УДК 004.056:004.738.5
DOI https://doi.org/10.32782/2663-5941/2025.4.2/03

Бачкала Б.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Тимошин Ю.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

МОДЕЛІ ЗАБЕЗПЕЧЕННЯ ВИСОКОЇ ДОСТУПНОСТІ ТА ВІДМОВОСТІЙКОСТІ РОЗПОДІЛЕНИХ СИСТЕМ ЦИФРОВОЇ ІНФРАСТРУКТУРИ

З розвитком інформаційних технологій та стрімким зростанням обсягу даних, сучасні організації все частіше надають перевагу використанню хмарних обчислень для підтримки цифрової інфраструктури. Це сприяє поширенню складних розподілених комп'ютерних систем, побудованих на основі хмарних, мультихмарних, туманних та периферійних рішень. Такі системи зазвичай функціонують в умовах можливих технічних (мережевих, апаратних та програмних збоїв), і зовнішніх впливів, а отже мають залишатися працездатними у разі відмови як окремих компонентів, так і цілих вузлів таких систем. Сучасний бізнес висуває жорсткі вимоги до надійності таких систем, оскільки навіть мінімальний час простою може призвести до значних фінансових втрат та зниження довіри з боку клієнтів. Щоб організувати безвідмовну роботу хмарних розподілених систем необхідно реалізувати моделі автоматичного аналізу параметрів стану вузлів обробки, їх відновлення та забезпечити високу доступність за рахунок реплікації даних, архітектурних рішень та резервування компонентів.

У статті проведено аналіз моделей забезпечення відмовостійкості та високої доступності сучасних розподілених систем, зокрема хмарних, мультихмарних, туманних та периферійних систем Інтернету Речей (IoT). У статті визначені та проаналізовані сучасні методи забезпечення відмовостійкості та високої доступності для різних типів розподілених систем. Надано рекомендації щодо застосування таких моделей та методів залежно від специфіки середовища та вимог щодо якості обслуговування. У статті запропоновано модель адаптивного резервування даних для периферійних систем з урахуванням мережевих відмов на основі розподілу Вейбула. Викладено результати моделювання співвідношень цільового рівня надійності системи та інтервалу резервування. Основною мотивацією проведених досліджень є систематизація підходів до підвищення надійності розподілених систем та розробка стратегії резервування, адаптованої для периферійних систем з характерними мережевими відмовами компонентів.

Ключові слова: *висока доступність, відмовостійкість, розподілені системи, хмарні обчислення, мультихмара, периферійні обчислення, IoT.*

Постановка проблеми. Попри широкий спектр існуючих інформаційних технологій, нині не існує універсального рішення, здатного забезпечити безперервну роботу розподілених систем цифрової інфраструктури для обробки прикладних задач з урахуванням багатьох умов їх експлуатації та різних внутрішніх та зовнішніх факторів технічного, організаційно-управлінського та операційного характеру. Отже, актуальним є завдання аналізу моделей забезпечення високої доступності та відмовостійкості таких систем, а також розроблення рекомендацій щодо вибору оптимальних архітектурних рішень та операційних характеристик, залежно від особливостей конкретного середовища, включаючи хмарні сервіси IaaS, та вимог щодо якості обслуговування. Для периферійних систем актуальною є проблема втрати даних та зниження продуктивності через відсутність можливості своєчасно копіювати чи реплікувати дані, а також через специфіку роботи таких систем у середовищах із обмеженим або нестабільним мережевим з'єднанням. Тому під час реалізації методів підвищення надійності периферійних систем необхідно враховувати як специфіку мережевої інфраструктури, так і особливості локальної обробки даних, резервування

каналів зв'язку, розміщення резервних ЦОД, ізоляції «зон доступності» та можливість автономної роботи периферійних пристроїв за умов періодичних мережевих чи апаратних відмов.

Аналіз останніх досліджень і публікацій. Велика кількість наукових праць присвячена аналізу та класифікації підходів до забезпечення безперервності роботи ІТ-сервісів [1, с. 76-86]. Забезпечення високої доступності зазвичай не потребує повного дублювання всіх компонентів, але включає резервування критичних вузлів, кешування важливих даних, механізми автоматичного відновлення та балансування навантаження. В свою чергу, відмовостійкість фокусується на повному уникненні простою за будь-яких одиничних відмов – тобто система продовжує працювати без перерви навіть у разі відмови одного чи кількох компонентів. У роботі [2] проаналізовані підходи до забезпечення надійності та високої доступності в середовищах хмарних обчислень, які відкривають широкі можливості для підвищення надійності та забезпечення безперервності бізнес-процесів завдяки своїй гнучкості, масштабованості та можливості оперативного відновлення роботи сервісів після відмов, що дає змогу зменшити ризики і витрати на підтримку ІТ-інфраструктури.

У статті [3, с. 1-14] запропоновано новий метод адаптивного управління хмарними ресурсами (FT-ERM) з вбудованим механізмом відмовостійкості для забезпечення високої доступності серверів та віртуальних машин (ВМ), що передбачає проактивне резервування додаткових ресурсів для попередження аварійних ситуацій шляхом завчасної міграції навантаження, показано підвищення загальної доступності серверів на 34.47% та значне зменшення надлишкових міграцій віртуальних машин порівняно з традиційними підходами. Однак динамічні характеристики хмарного середовища підвищують ймовірність виникнення додаткових помилок та відмов. У дослідженні [4, с. 75920-75940] представлено нову гібридну модель відмовостійкого планування та балансування навантаження (HFSLM). Це рішення дозволило усунути проблему відмов та нерівномірності розподілу задач між віртуальними машинами і скоротити час виконання задач на 8% порівняно з базовими алгоритмами планування.

У дослідженні [5] запропоновано нову платформу FogROS2-FT для підвищення якості обслуговування та відмовостійкості за рахунок мультихмарного розгортання обчислювальних модулів. Платформа автоматично реплікує незалежні роботизовані сервіси, направляє запити до цих реплік і повертає назад першу отриману відповідь. Авторами роботи [6, с. 10501-10511] запропоновано новий підхід для підвищення відмовостійкості промислових хмарних платформ під назвою TOLERANCER у вузлах систем Індустріального IoT (IIoT). Запропонована система здатна ідентифікувати відмови програмного чи апаратного рівня на крайових обчислювальних серверах та автоматично перерозподіляти навантаження між працездатними вузлами, гарантуючи мінімальний час простою. У статті [7] представлено метод динамічного балансування навантаження (FTSM), який віддає пріоритет обробці чутливих до затримок запитів у середовищі периферійних та туманних обчислень, враховує час виконання, критичність задач та поточне навантаження на вузлах для оптимального перенесення процесів у випадку відмов.

Технологія розподіленої обробки даних та додатків з використанням динамічно інтерпретованих метамоделей, а також використання такого прикладного середовища для міжкорпоративної обробки бізнес-об'єктів, що описані у дослідженнях [8, с. 128-138; 9, с. 137-150], дозволяють зменшити вектор параметрів управління для систем розподіленої обробки, адаптувати і покращити балансування навантаження на вузли, збільшити надійність ІТ-систем.

Питанням забезпечення відмовостійкості хмарних систем на основі оцінок метрик надійності присвячена робота [10, с. 29-31], в якій досліджено вплив метрик RPO (Recovery Point Objective) на потенційні втрати даних та RTO (Recovery Time Objective) на допустимий час простою в середовищі хмарних сервісів моделі IaaS.

В роботі [11, с. 158-161] запропоновано модель оцінки параметрів обробки інформації в розподілених системах з динамічною структурою, які реалізують неоднорідні розподілені бази даних, що дозволяє формалізувати структуру керування обробкою запитів.

Робота [12, с. 1-8] присвячена процесу моделювання аналізу ситуацій та прийняття рішень щодо функціонування вузлів розподілених комп'ютерних систем (РКС) на основі спеціального механізму стохастичних RA-мереж, що дозволяє моделювати потік даних з різними, у тому числі суттєво різними інтенсивностями, що особливо важливо для аналізу ситуацій та прийняття рішень у динаміці керування параметрами вузлів РКС.

У роботі [13, с. 161-166] розглянуто методи підвищення ефективності виявлення загроз у роз-

поділених базах даних за допомогою системи моніторингу подій. Зазначено, що система працює на основі моделі моніторингу подій різнорідних нереляційних розподілених баз даних.

У статті [14, с. 320-325] описана складна адаптивна система, яка базується на нейронній мережі та інтелектуальних агентах. Розроблена система дозволяє ідентифікувати та класифікувати атаки та запобігати їм в режимі реального часу. В роботі [15, с. 195-203] показано, що автоматичне масштабування є ключовим механізмом адаптивного управління ресурсами, що дозволяє системі своєчасно реагувати на зміни навантаження. Своєчасному виявленню відхилень, їх діагностиці, реконфігурації та відновленню під час аварійних ситуацій присвячена робота [16, с. 104-111].

В роботі [17, с. 84-91] показано, що мультихмарна архітектура дозволяє організаціям використовувати неінтегровані послуги кількох публічних хмарних платформ. Створення ефективного мультихмарного середовища потребує системного підходу до вибору оптимальних платформ і рішень – цьому питанню присвячено дослідження [18, с. 32-39] [19, с. 113-121] [20, с. 51-62], включаючи географічно розподілені архітектури [21, с. 39-40].

Для забезпечення високої доступності у туманних обчисленнях застосовуються проактивні та реактивні стратегії управління відмовами, чому присвячена робота [22, с. 415-422]. В роботі [23, с. 49-52] показано як реплікація даних та сервісів забезпечує безперервність роботи при виході з ладу окремих компонентів периферійних вузлів.

Публікації [24, с. 217-245; 25, с. 233-246] присвячені порівнянню ймовірнісних моделей для опису процесу відмови та як вони відповідають експоненціальному, Вейбуллівському, нормальному та логнормальному розподілам ймовірностей для задач надійності, включаючи моделювання для реальних об'єктів.

В роботі [26, с. 1520-1532] обговорюються питання віртуалізації мережевих функцій, що забезпечує ефективне планування хмарних ресурсів шляхом віртуалізації мережевих сервісів та програм у алгоритми та відповідне ПЗ, яке працює на стандартних серверах. Провайдер повинен забезпечувати доступність сервісів мережі паралельних віртуалізованих мережевих функцій (ВМФ).

Вище описане підтверджує, що основними факторами, які впливають на стійкість та надійність розподілених систем обробки є технічні компоненти інфраструктури ЦОД, їх конфігурація, ізоляція сегментів у просторі та працездатність LAN- та інтернет-з'єднань, ефективне балансування потоків даних всередині вузлів та між ними, спеціалізоване ПЗ для управління розподіленими системами обробки та моделі відновлення, впровадження оптимальних і адаптивних моделей управління з резервуванням ресурсів і проактивними заходами виявлення та запобігання нетипових та аварійних ситуацій для забезпечення відмовостійкості.

Постановка завдання. Цілі пропонованого дослідження полягають у тому, щоб:

(I) розробити математичну модель адаптивної стратегії резервування для периферійних систем. Побудувати аналітичні залежності узгодження цільового рівня надійності системи із інтервалом резервування з урахуванням ймовірності мережевих відмов;

(II) побудувати графіки навантаження для гібридного хмарного середовища з метою визначення параметричних діапазонів стійкої високонадійної роботи з різними законами відмов компонентів периферійної системи;

(III) надати рекомендації щодо вибору конкретних моделей залежно від специфіки середовища та вимог щодо якості обслуговування для забезпечення відмовостійкості та високої доступності розподілених систем.

Виклад основного матеріалу. Висока доступність та відмовостійкість у хмарних середовищах реалізується повним або частковим резервуванням всіх компонентів інфраструктури. Використання географічно розподілених центрів обробки даних (ЦОД) або зон доступності дозволяє захистити IT-сервіси від зовнішніх факторів (стихійних лих, перебоїв з електроенергією). Балансування навантаження є ще одним ключовим механізмом забезпечення високої доступності у хмарній інфраструктурі. Цей підхід дозволяє рівномірно розподіляти запити та інші типи навантаження між декількома фізичними або віртуальними серверами, забезпечуючи стабільність та високу продуктивність системи. Балансування може здійснюватися за допомогою програмних або апаратних засобів. У разі відмови одного із серверів балансувальник навантаження автоматично виключає його з процесу обробки та перенаправляє трафік на інші справні вузли. Важливою складовою забезпечення високої доступності у хмарі є також реалізація механізмів аварійного відновлення, які дозволяють мінімізувати показники RTO (цільовий час відновлення – допустимий час простою системи) та RPO (цільова точка

відновлення – допустимий обсяг втрати даних) при плануванні стратегії резервного копіювання. В результаті поєднання цих підходів сучасні хмарні інфраструктури досягають рівня доступності 99.99% («чотири дев'ятки»), що є стандартом для критичних бізнес-застосунків.

Порівняно з мультирегіональними рішеннями у межах однієї хмари, мультихмарний підхід забезпечує вищий рівень відмовостійкості та доступності. Цей підхід усуває залежність від одного постачальника: глобальний збій однієї платформи не спричинить повної зупинки системи. Крім того, різні хмари можуть виконувати різні завдання. Наприклад, одна платформа може обробляти високонавантажені транзакції, а інша – забезпечувати резервне зберігання даних чи аналітику, що підвищує ефективність і стійкість системи. Одним з базових механізмів забезпечення високої доступності в мультихмарних середовищах є реплікація даних між хмарами. На відміну від класичних хмарних середовищ, де реплікація обмежена дата-центрами однієї хмари, мультихмарна реплікація підтримує декілька копій даних на різних платформах. Такий підхід дозволяє підтримувати актуальність та цілісність даних у всіх хмарах одночасно, гарантуючи їх доступність навіть за умов виходу з ладу одного з постачальників. Наступним важливим компонентом високої доступності мультихмарних архітектур є автоматичне балансування навантаження між хмарними платформами. Для цього застосовуються спеціалізовані програмні рішення, такі як глобальні балансувальники навантаження та динамічні механізми розподілу трафіку. Балансування навантаження дозволяє рівномірно розподілити запити між декількома платформами, забезпечуючи оптимальне використання ресурсів та уникнення перевантаження окремих вузлів. Крім цього, балансувальники здатні виконувати автоматичне перенаправлення трафіку на резервні хмарні майданчики у разі виявлення відмов чи зниження продуктивності однієї з платформ, завдяки чому забезпечується мінімізація часу відновлення роботи сервісів. Механізм автоматичного відновлення між кількома хмарами передбачає моніторинг стану хмарних платформ і своєчасне перенесення навантаження на резервні ресурси іншої платформи у разі виявлення аварійної ситуації. Мультихмарні архітектури пропонують вищий ступінь диверсифікації ризиків, оскільки резервні компоненти розташовані у ізольованих та незалежних хмарних платформах.

У середовищі периферійних та туманних обчислень висока доступність та відмовостійкість досягається шляхом децентралізації обчислювальних ресурсів, наближенням їх до кінцевих пристроїв та зниженням затримок, що особливо важливо для систем IoT та Індустріального Інтернету речей (IIoT). Аналіз та фільтрація даних на границі зменшує витрати на зв'язок, хмарну обробку та зберігання даних. Архітектури на базі периферійних та хмарних обчислень дозволяють створювати географічно розподілені системи з низькою затримкою, визначенням місцезнаходження та підтримкою мобільності. Головною метою цього підходу є покращення якості надання послуг, а саме: зменшення затримок передачі даних, забезпечення локальної обробки даних у реальному часі та гарантування роботи сервісів навіть за нестабільного зв'язку. За рахунок географічного рознесення вузлів і наближення їх до кінцевих пристроїв досягається стійкість до регіональних збоїв – наприклад, якщо виходить з ладу центр обробки даних у певній локації, пристрої на периферії можуть тимчасово виконувати основні функції самостійно або переключатися на резервний центр в іншому регіоні.

Для забезпечення високої доступності у туманних обчисленнях застосовуються проактивні та реактивні стратегії управління відмовами, причому проактивні методи передбачають використання систем моніторингу, які збирають інформацію про стан вузлів мережі та аналізують її для виявлення потенційних проблем ще до їх виникнення. Для цього часто використовують методи статистичного аналізу, моделі машинного навчання або алгоритми штучного інтелекту, що дають змогу завчасно спрогнозувати відмову вузла або зниження його продуктивності. На основі цих прогнозів система може адаптивно перерозподіляти навантаження, мігрувати критично важливі сервіси на більш стабільні вузли або резервувати додаткові ресурси, мінімізуючи таким чином ймовірність простою та забезпечуючи високу доступність. Реактивні ж стратегії забезпечують оперативне реагування після фактичного виникнення відмови. Серед таких методів особливо ефективними є автоматичне переключення маршрутів передачі даних та швидке переналаштування роботи системи для мінімізації негативних наслідків. Ще одним напрямом забезпечення відмовостійкості периферійних та туманних обчислень є реплікація даних та сервісів, що забезпечує безперервність роботи при виході з ладу окремих компонентів. Це дозволяє

забезпечити гнучкість і високу надійність систем навіть за умов значних навантажень або частих змін у структурі мережі.

У таблиці 1 наведено узагальнені рекомендації щодо вибору моделей забезпечення відмовостійкості розподілених систем.

Для розглянутих моделей та обчислювальних середовищ спільною рисою є використання механізмів резервування ресурсів та даних для забезпечення високої надійності розподілених систем. При виборі стратегії резервування для периферійних та туманних середовищ необхідно враховувати випадковий характер відмови мережевих компонентів та вузлів системи. Далі пропонується математична модель адаптивного резервування даних на прикладі периферійних систем.

Для врахування ймовірності відмов компонентів периферійної системи обрано двопараметричний розподіл Вейбула. Цей розподіл описує різні режими відмов: при значенні параметра форми розподілу $k<1$ інтенсивність відмов зменшується з часом; при $k=1$ інтенсивність відмов є постійною; при $k>1$ інтенсивність відмов зростає з часом. Параметр масштабу розподілу (λ) визначає характерний час, протягом якого більшість компонентів системи зазнає відмов. Формально функція надійності визначає ймовірність безвідмовної роботи системи до часу t:

$$R(t)=exp\left[-\left(\frac{t}{\lambda}\right)^k\right] \quad (1)$$

де $k>0$ – параметр форми розподілу Вейбула; $\lambda>0$ – параметр масштабу розподілу Вейбула.

Адаптивне резервування передбачає динамічну зміну інтервалів між резервними копіями даних відповідно до вимог надійності системи. Нехай після чергового резервного копіювання минув час t, визначимо час Δt як додатковий інтервал часу до наступного резервування. Введемо поняття відносної надійності на основі (1), що визначає ймовірність безвідмовної роботи системи на додатковому інтервалі Δt за умови, що система вже пропрацювала час t:

$$R(t,t+\Delta t)=\frac{R(t+\Delta t)}{R(t)}=exp\left[-\left(\frac{t+\Delta t}{\lambda}\right)^k+\left(\frac{t}{\lambda}\right)^k\right] \quad (2)$$

Критерієм вибору оптимального інтервалу резервування є забезпечення відповідності ймовірності безвідмовної роботи системи цільовому рівню надійності R_{req}:

$$R(t,t+\Delta t)=R_{req} \quad (3)$$

Підставляючи вираз (2) у (3) отримаємо рівняння:

$$exp\left[-\left(\frac{t+\Delta t}{\lambda}\right)^k+\left(\frac{t}{\lambda}\right)^k\right]=R_{req} \quad (4)$$

Застосувавши логарифмічне перетворення до рівняння (4) отримаємо вираз для визначення оптимального інтервалу резервування Δt:

$$\Delta t=\lambda\left[\left(\frac{t}{\lambda}\right)^k-\ln\left(R_{req}\right)\right]^{\frac{1}{k}}-t \quad (5)$$

Формула (5) дозволяє визначити оптимальний інтервал резервування на основі поточного часу роботи системи та цільового рівня надійності.

Розглянемо приклад застосування формули (5) для систем класу Industrial IoT (IIoT) з наступними параметрами:

$$k=1.8;\ \lambda=6\ років;\ R_{req}=0.999$$

Початковий інтервал резервування визначимо за умови $t=0$:

$$\Delta t=6\times\left[-\ln(0.999)\right]^{\frac{1}{1.8}}\approx 0.13\ років\approx 47\ днів$$

Таблиця 1

Моделі забезпечення відмовостійкості розподілених систем

Середовище	Ключові характеристики	Рекомендовані моделі
Хмара	Надлишковість ресурсів; багатозонне і багаторегіональне резервування; централізоване керування та автоматичне перемикання при збоях.	FT-ERM, HFSLM
Мультихмара	Відсутність єдиної точки відмови; автоматичний перерозподіл навантаження між різними хмарами; інтегрована міжхмарна синхронізація.	FogROS2-FT
Периферія	Децентралізація обчислень; розподіл задач між сусідніми вузлами при відмові; низькі затримки обробки даних.	TOLERANCER
Туманні обчислення	Проактивне виявлення відмов; автоматичний перехід на альтернативні вузли у разі відмов.	FTSM

Визначимо інтервал резервування через 1 рік безвідмовної роботи:

$$\Delta t = 6 \times \left[\left(\frac{1}{6} \right)^{\frac{1}{1.8}} - \ln\left(0.999\right) \right]^{\frac{1}{1.8}} - 1 \approx 0.014 \; років \approx 5 \; днів$$

Цей приклад демонструє адаптивність запропонованої моделі: з часом експлуатації системи у режимі зношування ($k > 1$) інтервал резервування зменшується з 47 днів до 5 днів, забезпечуючи незмінний рівень надійності.

Для аналізу впливу параметрів розподілу Вейбула на вибір інтервалу резервування було проведено моделювання засобами мови програмування Python. На графіках рис. 1–2 показані залежності цільового рівня надійності системи від інтервалу резервування за різних значень параметрів масштабу та форми розподілу Вейбула.

При малих значеннях середнього часу до відмови (рис. 1) спостерігається швидке зниження надійності системи зі збільшенням інтервалу резервування, особливо для систем у режимі ранніх відмов. Збільшення середнього часу до відмови (рис. 2) дозволяє підтримувати високий рівень надійності при більш тривалих інтервалах резервування. Для обох графіків характерною є наявність спільної точки перетину кривих, яка відповідає однаковому значенню надійності для різних режимів відмов у момент, коли інтервал резервування дорівнює середньому часу до відмови системи.

На рис. 3 графічно визначено параметричні діапазони інтервалів резервування, що забезпечують стійку високонадійну роботу периферійних систем для режимів ранніх відмов, випадкових відмов та зношування.

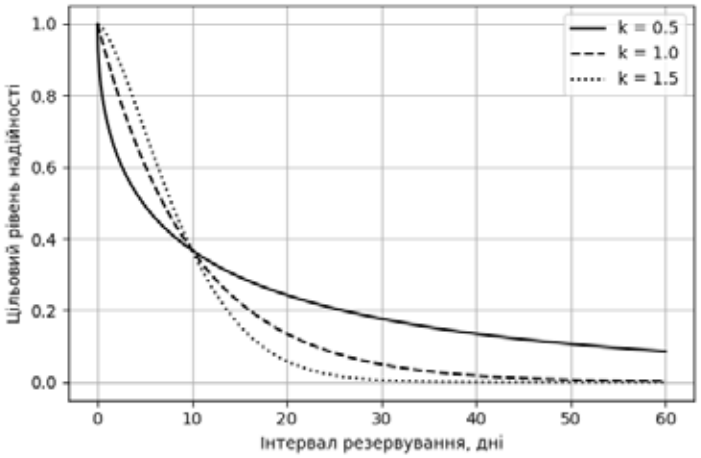

Рис. 1. Залежність цільового рівня надійності від інтервалу резервування за різних режимів відмов із середнім часом до відмови (λ) 10 днів

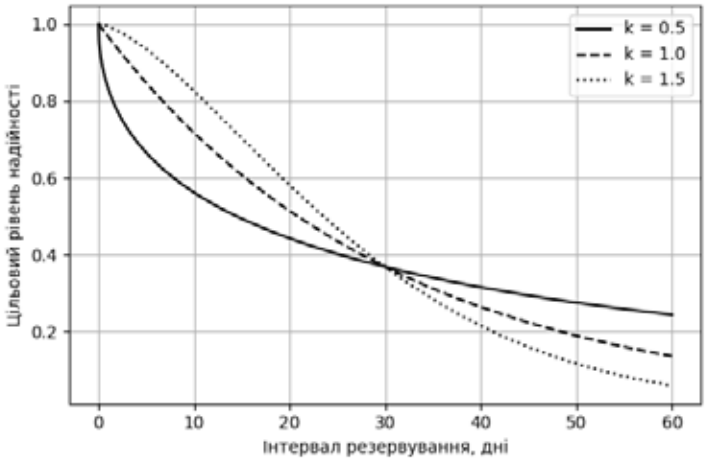

Рис. 2. Залежність цільового рівня надійності від інтервалу резервування за різних режимів відмов із середнім часом до відмови (λ) 30 днів

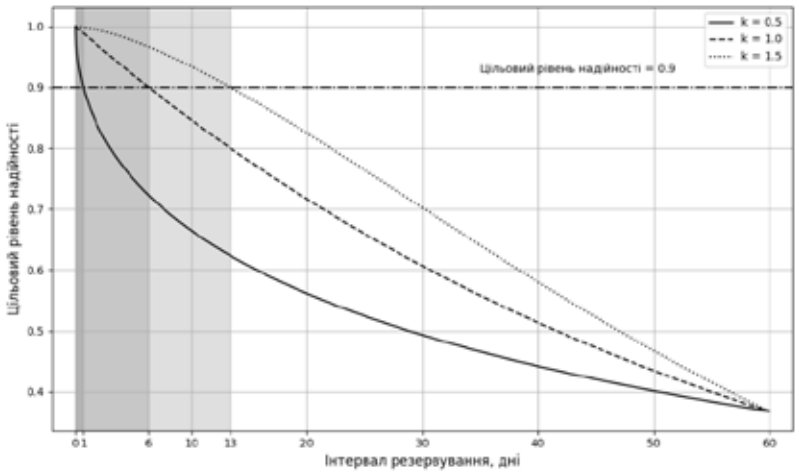

Рис. 3. Порівняння інтервалів резервування за різних режимів відмов із середнім часом до відмови (λ) 60 днів

На графіку (рис. 3) зафарбованими дільницями показано діапазони інтервалів резервування, що забезпечують цільовий рівень надійності для різних режимів відмов. Рівень надійності (0.9) було обрано відповідно до типових галузевих вимог до допустимих втрат даних у таких системах та для ілюстрації впливу різних режимів відмов. Порівнюючи значення інтервалів резервування при зменшенні ($k = 0.5$) та збільшенні ($k = 1.5$) інтенсивності відмов, стає очевидним, що системи з ранніми відмовами потребують значно коротших інтервалів резервування (до 24 годин) для підтримки цільового рівня надійності порівняно із системами, які функціонують у режимі зношування, де допустимі інтервали резервування можуть досягати 13 днів. Наведені графоаналітичні залежності можуть бути використані інженерами для оцінки допустимих меж інтервалів резервування залежно від характеру зміни інтенсивності відмов та для вибору стратегій резервування.

Висновки. У статті запропоновано та обґрунтовано комплекс моделей забезпечення високої доступності та відмовостійкості розподілених систем цифрової інфраструктури з акцентом на хмарні, мультихмарні, периферійні та туманні обчислення. Розглянуто специфіку реалізації механізмів надійності у кожному з цих середовищ, виявлено їх переваги та недоліки. Аналіз літературних джерел і практичних підходів свідчить, що ефективність таких систем досягається шляхом впровадження комплексних архітектурних рішень – резервування, реплікації даних, автоматичного масштабування, балансування навантаження та проактивного моніторингу стану компонентів.

Розроблена математична модель адаптивного резервування дозволяє дослідникам та інженерам контролювати ризики втрати даних через динамічне керування інтервалом резервування відповідно до вимог надійності у периферійних середовищах. Надані графоаналітичні залежності дозволяють враховувати вплив мережевих відмов на чисельні співвідношення частоти резервування із цільовим рівнем надійності периферійних систем.

Впровадження обґрунтованих підходів до забезпечення високої доступності та відмовостійкості є необхідною умовою для стабільного функціонування сучасної цифрової інфраструктури та підтримки критично важливих бізнес-процесів. Поєднання комплексних архітектурних рішень з адаптивними методами управління є ефективною стратегією для забезпечення надійності реальних розподілених систем.

Список літератури:

1. Panda J., Das S., Pattnaik D. Resilient IT Infrastructure: Strategies for Minimizing Downtime and Ensuring Business Continuity. *Journal of Humanities and Social Sciences Research*. 2024. Vol. 6 (S). P. 76–86. DOI: 10.37534/bp.jhssr.2024.v6.nS.id1255.p76

2. Mesbahi M. R., Rahmani A. M., Hosseinzadeh M. Reliability and high availability in cloud computing environments: a reference roadmap. *Human-centric Computing and Information Sciences*. 2018. Vol. 8, Article 20. DOI: 10.1186/s13673-018-0143-8

3. Saxena D., Gupta I., Singh A. K., Lee C.-N. A Fault Tolerant Elastic Resource Management Framework Towards High Availability of Cloud Services. *IEEE Transactions on Network and Service Management*. 2022. P. 1–14. DOI: 10.1109/TNSM.2022.3170379

4. Mushtaq S. U., Sheikh S., Idrees S. M. Next-Gen Cloud Efficiency: Fault-Tolerant Task Scheduling With Neighboring Reservations for Improved Resource Utilization. *IEEE Access*. 2024. Vol. 12. P. 75920–75938. DOI: 10.1109/ACCESS.2024.3404643

5. Chen K., Hari K., Chung T., Wang M., Tian N., Juette C., Ichnowski J., Ren L., Kubiatowicz J., Stoica I., Goldberg K. FogROS2-FT: Fault Tolerant Cloud Robotics. arXiv preprint arXiv:2412.05408 [cs.RO], 2024. DOI: 10.48550/arXiv.2412.05408

6. Al-Dulaimy A., Sicari C., Papadopoulos A. V., Galletta A., Villari M., Ashjaei M. TOLERANCER: A Fault Tolerance Approach for Cloud Manufacturing Environments. *IEEE Transactions on Industrial Informatics*. 2023. Vol. 19, no. 10. P. 10501–10511. DOI: 10.1109/ETFA52439.2022.9921606

7. Alarifi A., Abdelsamie F., Amoon M. A fault-tolerant aware scheduling method for fog-cloud environments. *PLOS ONE*. 2019. Vol. 14, No. 10. e0223902. DOI: 10.1371/journal.pone.0223902

8. Тимошин Ю. А. Технологія розподіленої обробки даних і додатків з використанням динамічно інтерпретованих метамоделей. *Адаптивні системи автоматичного управління*. 2014. № 24. С. 128–138. DOI: 10.20535/1560-8956.24.2014.38199.

9. Стенін О. А., Тимошин Ю. А., Шемседінов Т. Г. Динамічне прикладне середовище з інтерпретацією метамоделей для міжкорпоративної обробки бізнес-об'єктів. *Адаптивні системи автоматичного управління*. 2016. № 28. С. 137–150. DOI: 10.20535/1560-8956.28.2016.82401.

10. Тимошин Ю., Бачкала Б. Проблеми забезпечення відмовостійкості хмарних систем на основі оцінок метрик надійності. *Матеріали XIII Міжнародної науково-практичної конференції з інформаційних систем та технологій «Infocom Advanced Solutions – 2025».* Київ, 21–22 травня 2025. С. 29–31. URL: https://ist.kpi.ua/wp-content/uploads/2025/05/infocom-advanced-solutions-2025_compressed-3.pdf (дата звернення: 06.06.2025).

11. Mukhin V., Kornaga Y., Bondarenko V., Zavgorodnii V., Herasymenko O., Sholokhov O. Mathematical Model for Heterogeneous Databases Parameters Estimation in Distributed Systems with Dynamic Structure. *2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT)*. Kyiv, Ukraine, 2020. P. 158–161. DOI: 10.1109/ATIT50783.2020.9349331.

12. Hu Z., Mukhin V., Loutskii H., Kornaga Y. Stochastic RA-Network for the Nodes Functioning Analysis in the Distributed Computer Systems. *International Journal of Computer Network and Information Security*. 2016. Vol. 8, No. 6. P. 1–8. URL: https://sciup.org/15015169

13. Корнага Я. І., Барабаш А. О. Математична модель моніторингу подій в системі управління розподіленими базами даних на основі сенсорів подій. *Інфокомунікаційні та комп'ютерні технології*. 2023. Т. 1, № 05. С. 161–166. DOI: 10.36994/2788-5518-2023-01-05-17.

14. Mukhin V., Kornaga Y., Steshyn V., Mostovoy Y. Adaptive Security System Based on Intelligent Agents for Distributed Computer Systems. *Proceedings of the 13th International Conference on Development and Application Systems (DAS 2016)*. Suceava, Romania, 2016. P. 320–325. DOI: 10.1109/DAAS.2016.7492595.

15. Bondi A. B. Characteristics of scalability and their impact on performance. *Proceedings of the 2nd International Workshop on Software and Performance*. 2000. P. 195–203. DOI: 10.1145/350391.350432.

16. Яскевич В. О., Клочко О. Ю. Методи підвищення відмовостійкості інтернет сервісів. *Кібербезпека: освіта, наука, техніка*. 2019. № 3 (3). С. 104–111. DOI: 10.28925/2663-4023.2019.3.104111

17. Касерес А., Глоба Л. С. Підхід до формування мультихмарного середовища на основі багатокритеріального аналізу та онтологічного моделювання. *Системи управління, навігації та зв'язку*. 2024. № 4. С. 84–91. DOI: 10.26906/SUNZ.2024.4.084

18. Гіоргізова-Гай В. Ш., Кирюша Б. А. Хмарні послуги в корпоративній інфраструктурі. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2024. Т. 35 (74), № 6. С. 32–39. DOI: 10.32782/2663-5941/2024.6.2/05

19. Mankotia A. Comparing High Availability and Disaster Recovery in Multi-Cloud Environments. *International Journal of Computer Trends and Technology*. 2024. Vol. 72, No. 7. P. 113–121. DOI: 10.14445/22312803/IJCTT-V72I7P115

20. Yalate A. Demystifying Multi-Cloud Architecture: Foundational Concepts and Design Patterns. *European Journal of Computer Science and Information Technology*. 2025. Vol. 13, No. 24. P. 51–62. DOI: 10.37745/ejcsit.2013/vol13n245162

21. Недоступ Д. М., Солом'яний М. В. Аналіз підходів забезпечення відмовостійкості архітектур Extended Cloud. *Матеріали VIII Міжнародної науково-технічної конференції «Проблеми електромагнітної сумісності перспективних безпроводових мереж зв'язку»*. Харків, 2022. С. 39–40. URL: http://openarchive.nure.ua/handle/document/23319 (дата звернення: 06.06.2025).

22. Shah Y., Thakkar E., Bhavsar S. Fault Tolerance in Cloud and Fog Computing–A Holistic View. *Data Science and Intelligent Applications*. 2021. P. 415–422. DOI: 10.1007/978-981-15-4474-3_46

23. Agrawal A., Toshniwal D. Fault Tolerance in IoT: Techniques and Comparative Study. *Asian Journal For Convergence In Technology*. 2021. Vol. 7, No. 1. P. 49–52. DOI: 10.33130/AJCT.2021v07i01.011

24. Vardhan A. S. S., Verma A., Ogale J., Saket R. K., Galloway S. Modern aspects of probabilistic distributions for reliability evaluation of engineering systems. In: *Reliability Analysis of Modern Power Systems*. Wiley, 2024. P. 217–245. DOI: 10.1002/9781394226771.ch12

25. Ocansey S. A., Bikdash M. Temperature-Driven Reliability Analysis of Power Grid Failures: A Weibull Distribution Approach To Outage Prediction and Mitigation. *Journal of Statistical Theory and Applications*. 2025. Vol. 24, No. 1. P. 233–246. DOI: 10.1007/s44199-025-00109-y

26. Guo Z., Li J., Ramesh R. Scalable, adaptable, and fast estimation of transient downtime in virtual infrastructures using convex decomposition and sample path randomization. INFORMS Journal on Computing. 2020. Vol. 32, No. 2. P. 321–345. DOI: 10.1287/ijoc.2019.0888

Bachkala B.O., Tymoshyn Yu.A. MODELS FOR ENSURING HIGH AVAILABILITY AND FAULT TOLERANCE IN DISTRIBUTED DIGITAL INFRASTRUCTURE SYSTEMS

As information technologies advance and data volumes grow at a rapid pace, modern organizations increasingly favor the use of cloud computing to support their digital infrastructure. This trend has led to the spread of complex distributed computer systems built on cloud, multi-cloud, fog, and edge computing solutions. Such systems typically operate under conditions of potential technical failures (network, hardware, and software malfunctions) and external factors, and therefore must remain operational in the event of failures of individual components or even entire nodes. Modern businesses impose strict reliability requirements on these systems, since even minimal downtime can result in significant financial losses and a decline in customer trust. To achieve continuous operation of cloud-based distributed systems, it is necessary to implement models for automatic analysis of processing node status parameters and their recovery, and ensure high availability through data replication, architectural design, and component redundancy.

This article contains the analysis of models for ensuring fault tolerance and high availability in modern distributed systems, including cloud, multi-cloud, fog, and edge Internet of Things (IoT) systems. The article identifies and examines current methods for achieving fault tolerance and high availability for various types of distributed systems and provides recommendations for applying these models and methods based on the specifics of the environment and quality of service (QoS) requirements. In addition, the article proposes an adaptive data backup model for edge systems, which accounts for network failures using the Weibull distribution. The results of modeling the relationship between the target system's reliability level and the backup interval are presented. The main motivation of the conducted research is to systematize approaches to improving the reliability of distributed systems and to develop a backup strategy adapted for edge systems characterized by network-related component failures.

Key words: *High availability, fault tolerance, distributed systems, cloud computing, multi-cloud, edge computing, IoT.*

Дата надходження статті: 10.06.2025
Дата прийняття статті: 16.06.2025
Опубліковано: 27.10.2025