуючого приладу. В моделях використано довільний закон розподілу випадкових величин обслуговування та відновлення вузла ТО, що надає додаткові можливості при дослідженні конкретних туманних обчислень. Моделі базуються на роботах [3–5].

Як раніше вже відмічалось, всі обчислювальні, регулюючі, управляючі функції у туманних мережах виконуються у вузлах, які є аналогами серверів хмарних структур. Більш того, завдяки реалізованого в ТО принципу (Pillar) автономності централізоване прийняття рішень у хмарі може здійснюватися вузлами на основі даних, що обробляються в них. Тому вузол ТО (Fog Node) в моделюючій системі масового обслуговування (СМО) розглядається як обслуговуючий прилад.

На вхід вузла, як одноканальної СМО з очікуванням, надходять  $N$  пуассонівських потоків різнотипних заявок на вирішення задач з інтенсивністю  $\lambda_i, i = 1, N$ . Потоки занумеровані в порядку убавання важливості заявок, тобто заявки  $i$ -го потоку володіють  $i$ -м пріоритетом в обслуговуванні. Час обслуговування заявок є випадковою величиною з функцією розподілу  $B_i(t)$  і двома кінцевими моментами  $b_i$  і  $b_i^{(2)}, t = 1, N$ .

Природний процес обслуговування заявок у СМО порушується відмовленнями обслуговуючого приладу. Обслуговуючий прилад ненадійний і може виходити з ладу за пуассонівським законом з параметром  $\lambda_0$ . Час відновлення приладу — випадкова величина з функцією розподілу  $B_0(t)$  і двома кінцевими моментами

$b_0$  і  $b_0^{(2)}$ . Прилад може вийти з ладу як при обслуговуванні заявок (при цьому можливі два випадки: заявки повертаються в чергу; заявки губляться), так і у вільному стані. Відмовлення обслуговуючого приладу приводять до росту черги заявок і додаткових затримок у їхньому обслуговуванні. Один із можливих способів адаптації СМО до непродуктивних відмов обчислювальних ресурсів — пріоритетний прийом заявок від різних джерел у чергу до цих ресурсів на час їхнього відволікання. Таке регулювання надходженням потоків може бути досягнуте за рахунок зворотного зв'язку вузлів з джерелами заявок.

У період відновлення обслуговуючого приладу заявки одних потоків у чергу приймаються, а інші — не приймаються. Ця умова задається матрицею-рядком коефіцієнтів  $n_i, i = 1, N$ , причому  $n_i = 1$  в тому випадку, якщо заявки  $i$ -го потоку в чергу приймаються, і  $n_i = 0$ , якщо заявки одержують відмовлення.

Після відновлення обслуговуючого приладу можливі дві дисципліни поновлення обслуговування: із заявок старшого пріоритету та із заявок, обслуговування яких було перервано відмовленням приладу (за умови, що вони не губляться під час відмовлення).

Обслуговування заявок у системі може бути організоване за правилами відносних, абсолютних, змішаних і комбінованих пріоритетів.

При відносних пріоритетах можливість заявок враховується тільки в момент їхнього призначення на обслуговування. На прилад, що звільнився, призначається заявка з найбільшим пріоритетом, і її подальше обслуговуванням не переривається іншими заявками.

Абсолютні пріоритети припускають переривання обслуговування заявок низького пріоритету заявками, що надходять на вхід системи, більш високого пріоритету. Перервані заявки в цьому випадку повертаються в чергу і очікують свого дообслуговування.

Проміжною стосовно розглянутого вище є дисципліна обслуговування з комбінованими пріоритетами. При комбінованих пріоритетах час обслуговування всіх заявок, крім заявок старшого пріоритету, розбито на два часових відрізки: на першому відрізку діє абсолютний пріоритет, на другому — відносний.

Змішані пріоритети являють собою сполучення абсолютних і відносних пріоритетів, причому для окремих заявок може бути використане безпріоритетне обслуговування.

Потрібно визначити наступні характеристики обслуговування заявок у вузлах ТО:  $w_i$  — середній час чекання початку обслуговування заявок  $i$ -го потоку в  $i$ -й черзі;  $v_i$  — середній час перебування заявок  $i$ -го потоку в системі (час відгуку системи);  $q_i$  — середнє число заявок  $i$ -го потоку в  $i$ -й черзі;  $l_i$  — середнє число заявок  $i$ -го потоку в системі.

Сполучення однієї із дисциплін обслуговування з однією із дисциплін поновлення обслуговування після відновлення приладу, що відмовив, і поведженням заявки, обслуговування якої було перервано відмовленням, задає умови самостійних задач для пріоритетних СМО наступних типів:



- систем з відносними пріоритетами та поновленням обслуговування із заявок, обслуговування яких було перервано відмовленнями (системи з відносними пріоритетами першого типу);
- систем з абсолютними пріоритетами;
- систем з відносними пріоритетами та поновленням обслуговування із заявок систем старшого пріоритету (системи з відносними пріоритетами другого типу);
- систем зі змішаними пріоритетами;
- систем з комбінованими пріоритетами;
- пріоритетних систем із втратами.

Таким чином, система масового обслуговування, що моделює обчислювальний процес функціонування вузла ТО з адаптацією до відмов, характеризується наступними параметрами: вхідним потоком заявок і потоком відмов обслуговуючого приладу; дисципліною обслуговування заявок; дисципліною поновлення обслуговування заявок після відновлення приладу, що відмовив, і дисципліною прийому заявок до черги під час цього відновлення.

### Модель системи з відносними пріоритетами першого типу

Характеристики обслуговування заявок у сталому режимі зв'язані між собою формулами Літтла, що для систем із пріоритетним прийомом заявок до черги під час відновлення приладу, що відмовив, мають наступний вигляд:

$$l_i = \lambda_i^* v_i,$$

$$q_i = \lambda_i^* w_i,$$

де  $\lambda_i^* = K_r \lambda_i (1 + n_i \rho_0)$  — інтенсивність надходження заявок  $i$ -го потоку в систему з урахуванням дисципліни прийому до чергу під час відновлення приладу, що відмовив;

$$K_r = \frac{1}{1 + \rho_0}$$

— імовірність того, що обслуговуючий прилад знаходиться у справному стані;

$$\rho_0 = \lambda_0 b_0$$

— «завантаження» системи відмовленнями.

Умовою сталого режиму в системах даного класу без втрат є  $\sum_{i=1}^N \rho_i^* < K_r$ , де

$$\rho_i^* = \lambda_i^* b_i$$

— імовірність зайнятості приладу обслуговуванням заявки  $i$ -го потоку.

Крім того, середній час перебування заявки  $i$ -го потоку в розглянутій системі —  $v_i = w_i + b_i(1 + \rho_0)$ . Тому для знаходження необхідних характеристик досить обчислити  $w_i, i = \overline{1, N}$ .

Простежимо за рухом деякої заявки, що надходить у систему,  $j = \overline{1, N}$   $j$ -го потоку, Перед тим як вона потрапить на прилад, повинно бути виконано наступне:

— закінчене довідновлення приладу (за умови, що заявки  $j$ -го потоку приймаються в чергу під час відновлення приладу, що відмовив);

— дообслугована заявка, що знаходиться на приладі (якщо він справний) або в черзі (якщо обслуговування заявки було перервано відмовленням, і прилад в цей час відновлюється);

— обслуговано всі наявні заявки з черг із номерами від 1 до  $j$ ;

— обслуговано заявки, що знову надходять з потоків з номерами від 1 до  $j - 1$ .

Для середніх тривалостей перерахованих подій рівняння рівноваги має вигляд:

$$w_j = n_j \sigma_0 + \sum_{i=1}^N \sigma_i + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \quad (1)$$

де  $\delta_0 = K_r \rho_0 \Delta_0$ ,  $\delta_i = \rho_i (1 + n_i \rho_0) \Delta_i (1 + \rho_0)$ ;

$n_j K_r \rho_0$  — імовірність того, що в момент надходження заявки  $j$ -го потоку обслуговуючий прилад відновлюється;

$\Delta_0 = \frac{b_0^{(2)}}{2b_0}$  — середній час довідновлення приладу, що відмовив;

$\rho_i (1 + n_i \rho_0)$  — імовірність того, що заявка, що надходить у систему  $j$ -го потоку, застане в ній необхідну заявку  $i$ -го потоку;

$\Delta_i = \frac{b_i^{(2)}}{2b_i}$  — середній час дообслуговування заявки  $i$ -го потоку без обліку відмовлень обслуговуючого приладу.

Після нескладних перетворень рівняння (1) отримаємо наступне рекурентне співвідношення:

$$w_j = \frac{1}{K_r - R_j} [n_j K_r \sigma_0 + \sum_{i=1}^N \rho_i (1 + n_i \rho_0) \Delta_i + \sum_{i=1}^{j-1} w_i \rho_i^*], \quad (2)$$

де  $R_j = \sum_{i=1}^j \rho_i^*$ .

Для рекурентного співвідношення вигляду

$$w_j = \frac{1}{K_r - R_j} (A_j + \sum_{i=1}^{j-1} w_i \rho_i^*)$$

неважко показати, що

$$w_j = \frac{K_r A_1 + \sum_{i=2}^j (A_i - A_{i-1})(K_r - R_{i-1})}{(K_r - R_j)(K_r - R_{j-1})}. \quad (3)$$

Використовуючи вираз (3) зі співвідношення (2), визначаємо  $w_j$  в явному вигляді:

$$w_j = \frac{1}{2(K_r - R_j)(K_r - R_{j-1})} [n_1 K_r^2 \lambda_0 b_0^{(2)} + \sum_{i=1}^N \lambda_i b_i^{(2)} (1 + n_i \rho_0) + K_r \lambda_0 b_0^{(2)} \sum_{i=2}^j (n_i - n_{i-1})(K_r - R_{i-1})]. \quad (4)$$

### Модель системи з абсолютними пріоритетами

Абсолютні пріоритети припускають переривання обслуговування заявок молодшого пріоритету заявками старшого пріоритету. Тому середній час перебування  $j$ -го потоку в системі  $v_j = w_j + u_j + b_j(1 + \rho_0)$ , де  $u_j$  — середній час усіх перерв в обслуговуванні заявки  $j$ -го потоку, що обумовлені дисципліною обслуговування.

Одержимо вираз для визначення  $u_j$ . Для цього попередньо знайдемо середню тривалість періоду зайнятості ненадійного приладу для одного  $i$ -го потоку заявок.

У сталому режимі прилад може бути вільний, коли в системі немає заявок, зайнятий обслуговуванням чи відновленням дообслуговування.

Відомо, що середній час від фіксованого моменту закінчення періоду зайнятості приладу до першого надходження у вільну систему заявки  $i$ -го потоку складає  $\frac{1}{\lambda_i^*}$ . За великий час  $T$  прилад зайнятий обслуговуванням  $T\rho_i^*$  одиниць часу, відновлення дообслуговування  $T\rho_i^*\rho_0$  одиниць часу і вільний  $T - T\rho_i^*(1 + \rho_0)$  одиниць часу. За цей час має місце в середньому  $\lambda_i^* T [1 - \rho_i^*(1 + \rho_0)]$  циклів обслуговування. Тому середній період зайнятості ненадійного приладу буде:

$$\pi_i = \frac{T\rho_i^*(1 + \rho_0)}{T[1 - \rho_i^*(1 + \rho_0)]\lambda_i^*} = \frac{b_i}{K_r - \rho_i^*}. \quad (5)$$

За час обслуговування заявки  $j$ -го потоку відбудеться  $\Lambda_{j-1} b_j$  переривань,  $\Lambda_{j-1} = \sum_{i=1}^{j-1} \lambda_i (1 + n_i \rho_0)$ . Кожне таке переривання — це період зайнятості приладу заявками з потоків з номерами від 1 до  $j - 1$ . Середній час обслуговування заявки із сумарного потоку буде

$$\sum_{i=1}^{j-1} \frac{\lambda_i^*}{\Lambda_{j-1}^*} b_i = \frac{R_{j-1}}{\Lambda_{j-1}^*},$$

де  $\Lambda_{j-1}^* = \sum_{i=1}^{j-1} \lambda_i^*$ .

Тоді відповідно до формули (5) період зайнятості приладу заявками з потоків з номерами від 1 до  $j - 1$ :

$$\pi_{j-1} = \frac{R_{j-1}}{\Lambda_{j-1}^*(K_r - R_{j-1})}, \quad (6)$$

а середній час усіх перерв в обслуговуванні заявки  $j$ -го потоку, обумовлених дисципліною обслуговування:

$$u_j = \Lambda_{j-1} b_j \pi_{j-1} = b_j \frac{R_{j-1}}{K_r (K_r - R_{j-1})}. \quad (7)$$

З урахуванням виразу (7):

$$v_j = w_j + \frac{b_j}{K_r - R_{j-1}}.$$

Визначимо  $w_j$  з рівняння рівноваги, що для системи з абсолютними пріоритетами має вигляд

$$w_j = n_j \sigma_0 + \sum_{i=1}^j \sigma_i + \eta_i + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \quad (8)$$

де, на відміну від рівняння (1), другий доданок враховує час дообслуговування заявок з потоків з номерами від 1 до  $j$ ;

$$\eta_i = \sum_{k=2}^j P_k \Delta_k (1 + \rho_0) \text{ — середній час дообслуговування заявок з черг із номерами від 2 до } j, \text{ обслуговування яких було перервано надходженням заявок старших пріоритетів;}$$

$P_k$  — імовірність того, що в черзі є заявка  $k$ -го потоку, перервана заявками із сумарного потоку, що поєднує перші  $(k-1)$  потоки. Очевидно, що

$$P_k = \lambda_k^* u_k. \quad (9)$$

Виконавши нескладні перетворення в рівнянні (8), з урахуванням виразів (7) і (9) одержимо рекурентне співвідношення для  $w_j$ :

$$w_j = \frac{1}{K_r - R_j} \left[ n_j K_r \sigma_0 + \sum_{i=1}^j \rho_i (1 + n_i \rho_0) \Delta_i + \sum_{k=2}^j \rho_k^* \frac{R_{k-1}}{K_r (K_r - R_{k-1})} \Delta_k + \sum_{i=1}^{j-1} \rho_i^* w_i \right]. \quad (10)$$

Зі співвідношення (10), використовуючи рівність (5), визначаємо  $w_j$  в явному вигляді:

$$w_j = \frac{K_r}{2(K_r - R_j)(K_r - R_{j-1})} \times \\ \times [n_1 K_r^2 \lambda_0 b_0^{(2)} + \sum_{i=1}^j \lambda_i b_i^{(2)} (1 + n_i \rho_0) + K_r \lambda_0 b_0^{(2)} \sum_{i=2}^j (n_i - n_{i-1})(K_r - R_{i-1})].$$

### Модель системи з відносними пріоритетами другого типу

У розглянутій системі переривання обслуговування заявок молодшого пріоритету заявками старшого пріоритету обумовлені дисципліною поновлення обслуговування після відновлення приладу, що відмовив. Тому середній час перебування заявки  $j$ -го потоку в системі

$$v_j = w_j + u_j^* + b_j(1 + \rho_0),$$

де  $u_j^*$  — середній час усіх перерв в обслуговуванні заявки  $j$ -го потоку, обумовлених дисципліною поновлення обслуговування після відновлення приладу, що відмовив.

Для визначення  $u_j^*$  відзначимо, що, на відміну від системи з абсолютними пріоритетами заявки старших пріоритетів, що надійшли в систему під час обслуговування заявки молодшого пріоритету, замінять її на приладі лише в тому випадку, якщо обслуговування останньої перерветься відмовленням. Отже,

$$u_j^* = P_{ОТК}(B_j)u_j,$$

де  $P_{ОТК}(b_j)$  — імовірність переривання відмовленням процесу обслуговування заявки  $j$ -го потоку.

Для перебування  $P_{ОТК}(b_j)$  визначимо середній час перебування заявки  $j$ -го потоку на приладі до першого переривання відмовленням її обслуговування ( $b_{cpj}$ ), тому що ці величини зв'язані між собою наступним співвідношенням:

$$b_{cpj} = b_j[1 - P_{ОТК}(b_j)].$$

Функція розподілу часу перебування заявки  $j$ -го потоку на приладі має вигляд

$$P_j(t) = [1 - B_j(t)]e^{-\lambda_0 t},$$

де  $[1 - B_j(t)]$  — імовірність того, що за час  $t$  заявка  $j$ -го потоку не буде обслугована на приладі;  $e^{-\lambda_0 t}$  — імовірність того, що за час  $t$  не відбудеться жодного відмовлення.

Тоді

$$b_{cpj} = \int_0^{\infty} [1 - B_j(t)]e^{-\lambda_0 t} dt, \quad (11)$$

$$P_{ОТК}(b_j) = 1 - \frac{\int_0^{\infty} [1 - B_j(t)]e^{-\lambda_0 t} dt}{b_j}. \quad (12)$$

Рівняння рівноваги відносно  $w_j$  для розглянутої системи має наступний вигляд:

$$w_j = n_j \sigma_0 + \sum_{i=1}^j \sigma_i + \sum_{i=j+1}^N \rho_i^* \Delta_i^* + \eta_j^0 + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \quad (13)$$

де, на відміну від виразу (8), третій доданок враховує час дообслуговування заявок з потоків з номерами від  $j + 1$  до  $N$ ;

$\Delta_i^*$  — середній час зайнятості приладу дообслуговування заявки  $i$ -го потоку та відновленням після відмовлення, що настало під час дообслуговування цієї заявки;

$$\eta_j^* = \sum_{k=2}^j P_k^* \Delta_k^* (1 + \rho_0) \text{ — середній час дообслуговування заявок з черг із номерами від 2 до } j, \text{ переривання обслуговування яких обумовлені дисципліною поновлення обслуговування після відновлення приладу, що відмовив;}$$

$P_k^* = \lambda_k^* u_k^*$  — імовірність того, що в черзі є заявка  $k$ -го потоку, яку на відновленому після відмовлення приладі замінила заявка із сумарного потоку, що поєднує перші  $(k - 1)$  потоки.

З визначення  $\Delta_i^*$  очевидно, що

$$\Delta_i^* = [1 - P_{OTK}(\Delta_i)] \Delta_i + P_{OTK}(\Delta_i) b_0,$$

де  $P_{OTK}(\Delta_i)$  — імовірність переривання відмовленням процесу дообслуговування заявки  $i$ -го потоку.

Для визначення  $P_{OTK}(\Delta_i)$  необхідно мати функцію розподілу ймовірності часу дообслуговування заявки  $i$ -го потоку на приладі. Відомо, що щільність розподілу ймовірності цього часу  $f_i(t) = \frac{1 - B_i(t)}{b_i}$ .

Тоді з урахуванням цієї формули і виразу (12):

$$P_{OTK}(\Delta_i) = 1 - \frac{\int_0^{\infty} \left[ 1 - \int_0^t \frac{1 - B_i(x)}{b_i} dx \right] e^{-\lambda_0 t} dt}{\Delta_i}.$$

Після нескладних перетворень у рівнянні (13), використовуючи формулу (3), знаходимо  $w_j$  в явному вигляді:

$$w_j = \frac{K_r}{(K_r - R_j)(K_r - R_{j-1})} \left\{ n_1 K_r \sigma_0 + K_r \sum_{i=1}^j \sigma_i + K_r \sum_{i=j+1}^N \rho_i^* \Delta_i^* + \right. \\ \left. + \sigma_0 \sum_{i=2}^j (n_i - n_{i-1})(K_r - R_{i-1}) + \frac{1}{K_r} \sum_{i=2}^j [\rho_i^* \Delta_i^* - \sigma_i \overline{P_{OTK}(b_i)}] R_{i-1} \right\},$$

$$\text{де } \overline{P_{OTK}(b_i)} = 1 - \frac{P_{OTK}(b_i)}{K_r}.$$

### Моделі систем зі змішаними пріоритетами

У системах зі змішаними пріоритетами сусідні потоки заявок у  $M$  груп, між якими діє абсолютний, а усередині кожної — відносний пріоритет в обслуговуванні. При цьому кожна  $m$ -та група потоків заявок містить у собі потоки з номерами від  $(s_{m-1} + 1)$  до  $s_m, m = \overline{1, M}$ . Можливі дві дисципліни поновлення обслуговування заявок з потоків, об'єднаних в одну групу: із заявок, обслуговування яких було перервано відмовленням приладу (система I типу), та із заявок старшого пріоритету (система II типу).

**Модель системи I типу.** Неважко помітити, що на відміну від системи з абсолютними пріоритетами заявка  $j$ -го потоку  $m$ -ї групи, що надходить на вхід системи, повинна очікувати в черзі дообслуговування заявок з потоків з номерами від 1 до  $s_m$ , і її обслуговування може бути перервано надходженням заявок старших пріоритетів у тому випадку, якщо  $j > s_1$  (при цьому сумарний потік заявок, що перериває її обслуговування, включає потоки з номерами від 1 до  $s_{m-1}$ ). Тому рівняння рівноваги відносного середнього часу очікування початку обслуговування заявок  $j$ -го потоку  $m$ -ї групи буде мати наступний вигляд:

$$w_j^{(m)} = n_j \sigma_0 + \sum_{i=1}^{s_m} \sigma_i + \eta^{(m)} + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \quad (14)$$

де  $\eta^{(m)} = \sum_{r=1}^{m-1} \sum_{k=s_{r+1}}^{s_{r+1}} \rho_k^* \frac{R_{s_r}}{K_r - R_{s_r}} \Delta_k (1 + \rho_0)$ , а середній час перебування заявки  $j$ -го потоку  $m$ -ї групи в системі:

$$v_j^{(m)} = w_j^{(m)} + u_j^{(m)} + b_j (1 + \rho_0), \quad (15)$$

де  $u_j^{(m)} = b_j \frac{R_{s_{m-1}}}{K_r (K_r - R_{s_{m-1}})}$ .

Виконавши нескладні перетворення у виразі (15), отримуємо:

$$v_j^{(m)} = w_j^{(m)} + \frac{b_j}{K_r + R_{s_{m-1}}}.$$

З рівняння (14), з урахуванням рівності (3) знаходимо вираз для  $w_j$  в явному вигляді:

$$w_j^{(m)} = \frac{K_r}{2(K_r - R_j)(K_r - R_{j-1})} [n_1 K_r^2 \lambda_0 b_0^{(2)} + \sum_{i=1}^{s_m} \lambda_i b_i^{(2)} (1 + n_i \rho_0) + K_r \lambda_0 b_0^{(2)} \sum_{i=2}^j (n_i - n_{i-1})(K_r - R_{i-1})].$$

**Модель системи II типу.** На відміну від системи I типу, переривання обслуговування заявок у розглянутій системі обумовлені як дисципліною обслуговування, так і дисципліною поновлення обслуговування після відновлення приладу, що відмовив. Тому

$$v_j^{(m)} = w_j^{(m)} + u_j^{(m)} + u_j^{*(m)} + b_i(1 + \rho_0),$$

$$\text{де } u_j^{*(m)} = b_i P_{\text{ОТК}}(b_i) \frac{\Delta R_{j-1}^{(m)}}{K_r (K_r - \Delta R_{j-1}^{(m)})}, \quad \Delta R_{j-1}^{(m)} = R_{j-1} - R_{S_{m-1}},$$

а рівняння рівноваги для  $w_j$  має наступний вигляд:

$$\begin{aligned} \omega_j^{(m)} = & n_l \sigma_0 + \sum_{i=1}^j \sigma_i + \sum_{i=j+1}^{S_m} \rho_i^* \Delta_i^* + \eta^{(m)} + \eta^{*(m)} + \\ & + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \end{aligned} \quad (16)$$

де

$$\begin{aligned} \eta_j^{*(m)} = & \sum_{r=1}^{m-1} \sum_{k=S_{r-1}+2}^{S_r} P_k^{*(r)} \Delta_k (1 + \rho_0) + \sum_{k=S_{m-1}+2}^j P_k^{*(m)} \Delta_k (1 + \rho_0), \\ & P_k^{*(m)} = \lambda_k^* u_k^{*(r)}. \end{aligned}$$

Використовуючи рівняння (16) і (3), неважко одержати вираз для  $w_j^{(m)}$  в явному вигляді.

### Моделі систем з комбінованими пріоритетами

У системі з комбінованими пріоритетами час обслуговування всіх заявок, крім заявок першого потоку, розбито на два відрізки (етапи): на першому відрізку діє абсолютний пріоритет, на другому — відносний. Тривалість першого етапу обслуговування заявки  $k$ -го потоку — постійна величина:  $z_{ik}, i = 1, k - 1, k = 2, N$ . На другому етапі обслуговування заявок можливі дві дисципліни поновлення обслуговування: із заявок, обслуговування яких було перервано відмовленням приладу (система I типу), та із заявок старшого пріоритету (система II типу).

**Модель системи I типу.** Середній час перебування заявки  $j$ -го потоку в системі  $v_j = w_j + u_j + b_j(1 + \rho_0)$ , де  $u_j$  — середній час усіх переривань в обслуговуванні заявки  $j$ -го потоку, обумовлених дією абсолютного пріоритету на першому етапі її обслуговування.

Отримаємо вираз для визначення  $u_j$ . За час обслуговування заявки  $j$ -го потоку на першому етапі відбудеться  $\sum_{i=1}^{j-1} (z_{ij} - z_{i+1,j}) \Lambda_i$  переривань, причому на кож-



ному  $(z_{ij} - z_{i+1,j})$  тимчасовому відрізку обслуговування заявки, перериваючий потік поєднує  $i$  перших потоків. Використовуючи вираз (6), аналогічно системам з абсолютними пріоритетами, знаходимо  $u_j$ :

$$u_j = \sum_{i=1}^{j-1} (z_{ij} - z_{i+1,j}). \quad (17)$$

Рівняння рівноваги відносно  $w_j$  для даної системи має наступний вигляд:

$$w_j = n_j \sigma_0 + \sum_{i=1}^j \sigma_i + \sum_{i=j+1}^N P_{OTHji} \Delta_{OTHji} (1 + \rho_0) + \eta_j + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0), \quad (18)$$

де  $\eta_j = \sum_{k=2}^j \sum_{i=1}^{k-1} P_{ik} \Delta_{ik} (1 + \rho_0)$ ;

$P_{OTHji} = \lambda_i (1 + n_i \rho_0) x_{ji}$  — імовірність того, що заявка, яка надійшла до системи  $j$ -го потоку, застане в ній недообслуговану на *другому* етапі заявку  $i$ -го потоку;

$$\Delta_{OTHji} = \frac{x_{ji}^{(2)}}{2x_{ji}} \text{ — середній час дообслуговування на другому етапі заявок}$$

$i$ -го потоку щодо заявок  $j$ -го потоку;

$x_{ji}, x_{ji}^{(2)}$  — відповідно перший і другий моменти тривалості другого етапу обслуговування заявки  $i$ -го потоку щодо заявок  $j$ -го потоку;

$P_{ik}$  — імовірність того, що в черзі є заявка  $k$ -го потоку, обслуговування якої на першому етапі було перервано заявкою  $i$ -го потоку ( $i < k$ );

$$\Delta_{ik} = b_k - \frac{z_{ik}}{2} \text{ — середній час дообслуговування заявки } k\text{-го потоку, обслу-}$$

говування якої було перервано на першому етапі заявками із сумарного потоку, що поєднує перші  $i$  потоків.

Значення  $x_{ji}$  та  $x_{ji}^{(2)}$  можна знайти за наступними формулами:

$$x_{ji} = b_i - z_{ji},$$

$$x_{ji}^{(2)} = b_i^{(2)} - 2z_{ji} b_i + z_{ji}^{(2)}.$$

Визначимо  $P_{ik}$ . З урахуванням виразу (17), середній час переривань в обслуговуванні заявки  $k$ -го потоку, обумовлених перериваннями її обслуговування заявками із сумарного потоку, що поєднує перші  $i$  потоків:

$$u_{\Sigma i,k} = \sum_{n=1}^{i-1} (z_{nk} - z_{n+1,k}) \frac{R_n}{K_r(K_r - R_n)} + z_{ik} \frac{R_i}{K_r(K_r - R_i)}.$$

Тоді:

$$P_{ik} = \lambda_k^* (u_{\Sigma i,k} - u_{\Sigma(i-1),k}) = \lambda_k^* z_{ik} \frac{\rho_i(1 + n_i \rho_0)}{(K_r - R_i)(K_r - R_{i-1})}.$$

Використовуючи рівняння (3) і (18), неважко одержати  $w_j$  в явному вигляді:

$$\begin{aligned} w_j = & \frac{1}{2(K_r - R_j)(K_r - R_{j-1})} \sum_{i=1}^j [K_r^2 \lambda_0 b_0 (n_i - n_{i-1}) + \\ & + \lambda_i (1 + n_i \rho_0) b_i^{(2)} + \sum_{k=i+1}^N \lambda_k (1 + n_k \rho_0) x_{ik}^{(2)} - \\ & - \sum_{k=1}^N \lambda_k (1 + n_k \rho_0) x_{i-1,k}^{(2)} + \sum_{k=1}^{i-1} P_{ki} (2b_k - z_{ki})] (K_r - R_{i-1}). \end{aligned}$$

**Модель системи II типу.** На відміну від системи I типу, в системі, що розглядається,

$$v_j = w_j + u_j + u_j^* + b_i(1 + \rho_0),$$

де  $u_j^*$  — середній час переривань в обслуговуванні заявки  $j$ -го потоку, обумовлених дисципліною поновлення обслуговування після відновлення приладу, що відмовив, у зоні дії відносного пріоритету.

Очевидно, що

$$u_j = \sum_{i=1}^{j-2} (z_{ij} - z_{i+1,j}) \frac{\Delta R_{i+1}^{j-1}}{K_r(K_r - \Delta R_{i+1}^{j-1})} P_{OTK}(z_{ij}, z_{i+1,j}) + x_{1j} \frac{R_{j-1}}{K_r(K_r - R_{j-1})} P_{OTK}(x_{1j}),$$

де  $\Delta R_{i-1}^{j-1} = R_{j-1} - R_{i-1}$ ;  $P_{OTK}(z_{ij}; z_{i+1,j}), P_{OTK}(x_{1j})$  — відповідно ймовірності того, що на часових відрізках  $(z_{ij} - z_{i+1,j})$  та  $x_{1j}$  відбудеться відмовлення обслуговуючого приладу.

Ймовірності відмовлення  $P_{OTK}(z_{ij}, z_{i+1,j}), P_{OTK}(x_{1j})$  можна обчислити за наступними формулами:

$$\begin{aligned} P_{OTK}(z_{ij}; z_{i+1,j}) &= e^{-\lambda_0 z_{i+1,j}} - e^{-\lambda_0 z_{ij}}, \\ P_{OTK}(x_{1j}) &= 1 - \frac{\int_0^{\infty} [1 - F_{1j}(t)] e^{-\lambda_0 t} dt}{x_{1j}}, \end{aligned}$$

де  $F_{1j}(t)$  — функція розподілу тривалості другого етапу обслуговування  $j$ -го потоку щодо заявок першого потоку.

Функцію  $F_{1j}(t)$  можна визначити, використовуючи пряме перетворення Лапласа щільності розподілу тривалості другого етапу:  $f_{1j}(p) = b_i(p)e^{\tilde{z}_{1j}p}$ , де  $b_i(p)$  — пряме перетворення Лапласа функції  $\frac{d}{dt}B_j(t)$ .

Рівняння рівноваги для  $w_j$  у розглянутій системі має наступний вигляд:

$$w_j = n_j \sigma_0 + \sum_{i=1}^j \sigma_i + \sum_{i=j+1}^N P_{OTHji} \Delta_{OTHji}^* + \eta_j + \eta_j^* + \sum_{i=1}^j \lambda_i^* w_i b_i (1 + \rho_0) + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i (1 + \rho_0),$$

де  $\eta_i^* = \sum_{k=2}^j \sum_{i=1}^{k-1} P_{ik}^* \Delta_{ik}^*$  ;

$\Delta_{OTHji}^* = [1 - P_{OTK}(\Delta_{OTHji})] \Delta_{OTHji} + P_{OTK}(\Delta_{OTHji}) b_0$  — середній час зайнятості приладу дообслуговування заявки  $i$ -го потоку в зоні дії відносного пріоритету стосовно заявок  $j$ -го потоку та відновленням після відмовлення, що наступило під час цього дообслуговування;

$$P_{OTK}(\Delta_{OTHji}) = 1 - \frac{\int_0^{\infty} \left[ 1 - \int_0^t \frac{1 - F_{ij}(x)}{x_{ji}} dx \right] e^{-\lambda_0 t} dt}{\Delta_{OTHji}}$$

— імовірність відмовлення приладу в час, що залишився, дообслуговування заявки  $i$ -го потоку в зоні дії відносного пріоритету стосовно заявок  $j$ -го потоку;

$P_{ik}^*$  — імовірність того, що в черзі є заявка  $k$ -го потоку, яку на відновленому після відмовлення приладі замінила заявка  $i$ -го потоку ( $i < k$ );

$$\Delta_{ik}^* = \Delta_{OTHik}^* = \frac{x_{ik}^{(2)}}{2x_{ik}} \text{ — середній час дообслуговування заявки } k\text{-го потоку,}$$

що була замінена на відновленому після відмовлення приладі заявкою  $i$ -го потоку в зоні дії відносного пріоритету.

$P_{ik}^*$  визначається аналогічно  $P_{ik}$  для системи I типу:

$$P_{ik}^* = \lambda_k^* (u_{\Sigma i, k}^* - u_{\Sigma(i-1), k}^*) = \lambda_k^* (z_{i-1, k} - z_{ik}) \frac{\rho_i (1 + n_i \rho_0)}{K_r - \rho_i} P_{OTK}(z_{i-1, k}, z_{ik}) + P_{ik} \frac{x_{1k}}{x_{ik}} P_{OTK}(x_{1k}).$$

## Моделі пріоритетних систем з втратами

Відмінною рисою даного класу систем є втрата заявок, обслуговування яких було перервано відмовленням приладу. Тому ймовірність зайнятості приладу обслуговуванням заявки  $i$ -го потоку  $P_{cpi}^* = \lambda_i^* b_{cpi}$ , де  $b_{cpi}$  визначається з виразу (11).

Умовою сталого режиму в системах із втратами є  $\sum_{i=1}^N \rho_{cpi}^* < K_r$ .

Розглянемо наступні типи пріоритетних систем із втратами:

- систему з відносними пріоритетами;
- систему з абсолютними пріоритетами;
- систему зі змішаними пріоритетами.

**Модель системи з відносними пріоритетами.** Середній час перебування заявок  $j$ -го потоку в даній системі  $v_j = w_j + b_{cpi}$ .

Вираз для визначення  $w_j$  одержимо з рівняння

$$w_j = n_j \sigma_0 + \sum_{i=1}^N \rho_{cpi}^* + \sum_{i=1}^j \lambda_i^* w_i b_i^* + w_j \sum_{i=1}^{j-1} \lambda_i^* b_i^*,$$

де  $b_i^* = b_{cpi} + b_0 P_{отк}(b_i)$  — середній час зайнятості приладу обслуговуванням заявки  $i$ -го потоку та відновленням після відмовлення, що настало під час цього обслуговування;

$P_{отк}(b_i) = 1 - \frac{b_{cpi}}{b_i}$  — імовірність переривання відмовленням обслуговування

заявки  $i$ -го потоку.

Остаточно в явному вигляді:

$$w_j = \frac{n_j \sigma_0 + \sum_{i=1}^N \rho_{cpi}^* \Delta_i^* + \sigma_0 \sum_{i=2}^j (n_i - n_{i-1})(1 - R_{i-1}^*)}{(1 - R_j^*)(1 - R_{j-1}^*)},$$

де  $R_j^* = \sum_{i=1}^j \lambda_i^* b_i^*$ .

**Модель системи з абсолютними пріоритетами.** Використовуючи ту ж методику визначення  $w_j$  та  $v_j$ , що і для системи з абсолютними пріоритетами без втрат, можна одержати наступні результати. Середній період зайнятості приладу обслуговуванням заявки  $i$ -го потоку та відновленням після відмовлення, що наступило під час цього обслуговування:

$$\pi_i = \frac{b_i^*}{1 - \rho_i^{**}}, \text{ де } \rho_i^{**} = \lambda_i^* b_i^*.$$

Середній час усіх переривань в обслуговуванні заявки  $j$ -го потоку, обумовлених дисципліною обслуговування:

$$u_j = b_{cpj} \frac{R_{j-1}^*}{1 - R_{j-1}^*}, \quad v_j = w_j + \frac{b_{cpj}}{1 - R_{j-1}^*},$$

$$w_j = \frac{n_i \sigma_0 + \sum_{i=1}^{j-1} \rho_{cpi}^* \Delta_i^* + \sigma_0 \sum_{i=2}^j (n_i - n_{i-1})(1 - R_{i-1}^*)}{(1 - R_j^*)(1 - R_{j-1}^*)}.$$

**Модель системи зі змішаними пріоритетами.** За аналогією з однойменною системою без втрат (системою зі змішаними пріоритетами I типу) необхідні характеристики обслуговування заявок  $j$ -го потоку  $m$ -ї групи знаходять за формулами:

$$w_j^{(m)} = \frac{n_1 \sigma_0 + \sum_{i=1}^{S_m} \rho_{cpi}^* + \sigma_0 \sum_{i=1}^j (n_i - n_{i-1})(1 - R_{i-1}^*)}{(1 - R_j^*)(1 - R_{j-1}^*)},$$

$$v_j^{(m)} = w_j^{(m)} + \frac{b_{cpj}}{1 - R_{S_{m-1}}^*}.$$

## Висновки

1. Запропоновано аналітичні моделі для вирахування характеристик з використанням багатьох потоків і багатьох пріоритетів заявок на рішення задач, різних дисциплін обслуговування та їхніх комбінацій, з урахуванням відмов і різних дисциплін дообслуговування та накопичення в чергах на час відновлення обслуговуючого приладу (вузла ТО — *Fog Node*). В моделях використано довільний закон розподілу випадкових величин обслуговування та відновлення вузла ТО, що надає додаткові можливості при дослідженні конкретних туманних обчислень. Моделі базуються на роботах [3–5].

2. Розглянуто моделі, придатні для аналізу не тільки туманних але і граничних (периферійних, крайніх) вузлів, а також *логічно ізольованих виконавчих середовищ різних користувачів* туманних і граничних вузлів. Останню можливість надає програмованість реалізації у туманних і граничних структурах сервісних моделей.

3. Ефективність функціонування туманних вузлів може визначатись якістю організації у них обчислювальних процесів, які кількісно оцінюються показниками ефективності [3, 4, 6], що базуються на оцінці середнього чи максимального часу перебування задач у системі, часу відгуку обчислювальної системи, затримки щодо припустимих часових термінів і т.п. Отримані формульні вирази характеристик різних моделей придатні для оцінки ефективності туманних вузлів.

4. Деякі провайдери хмарних обчислень іноді орендують сервери за кордоном, що при великих даних може приводити до великого мережевого трафіка з відповідними часовими та ціновими наслідками для користувачів послуг.

1. OpenFog Reference Architecture for Fog Computing. OpenFog\_Reference\_Architecture\_2\_09\_17-FINAL-1.pdf. URL: <https://site.ieee.org/denver-com/files/2017/06/>

2. Алексей Чернобровцев. Стандартизация туманной инфраструктуры. *Журнал сетевых решений/LAN*. 2018. № 04. URL: <https://www.osp.ru/lan/2018/04/13054572>

3. Матов А.Я., Шпилев В.Н., Комов А.Д. и др. Организация вычислительных процессов в АСУ / под ред. А.Я. Матова. Киев, 1989. 200 с.

4. Матов О.Я., Храмова І.О. Проблеми користування і математичне моделювання хмарних обчислень для інтегрованої інформаційно-аналітичної системи державного управління. *Реєстрація, зберігання і оброб. даних*. 2010. Т. 12, № 2. С. 113–127.

5. Aleksandr Matov. Mathematical models of cloud computing with absolute-relative priorities of providing of computer resources to users in conditions of functioning features and failures // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018) P. 150–159. URL: <http://ceur-ws.org/Vol-2318/paper13.pdf>

6. Aleksandr Matov. Adaptation of cloud computing as optimization of the process of rendering services to users in the conditions of limited computing resources. Selected Papers of the XIX International Scientific and Practical Conference «Information Technologies and Security» (ITS 2019). CEUR Workshop Proceedings (ceur-ws.org). Vol-2577. P. 210–221. ISSN 1613-0073. URL: <http://ceur-ws.org/Vol-2577/paper17.pdf>

Надійшла до редакції 25.08.2021