ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
Навчально-науковий інститут інформаційних технологій
Кафедра Комп'ютерної інженерії

О.М. Ткаченко, Я.І. Торошанко, А.В. Лемешко,
В.О. Сосновий, С.С. Коротков

КОМП'ЮТЕРНІ МЕРЕЖІ: КОНТРОЛЬ ТА ПРОГНОЗУВАННЯ ПЕРЕВАНТАЖЕНЬ

Навчальний посібник

2021

УДК 004.77

Рекомендовано до видання вченою радою Навчально-наукового інституту інформаційних технологій Державного університету телекомунікацій (протокол № 8 від 31.03.2021 р.)

Укладачі: О.М. Ткаченко, Я.І. Торошанко, А.В. Лемешко,
В.О. Сосновий, С.С. Коротков.

Рецензенти:
д. т. н., проф. Бондарчук А.П.
к. т. н. Голубенко О.І.

КОМП'ЮТЕРНІ МЕРЕЖІ: КОНТРОЛЬ ТА ПРОГНОЗУВАННЯ ПЕРЕВАНТАЖЕНЬ. Навчальний посібник / О.М. Ткаченко, Я.І. Торошанко, А.В. Лемешко, В.О. Сосновий, С.С. Коротков., К. : ДУТ, 2021, 77с

Учбовий посібник містить короткі теоретичні відомості, загальні методичні рекомендації щодо контролю та прогнозування перевантажень в комп'ютерних мережах. Розглянуті рішення ґрунтуються на використанні розвинутої архітектури нейронної мережі, тому актуальним та важливим є вивчення алгоритмів навчання нейронної мережі для динамічних процесів передачі даних.

ЗМІСТ

ВСТУП

Сучасні інформаційно-комунікаційні системи характеризуються великою складністю і високою вартістю. Внаслідок цього евристичні підходи до формування архітектури, вибору основних конструктивних і експлуатаційних характеристик, оцінки їх параметрів поступово замінюються регулярними методами аналізу і синтезу.

Якщо раніше при розгортанні телекомунікаційних мереж досить було користуватися інтуїтивними міркуваннями і здоровим глуздом, то тепер стає необхідним володіти математичним апаратом, що дозволяє розрахувати оптимальну структуру окремих сегментів і пристроїв, а часом і вигляд всієї мережі в цілому [1-4]. У складних сучасних задачах за допомогою інтуїції і "фізичного сенсу" вдається сконструювати лише досить посередні структури, які, як правило, можуть бути замінені на більш досконалі, якщо звернутися до систематичної теорії. Необхідно застосовувати, перш за все, математичні методи синтезу, розрахунку та проектування з урахуванням специфіки телекомунікаційних мереж. З плином часу вони набувають все більшого значення.

Технологія моніторингу та аналізу представляє собою набір діагностичних засобів і методику їх використання, які дозволяють дати об'єктивну оцінку якості роботи прикладних програм в мережі і обґрунтувати рекомендації щодо поліпшення їх роботи.

Основний недолік більшості програм модернізації мережі – постійна заміна обладнання на більш продуктивне. При цьому значна частина проблем функціонування мережі криється зовсім не у вичерпанні ліміту пропускної здатності, а в проблемах взаємодії апаратури, конфігурації, організації мережі і роботи користувачів.

Інший недолік – використання адміністраторами мережі однієї-двох навмання вибраних частинних методик діагностики та моніторингу і, відповідно, необ'єктивне оцінювання стану мережі.

Методика безперервної діагностики мережі полягає в розбитті процесу на

наступні взаємопов'язані етапи: *визначення* частинних параметрів, *отримання* узагальнених параметрів на основі частинних параметрів, *визначення* на основі відповідної обробки попередніх еталонів узагальнених мережних показників, *вибір* та обгрунтування показників якості.

Якість надання послуг в телекомунікаційній мережі в значній мірі визначається алгоритмами маршрутизації, управління потоками даних і функціонування в умовах перевантаження. Різні методи маршрутизації – статичні чи динамічні, локальні чи централізовані, детерміністичні чи стохастичні – намагаються спрямувати повідомлення від джерела до місця призначення таким чином, щоб:

– затримка даних в мережі була мінімальною;

– управління трафіком даних повинно забезпечити уникнення чи мінімізацію появи перевантаження.

Сукупність ресурсів комп'ютерної мережі має обмежені (скінченні) можливості, які спричиняють виникнення конфліктів між користувачами системи. Ці конфлікти можуть спричинити зниження продуктивності системи до такого моменту, що система стає "засміченою". Причиною такої "засміченості" є багатократне дублювання даних і різке збільшення технологічної управляючої інформації. Внаслідок цього пропускна здатність суттєво зменшується, можливо і до нульової позначки. Це є типова поведінка "конкуруючих" систем [4]. Така ситуація може призвести до колапсу мережі.

Мережі не можуть обслуговувати весь запропонований їм трафік без деякого контролю. Повинні бути правила, які керують зовнішнім трафіком і координацією потоку всередині мережі.

Існує декілька визначень перевантаження, які, однак, не суперечать одне одному. Будемо користуватись наступним визначенням [4, 6]: "перевантаження – це втрата даних користувачем, спричинена збільшенням навантаження в мережі".

Отже, управління перевантаженням можна визначити як набір механізмів, що запобігають або зменшують таке погіршення. Якщо мережа нездатна запобігти втраті даних користувача, тоді потрібно спробувати максимально обмежити

втрати, і, в подальшому, спробувати бути справедливою до всіх постраждалих користувачів.

Перевантаження має значний вплив на ключові показники ефективності телекомунікаційної мережі і якість обслуговування користувачів. Вищесказане обумовлює актуальність і необхідність вивчення питань в цьому напрямку.

1 ЗАДАЧА КОНТРОЛЯ І ПРОГНОЗУВАННЯ ПЕРЕВАНТАЖЕНЬ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

1.1 Способи управління перевантаженнями в комп'ютерних мережах

Ефективність використання мережі в значній мірі визначається якістю управління в умовах перевантаження. Коли в мережу надходить надто великий обсяг даних, може виникнути перевантаження, і робочі характеристики мережі погіршуються. При надмірних завантаженнях пропускна здатність (продуктивність) мережі може стати нульовою. Така ситуація може призвести до колапсу мережі.

Сьогодні існує велика кількість алгоритмів боротьби з перевантаженнями даних алгоритмів, які задовольняють ті чи інші потреби користувачів.

В роботі [7] представлена класифікація систем боротьби з перевантаженнями RED, які використовуються в протоколі TCP (Transmission Control Protocol) мережі інтернет. На основі даної роботи була запропонована розгорнута класифікація, яка доповнена іншими реалізаціями алгоритмів боротьби з перевантаженнями, які використовуються на сьогоднішній день [8]. Описані характеристики алгоритмів, визначені переваги та недоліки їх застосування в певних умовах роботи мереж.

Досить поширеним для боротьби з перевантаженнями є алгоритм TCP Veno, який здатен ефективно працювати як в проводових мережах, так і в безпроводових. В багатьох роботах розглядається використання даного алгоритму при стандартних значеннях його параметрів. Подальші дослідження показали можливість збільшення пропускної здатності мережі шляхом оптимізації параметрів алгоритму TCP Veno [9].

Причиною перевантаження може бути недостатній об’єм пам'яті для вхідних буферів або невелика швидкодія процесора маршрутизатора. Однак, слід зауважити, що значне збільшення об’єму пам'яті вхідних буферів може призвести

до збільшення негативних наслідків, пов’язаних з перевантаженням [6]. Пояснення цього полягає в тому, що із зростанням об’єму буферної пам’яті збільшується кількість необроблених пакетів і час очікування їх обробки може перевищити допустимі норми тривалості тайм-ауту, при цьому з'являються повторно передані пакети, що призводить до подальшого зниження корисної пропускної здатності мережі.

Іншими словами, перевантаження може бути причиною виникнення лавинного процесу: переповнення буфера призводить до втрати пакетів, які доведеться передавати повторно або навіть кілька разів. Таким чином, обчислювальний вузол маршрутизатора-відправника отримує надлишкове паразитне завантаження.

Причиною перевантаження може бути повільний процесор або “вузьке горло” – низька пропускна здатність окремої дільниці мережі. Просте підвищення швидкодії процесора або інтерфейсу не завжди усуває проблему – вузьке місце, як правило, переноситься в інший сегмент мережі.

Перевантаження мережі, як правило, носить тимчасовий характер і означає, що вхідне навантаження на даному проміжку часу перевищило можливості ресурсів даної частини системи.

Рішення даної задачі може здійснюватись за двома напрямками: збільшення ресурсів системи і регулювання вхідного трафіку перевантаженої ділянки мережі [1, 3].

Перший шлях залежить від конкретної реалізації та наявності додаткових ресурсів. Наприклад, у супутникових системах підвищення продуктивності (пропускної спроможності) може бути досягнуто за рахунок збільшення потужності передавача.

В локальних підмережах можуть тимчасово використовуватись телефонні лінії з модемами між певними її точками. У випадку надзвичайних ситуацій може використовуватись резервне обладнання, запасні маршрутизатори тощо. Задача уникнення або зменшення рівня перевантаження може також вирішуватись на рівні підсистеми маршрутизації шляхом розподілу трафіку по декількох

маршрутах замість постійного використання одного і того ж, нехай навіть оптимального шляху.

У випадку, коли немає можливості збільшувати пропускну спроможність або вона уже збільшена до межі, єдиний спосіб боротьби з перевантаженням полягає в зменшенні навантаження шляхом зниження рівня обслуговування деяких або всіх користувачів, включаючи відмову в їх обслуговуванні.

Одним з поширених методів боротьби з перевантаженнями є управління зі зворотним зв'язком. Механізм і завдання управління вирішується на транспортному рівні засобами протоколу TCP [1, 10]. При виявленні перевантаження швидкість передачі знижується шляхом зменшення розміру ковзного вікна.

По суті, має місце управління з зворотним зв'язком, що запізнюється. При неправильному обрахуванні характеристик запізнювання система може втратити стійкість і перейти в незатухаючий коливальний режим, або коригування інтенсивності потоку буде здійснюватися занадто пізно [6, 10]. Компенсація затримки зворотного зв'язку може виконуватися методами передбачення, наприклад, з використанням моделі авторегресії і ковзного середнього (АРКС) або шляхом усереднення параметрів вікна. Другий варіант простіший, але, природно, забезпечує значно більш низьку якість сервісу.

Позитивного результату також можна досягти шляхом варіації значень тайм-аутів, зміною протоколів квитування і повторної передачі пакетів, зміною схеми буферизації.

Управління зі зворотним зв'язком широко використовується в архітектурі інтегрованих служб (Integrated Service Architecture – ISA) для підтримки служб з різними рівнями якості сервісу (Quality of Service – QoS) в Інтернеті і в приватних об'єднаних мережах [1].

Для систем без зворотного зв'язку вирішення проблеми вирівнювання швидкості передачі даних може бути вирішено за допомогою алгоритмів “дірявого відра” і “маркерного відра” [4, 11]. Алгоритм “дірявого відра”, являє собою механізм регулювання трафіку, коли частина потоку пакетів, що перевищує

пропускну здатність мережі, просто відкидається або позначається як надлишкова або низько пріоритетна.

В даний час одним з основних стандартів при побудові систем управління телекомунікаційними мережами є концепція управління *TMN* [1-4]. Мережа управління телекомунікаціями *TMN* представляє з себе спеціальну інфраструктуру, що забезпечує управління шляхом організації взаємодії з компонентами різних телекомунікаційних мереж за допомогою мережі передачі даних на основі єдиних інтерфейсів і протоколів обміну інформацією. Взаємозв'язок інфраструктури *TMN* з телекомунікаційною мережею показаний на рис. 1.1.

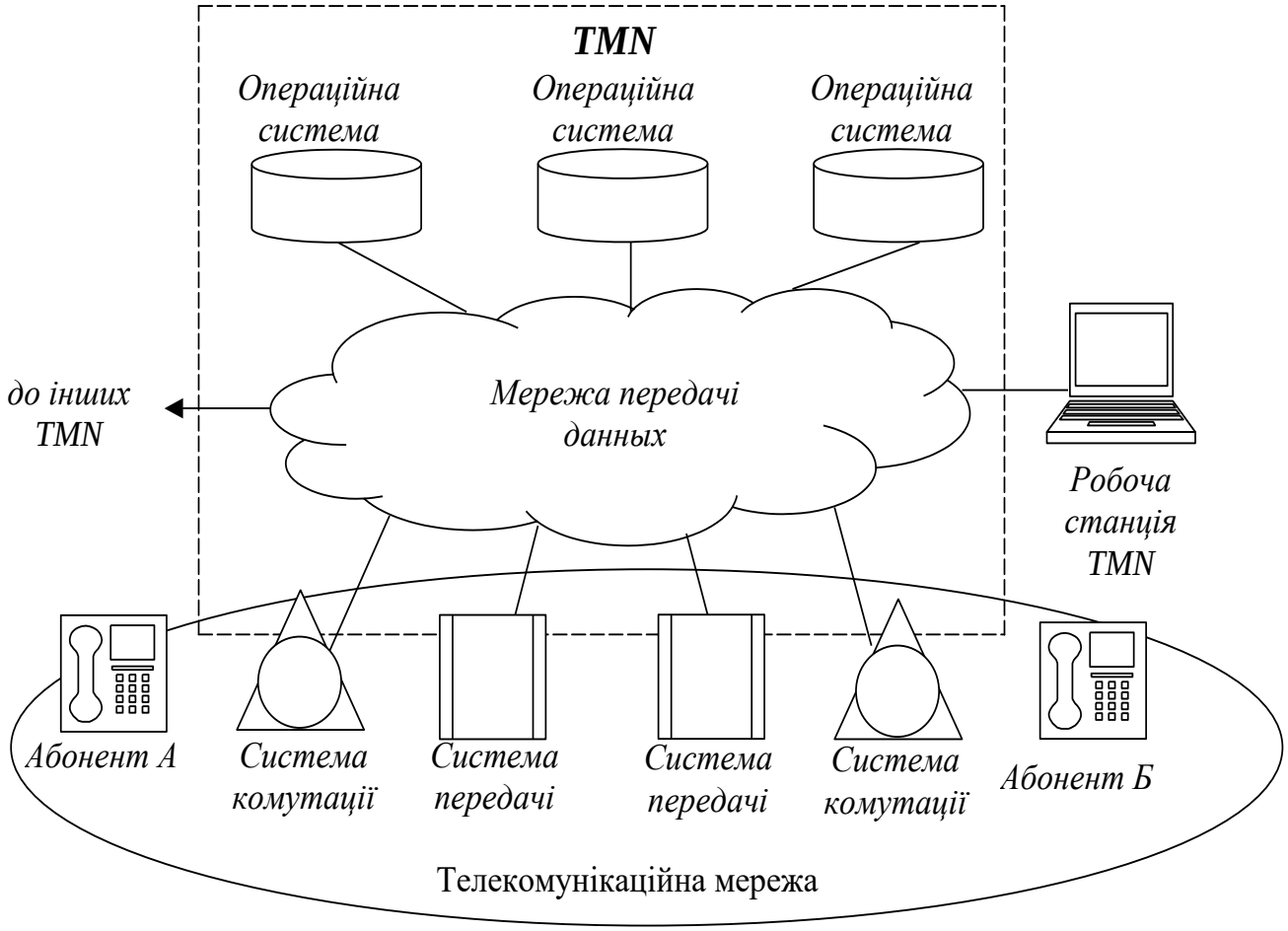

Рисунок 1.1 – Взаємозв'язок інфраструктури *TMN* з телекомунікаційною мережею

Організаційна структура *TMN* створюється для реалізації задач управління, експлуатації і технічного обслуговування різнорідного телекомунікаційного

обладнання, оперативного контролю і адміністрування мережних пристроїв, а також узгодженої взаємодії між різними типами систем управління в цілях надання послуг зв'язку із заданою якістю.

До сфери управління *TMN* входять практично всі існуючі в даний час види мереж і систем зв'язку, а також типи телекомунікаційного обладнання. Об'єктами управління *TMN* є телекомунікаційні ресурси, що фізично надають собою реальне обладнання зв'язку, на яке можливе здійснення цілеспрямованої управляючої дії. Фізично компоненти керованої мережі електрозв'язку (обладнання систем комутації, систем передачі і т.д., визначувані як мережні елементи), можуть бути як зосередженими (централізованими), так і розподіленими.

Реалізація прикладних процесів управління здійснюється операційними системами шляхом обміну управляючою інформацією з мережними елементами. При цьому операційні системи забезпечують обробку даних, що поступають від мережних елементів, підтримують інформаційну модель мережі електрозв'язку, забезпечують роботу прикладних програмних засобів управління. Крім того, операційні системи забезпечують підтримку терміналів користувача у вигляді їх робочих станцій. Таким чином, *TMN* здійснює моніторинг всієї мережі електрозв'язку, виробляє управляючі рішення, виходячи з реальних мережних умов і супутньої інформації.

Система *TMN* має наступні функціональні області управління:

- управління робочими характеристиками;
- управління надійністю та усунення несправностей;
- управління конфігурацією;
- управління розрахунками за послуги;
- управління безпекою.

У процесі управління робочими характеристиками (якістю роботи, якістю сервісу – *QoS*) генеруються команди контролю та управління, необхідні для визначення технічного стану мережних елементів і ефективності функціонування мережі електрозв'язку в цілому. Інформація про роботу мережі поступає із заданими інтервалами. За результатами обробки статистики роботи мережі

виробляються відповідні управляючі дії. Іншими словами, у даній функціональній області реалізується етап вимірювання робочих характеристик та оцінювання стану мережі.

Управління надійністю та усунення несправностей забезпечується шляхом виявлення, визначення несправності в мережі, їх реєстрацію, доведення відповідної інформації до обслуговуючого персоналу, видачу рекомендацій по усуненню несправностей.

Управління конфігурацією реалізується у процесі моніторингу мережних елементів (їх типів, місцезнаходження, ідентифікації параметрів та стану і т.п.), включення елементів в роботу, їх конфігурування і виходу з робочого стану, встановлення і змін фізичних з'єднань між елементами.

Управління розрахунками – це контроль степені використання мережних ресурсів і підтримання функції автоматичного нарахування оплати (білінгу).

Управління безпекою необхідне для захисту мережі від несанкціонованого доступу. Воно може включати обмеження доступу за допомогою паролів, видачу сигналів тривоги при спробах несанкціонованого доступу, відключення небажаних користувачів або, навіть, криптографічний захист інформації.

В комп'ютерних мережах без резервування ресурсів, управління повинно бути реакційним. Схема реакційного управління перевантаженням може бути впроваджена в двох місцях: – в комутаторах, де виникає перевантаження; – в джерелах, в яких контролюється надходження пакетів в мережу. Як правило, комутатор використовує деякий набір даних (наприклад, переповнення буферів) для визначення виникнення перевантаження, і непрямим чином чи недвозначно передає цю проблему джерелам, які зменшують їхній вхідний трафік.

Існують декілька альтернатив виявлення і уникнення перевантаження, основними з яких є [7, 8, 13, 14]:

– контроль заповненості та середнього часу зайнятості вхідних буферів, найбільш прийнятний при розділених вхідних чергах;

– контроль використання вихідних ліній і вихідних черг OQ: встановлено, що перевантаження виникає, коли використання мережі переходить поріг чутливості

приблизно 90%, і цей показник може бути використаний як сигнал попередження перевантаження.

– аналіз кругових затримок пакетів: зростання цих затримок свідчить про збільшення розміру черги і можливості перевантаження;

– постійне відстежування стану мережі, використовуючи певну схему дослідження.

Сигналізація і повідомлення про перевантаження від перевантаженого вузла на джерело може бути явним чи неявним. Коли повідомлення являється явним, комутатор посилає інформацію в заголовках пакета чи у відповідних керуючих пакетах (службові управляючі команди подавлення джерела, стримування, пакети оновлення стану, повідомленнях про управління швидкістю тощо). Сигналізація явних перевантажень викликає додаткове навантаження в мережі, так як мережа потребує передачі більшої кількості пакетів, ніж зазвичай. Це може призвести до втрати продуктивності, якщо часові витрати на сигналізацію не контролюються належним чином.

У випадку неявної сигналізації виявлення перевантаженого вузла здійснюється джерелом, – як правило за признаками перевищення заданого часу очікування (тайм-ауту) підтвердження виданих в мережу пакетів.

1.2 Альтернативна маршрутизація

Розглянемо один з випадків боротьби з перевантаженням способом альтернативної маршрутизації. В разі неможливості пропускання трафіку через перевантажений або пошкоджений з тих чи інших причин вузол потрібно знайти альтернативний шлях передачі пакетів [12].

Як приклад, розглянемо випадок системи обслуговування, коли до вузла надходять вхідні інформаційні потоки IS1, IS2 та IS3, які об’єднуються у спільну вхідну чергу IQ (рис. 1.2).

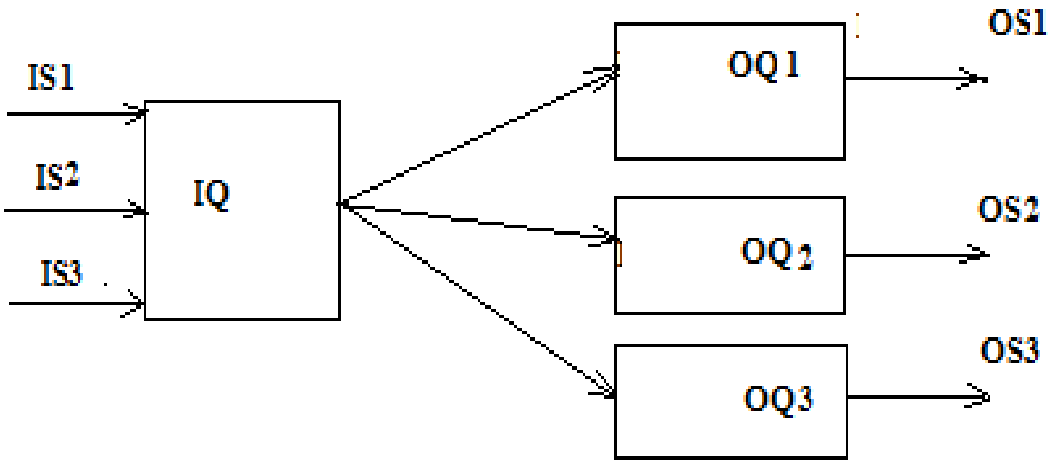

Рисунок 1.2 – Система обслуговування із спільною вхідною чергою

Із вхідної черги IQ згідно поточного алгоритму (таблиць) маршрутизації формуються вихідні черги OQ1, OQ2, OQ3 і відповідні їм вихідні потоки OS1, OS2, OS3.

На перші два потоки IS1 та IS2 поступає певна кількість пакетів даних, яка передається до вузлів призначення. В третьому потоці IS3 передається об'єм даних, де окрім даних джерела, яке зазвичай передає дані в цій мережі, також передаються дані, які передаються по альтернативному маршруту. Це створює додаткове навантаження і, як наслідок, спричиняє виникненню перевантаження вхідної черги IQ, а також тих чи інших вихідних черг.

Одним із ефективних способів запобігання перевантаженню є тимчасове відключення потоків, які не пов’язані із надзвичайною ситуацією. Таким чином, планується зменшити навантаження на вході вузла. Враховуючи, що потоки взаємопов'язані, відключення одного потоку призведе до відключення всіх потоків.

Як наслідок, через цей вузол не буде передано жоден з пакетів даних (рис. 1.3), що не вирішує проблему перевантаження без втрати пакетів і погіршення якості обслуговування, що тільки погіршує поточний стан мережі.

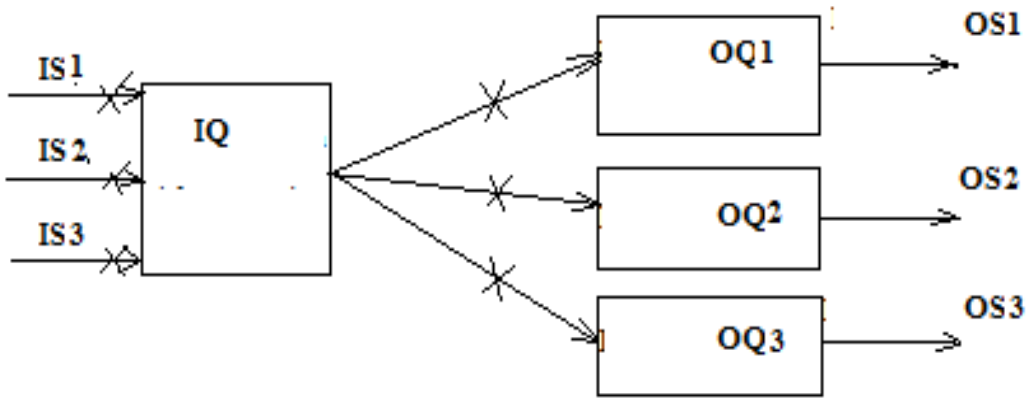

Рисунок 1.3 – Система з тимчасовим відключенням потоків спільної вхідної черги

Рішенням такої задачі може бути розділення спільної вхідної черги IQ на три незалежні черги IQ1, IQ2, IQ3 відповідно до вхідних потоків IS1, IS2, IS3 (рис. 1.4). Зауважимо, що можливі і змішані системи (при більшому числі вхідних потоків), коли певні групи вхідних потоків об’єднуються у спільні вхідні черги, кожна з яких розглядається незалежно одна від одної.

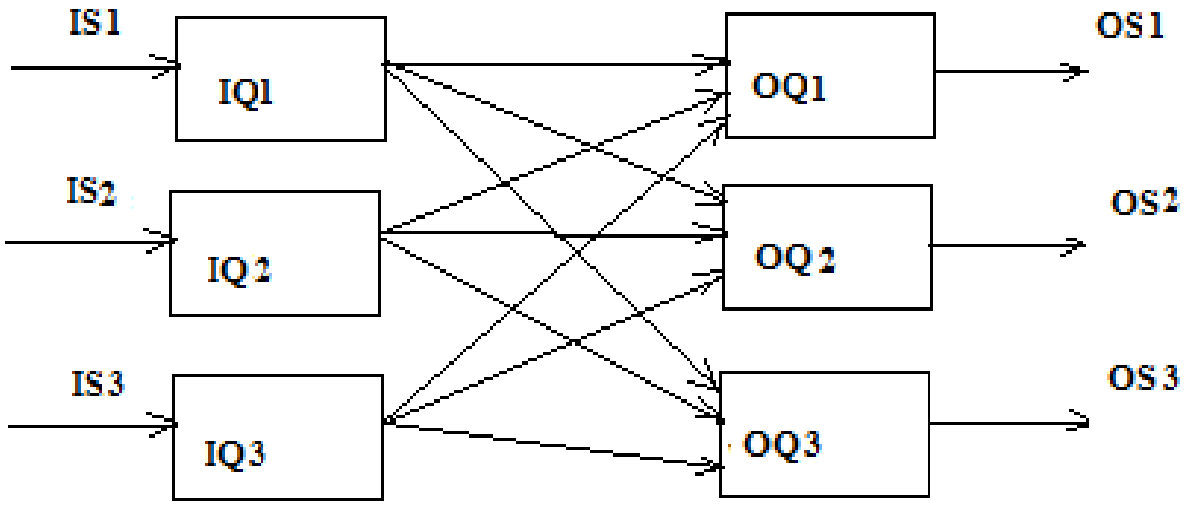

Рисунок 1.4 – Система з розділенням потоків вхідної черги

Застосувавши представлений вище метод, ми можемо відключити потоки IQ1 та IQ2 (рис. 1.5). Таким чином, зменшивши таким чином інформаційне навантаження на вузол не блокуючи потік IQ3, пов’язаний із надзвичайною

ситуацією. Після обробки інформації перевантаженої черги відключені потоки розблоковуються і вузол переходить у штатний режим функціонування в. мережі.

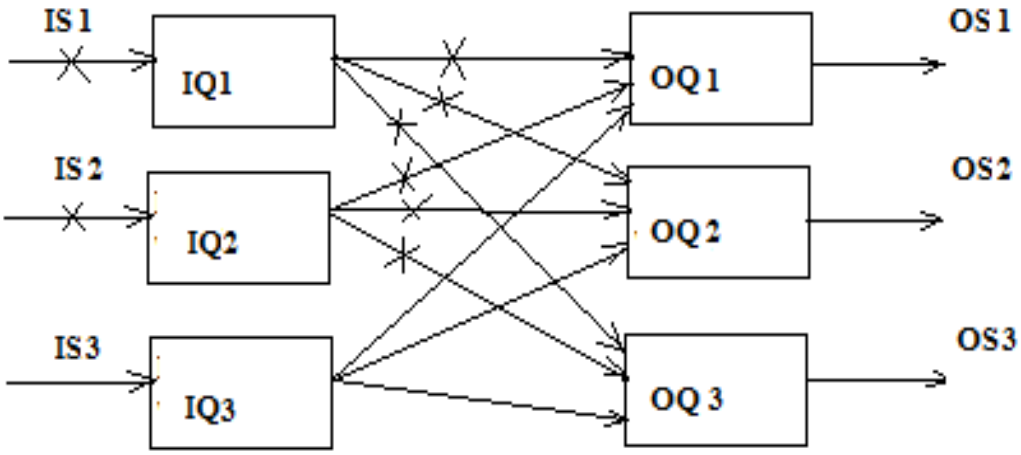

Рисунок 1.5 – Система з блокуванням потоків вхідної черги

Отже, зазначений спосіб є ефективним для другого випадку. Переваги цього способу є ефективність і простота реалізації, відсутність додаткових затрат на обладнання, універсальність (даний спосіб можна пристосувати і для вирішення інших задач).

Таким чином, показана можливість запобігання перевантаження на основі способу розділення вхідного потоку на окремі вхідні черги, що надає можливість більш гнучко регулювати вхідний трафік вузла в умовах надзвичайних ситуацій. Показана також можливість використання методів і кількісних характеристик чутливості для визначення стану мережі.

1.3 Захист від перевантажень на рівні керування потоками

Інтенсивність передачі пакетів джерелом повинна бути узгоджена з інтенсивністю обробки пакетів одержувачем. Якщо інтенсивність передачі перевищує інтенсивність обробки, пакети будуть накопичуватися в мережі, що призведе до зниження пропускної здатності мережі, а в результаті-до блокування ділянки мережі та мережі в цілому через переповнення буферної пам'яті, яка

використовується в вузлах для проміжного зберігання пакетів. Цю ситуацію можна виключити, якщо не допускати передачі пакетів з інтенсивністю, що перевищує інтенсивність обробки. Узгодження забезпечується за рахунок керування потоком пакетів між джерелом і одержувачем.

Керування потоками в обчислювальній мережі проводиться на декількох рівнях (рис. 1.6) і реалізується відповідними протоколами.

Основні функції керування потоками реалізуються протоколом керування каналом – найважливішим елементом тракту, що зв'язує процеси в головних термінальних комп'ютерах. Коректність потоків у кожному з каналів, що складають тракт, забезпечує коректне функціонування всього тракту процес-процес. Однак специфіка функціонування вузлів, СПД і комп'ютерів породжує необхідність в особливих елементах управління потоками на рівнях більш високих, ніж канал.

Керування потоком грунтується на механізмі квитанцій - повідомлень, що підтверджують прийняття пакету адресатом. Для управління використовуються різні способи квитування, що реалізуються протоколами відповідних рівнів.

Керування потоками в каналі, тобто між двома вузлами СПД, має забезпечувати ефективне використання пропускної здатності каналу і запобігати переповнення буферів, що приводить до блокування передачі пакетів у каналі. Основний принцип керування полягає в наступному. Рухаючись пакет зберігається в пам'яті вузла, який передає, чекаючи прийому квитанції про правильний прийом пакету вузлом-одержувачем.

Якщо пакет у процесі передачі по каналу був спотворений завадами, передавання повинне бути повторене одним із двох способів: посилкою у вузол-джерело сигналу перезапиту пакета (негативної квитанції) або за допомогою тайм-ауту.

Тайм-аут – проміжок часу, що відводиться для отримання сигналу, що підтверджує виконання відповідної дії. Якщо протягом тайм-ауту підтвердження у вузол-джерело не надійшло, пакет передається знову.

Спосіб повторної передачі пакету на основі тайм-ауту є найбільш зручним і надійним з наступних причин. Позитивні квитанції (підтвердження) і негативні

квитанції (перезапити) можуть бути втрачені (спотворені) в каналі, і, отже, працездатність інформаційного каналу, для керування яким використовується механізм негативних квитанцій, повинна підтримуватися механізмом тайм-ауту.

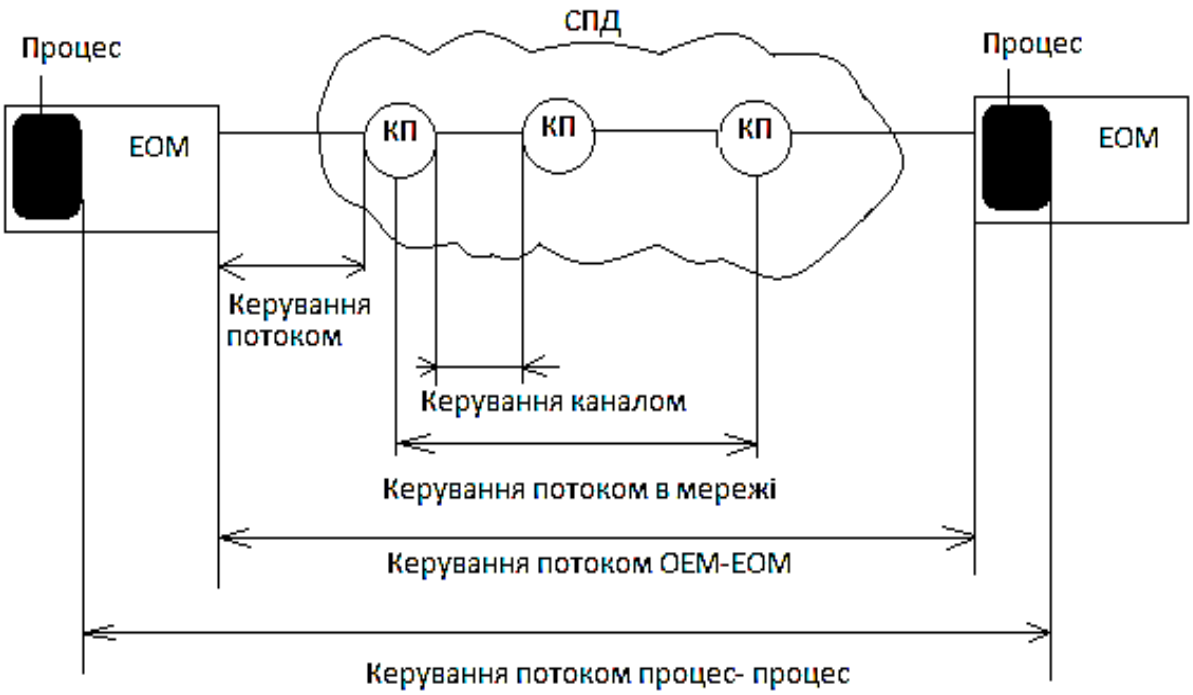

Рисунок 1.6 – Рівні керування потоками

Однак, використання тайм-ауту призводить до наступного небажаного явища: якщо квитанція-підтвердження втрачена в каналі або надійшла надто пізно, вузол-джерело після закінчення тайм-ауту повторно передає пакет, в результаті чого у вузлі-одержувачі утворюється копія пакету. Це явище виключається за рахунок введення в пакети змінного біта, значення якого встановлюється для послідовно переданих пакетів в 0, 1, 0, 1…. Приймаючий вузол контролює значення змінного біта: значення, які чергуються свідчать про коректний процес передачи: поява підряд значень 0 або 1 свідчить про те, що новий пакет – копія раніше прийнятого.

У розглянутому протоколі передачі пакетів по каналу припускається, що наступний пакет передається тільки після підтвердження про прийом попереднього. У такому випадку пропускна здатність каналу використується лише частково і втрати її, зростають зі збільшенням швидкості передачі і протяжності

каналу. Щоб уникнути цього використовується багатопакетний протокол передачі, що допускає передачу сукупності пакетів, перш ніж прийде квитанція, що підтверджує прийом першого з них. Щоб співвідносити пакети і квитанції, послідовність переданих пакетів повинна бути пронумерована: першому пакету присвоюється номер 0, другого – 1 і т. д. Гранична кількість пакетів, які передаються до отримання квитанції, називається шириною вікна.

Перед початком передачі у вузлі-джерелі і вузлі-одержувачі повинні бути встановлені однакові початкові номери 0. У результаті цього перший, який передається і перший,який приймається пакети будуть мати номери 0. У момент часу t порядковий номер пакету, який передається дорівнюватиме N_t, а номер пакету, який очіується у вузлі-одержувача – N_r.

У початковий період часу вузол передає пакети з номерами $0 \leq N_t < W$, де W – ширина вікна. Вузол-одержувач всі отримані пакети перевіряє на наявність помилок і спотворені пакети вважаються втраченими. Якщо надійшов пакет з номером $N=N_t$, то він буде прийнятий в правильній послідовності, про що надсилається квитанція, яка містить номер N.

Якщо $N>N_r$ то прийнятий пакет випереджає правильну послідовність і відповідно пакет який надійшов загублений. У цьому випадку квитанція про прийом пакета N не висилається. Якщо $N<N_r$, то пакет N є копією попереднього пакету, повторна передача якого сталася через втрату квитанції. Тому вузол-одержувач посилає квитанцію, що підтверджує прийом пакета N, а прийнята копія пакету знищується.

Вузол-джерело реагує на квитанції наступним чином. При потраплянні чергової квитанції з номером пакету N відповідний пакет знищується і передається черговий пакет $N_t = W$. Так, при підтвердженні прийому пакета 0 передається пакет $N_t = W$, при підтвердженні пакета 1 – пакет $N_t = W+1$ і т. д. З урахуванням порядку роботи вузла-одержувача отримання підтвердження про прийом пакета з номером N, великі номери попереднього підтвердження пакету, свідчить про прийом всієї послідовності пакетів включно. Копії цих пакетів знищуються. Для кожної збереженої у вузлі-джерелі копії пакета призначається тайм-аут. Якщо він

закінчується до отримання квитанції, пакет повторно направляється одержувачу і за ним подаються всі пакети з подальшими номерами, оскільки вузол-одержувач ігнорує всі пакети, що надійшли слідом за втраченим. Цей порядок передачі пакетів гарантує прийом пакетів в тій послідовності, в якій вони передавалися вузлом-джерелом. Недолік описаного протоколу – повторна передача правильно прийнятих пакетів, що знижує реальну пропускну здатність каналу.

Для кодування номерів пакетів відводиться кінцеве число бітів k і нумерація проводиться циклічно: 0, 1, 2…. 2^k -1, 0, 1, ... 2^k -1... . Ширина вікна W визначає число пакетів, копії яких зберігаються після передачі у вузлі-джерелі, і одночасно число буферів, які використовуються в вузлі для зберігання пакетів. Чим більше значення тим ефективніше використовується пропускна здатність каналу, але це досягається за рахунок збільшення у вузлах ємності пам'яті, що забезпечує роботу каналу. Ширина вікна повинна бути тим більше, чим більше пропускна здатність і протяжність каналу.

Механізм квитанції використовується для виконання ще однієї функції – захисту вузлів від перевантажень, що виникають, коли буферна пам'ять виявляється повністю заповненою пакетами. Для виключення перевантажень вузол призупиняє видачу квитанцій, в результаті чого припиняється прийом пакетів по вхідних у вузол напрямками.

Керування потоком в мережі (рис. 1.6) – між джерелом і адресатом – проводиться таким же чином, як і управління потоком в каналі, однак повинне забезпечувати:

– по-перше, усунення помилок, що вносяться до вузлів зв'язку, оскільки протоколи управління каналом їх не виявляють;

– по-друге, захист від перевантажень, що виникають, якщо вузол-джерело передає пакети з інтенсивністю, що перевищує інтенсивність обробки пакетів адресатом.

З метою правильної доставки пакетів вузол-джерело зберігає їх копії до отримання квитанції від вузла-адресата. Через відмову або зайнятість ресурсів комп'ютери втрачають здатність приймати адресовані їм пакети протягом якогось

часу. Якщо джерело продовжує відправляти пакети, то вони накопичуються в мережі і перевантажують її. У таких ситуаціях вузол-адресат повинен знищувати пакети, які надходять і не передавати квитанції про їх отримання, чим буде припинена передача пакетів вузлом-джерелом.

Управління потоком (рис. 1.6) між комп'ютером і вузлом СПД (для випадку, коли до вузла підключено єдиний комп'ютер) забезпечується описаними засобами управління каналом, який пов'язує комп'ютер з вузлом. Якщо до вузла підключені декілька комп'ютерів, для управління потоком потрібні спеціальні методи. Як правило, використовується наступний метод.

По-перше, для кожного комп'ютера надається певна кількість буферів, в яких розміщуються пакети, що надходять у вузол.

По-друге, кожному комп'ютеру присвоюються різні пріоритети на передачу пакетів, які залежать від ширини вікна, відведеного в каналі зв'язку комп'ютера (чим ширше вікно, тим вищий пріоритет комп'ютера на передачу у вузол).

Управління потоками на рівні комп'ютер-комп'ютер забезпечується в основному протоколом керування каналом між комп'ютерами і СПД і протоколом керування потоками в СПД. Для виключення переповнення мережі пакетами кожен комп'ютер, починаючи діалог з іншим комп'ютер, повинен переконатися в готовності комп'ютера, якого викликають до взаємодії - прийому пакетів. Така перевірка виконується при встановленні віртуального каналу і необхідна при передачі даних у формі дейтаграм.

Управління потоком між процесами, які реалізовуються в комп’ютерах, полягає в перевірці дозволу на доступ до процесу, який викликається - програма, база даних або термінал.

Розглянута система управління потоками, реалізована протоколами відповідних рівнів, забезпечує захист переданих даних від помилок, передачу пакетів, упорядкованих у ланцюжок за допомогою нумерації, і ефективне використання пропускної здатності каналу і мережі в цілому. Крім того, процедури перевірки готовності адресатів (комп’ютерів і процесів) до взаємодії з джерелом

виклику знижують навантаження на СПД і частково сприяють захисту вузлів зв'язку та СПД в цілому від перенавантажень.

У СПД з комутацією пакетів основні ресурси для функціонування мережі - це пропускна спроможність каналів і ємність буферної пам'яті у вузлах зв'язку. Пропускна здатність каналів, число буферів і топологія мережі визначають граничну пропускну здатність мережі. Реальна пропускна здатність мережі не перевищує граничної.

Зі збільшенням навантаження на мережу пропускна спроможність збільшується до граничного значення, але одночасно з цим зростає час доставки пакетів. Як у будь-якій транспортній системі, пропускна здатність мережі залежить від кількості пакетів, що знаходяться в мережі. Зі збільшенням числа пакетів, які передаються мережею, продуктивність спочатку зростає до максимального значення, а потім починає падати.

Стан мережі, при якому із-за великого числа переданих пакетів різко погіршуються її характеристики, називається перевантаженням. При числі пакетів $M \leq M^*$ в мережі існують умови для вільного просування пакетів до адресатів. При $M > M^*$ ці умови погіршуються і, зрештою, настільки, що мережа виявляється заблокованою пакетами, які в ній знаходяться і продуктивність мережі падає до нуля.

Блокування в мережі виникає через відсутність вільних буферів у вузлах зв'язку. Якщо два вузли А і В пов'язані між собою каналом і всі буфера вузла А зайняті пакетами, які повинні бути передані у вузол В, а всі буфери вузла В – пакетами, призначеними для вузла А, то виникає пряме блокування: жоден пакет з вузлів А і В не може бути переданий у вузол призначення. Непряме блокування виникає при кільцевій топології мережі, наприклад коли вузол А повинен передати пакети у вузол В, вузол В – у вузол С, а вузол С – у вузол А. У цьому випадку для пакетів може не виявитися вільних буферів і передачі повністю блокуються .

Для виключення перевантаження мережі необхідно не допускати надмірного надходження пакетів в мережу. Пакети повинні чекати своєї черги на передачу через мережу, знаходячись у пам'яті головних і термінальних комп’ютерів.

Найбільш простий спосіб захисту від перевантажень – введення для вузлів мережі системи дозволів на введення пакетів в мережу. При цьому кожному вузлу виділяється обмежена кількість дозволів, наприклад 5, на передачу пакетів в мережу. Якщо вузол вводить пакет у мережу, число дозволів зменшується на одиницю. Після того як всі дозволи будуть вичерпані, вузол припиняє прийом пакетів від комп'ютера-джерела.

Коли у вузол надходить пакет, адресований комп'ютеру, вузлу, який обслуговує, число дозволів збільшується на одиницю. Цей механізм виключає можливість переповнення мережі пакетами. Оскільки потоки в вузлах не збалансовані, число пакетів, які відправляються в загальному випадку не збігається з кількістю пакетів, що приймаються, в одних вузлах може бути надлишок дозволів, а в інших-їх дефіцит. Тому вузли повинні передавати надлишкові дозволи іншим вузлам, наприклад за допомогою спеціальних керуючих пакетів.

Оптимальне число дозволів, що надаються вузлом, визначається шляхом моделювання мережі. При цьому враховується гранична допустима кількість пакетів в мережі, залежна від топології, числа буферів і пропускної здатності каналів, а також затримка введення пакетів в мережу і час доставки пакетів.

Додатково до системи дозволів з метою поліпшення умов функціонування вводяться пріоритети для транзитних пакетів. Транзитні пакети передаються УС в першу чергу, що призводить до розвантаження мережі, а пакети, що надходять у вузли від комп'ютера, передаються в мережу тільки в тому випадку, коли всі транзитні пакети передані за відповідними напрямами.

За наявності системи, що захищає від перевантажень мережу в цілому, можуть виникнути локальні перевантаження - в області, яка охоплює кілька сусідніх вузлів. У результаті цього передача пакетів в області блокується.

Щоб відновити працездатність мережі, можна знищити заблоковані пакети за умови, що комп'ютер-джерело зберігає копії пакетів до отримання їх комп'ютером-адресатом. У цьому випадку передача знищених пакетів комп'ютером-джерелом виробляється повторно після закінчення тайм-ауту.

Однак такий спосіб відновлення роботи після перевантаження може призвести до різних небажаних наслідків. Так, можуть бути знищені керуючі пакети, що несуть квитанції, для чого потрібна буде повторна передача правильно доставлених пакетів. Оскільки важко передбачити всі можливі випадки, пов'язані з блокуваннями в мережі, основні функції щодо ліквідації блокувань покладаються на операторів, які здійснюють адміністративне управління мережею.

1.4 Статистичні моделі управління мережами

Важливою проблемою теорії комп'ютерних мереж є розробка основних математичних методів і рівнянь, зручних для вирішення конкретних практичних мережних задач [5, 6]. Подання мережі будь-якого масштабу у вигляді детермінованої системи і опис її відповідними рівняннями з детермінованими параметрами дасть вельми грубий, практично даремний результат.

По-перше, необхідно мати повну апріорну інформацію про параметри і стан мережі в кожний момент. Таке завдання є практично нездійсненним в переважній більшості випадків.

По-друге, відмови обладнання, аномальні ситуації, порушення в роботі мережі, перевантаження через перепади мережного та обчислювального навантаження є принципово випадковими подіями, які ми не можемо контролювати і якими неможливо управляти – їх можна тільки прогнозувати з певною точністю.

По-третє, навіть в ідеальному випадку наявності повної апріорної інформації про параметри, структурі і миттєвий стані мережі ці дані будуть практично марні. Системи рівнянь, якими описується мережу, будуть мати порядок, який можна порівняти з числом мережних і термінальних вузлів. Для чисельного рішення такої системи рівнянь в реальному часі буде потрібний практично нереальний обсяг обчислювальних ресурсів [15, 16]. Крім того, можна стверджувати, що помилки

розрахунків будуть неприпустимо великі, і отриманий результат буде, по суті, даремний.

Тому в даний час тільки статистичні методи опису мереж, процесів обміну даними, синтезу, оцінки параметрів і управління мережами можуть давати результати задовільної точності. При цьому і потрібні обчислювальні ресурси виявляються прийнятними.

Для розробки та застосування статистичних методів мережного моніторингу та аналізу, перш за все, необхідно побудувати математичні моделі мережного трафіку [15]. Технологія моніторингу та аналізу представляє собою набір діагностичних засобів і методику їх використання, які дозволяють дати об'єктивну оцінку якості роботи прикладних програм в мережі і обгрунтувати рекомендації щодо поліпшення їх роботи.

Основний недолік більшості програм модернізації мережі – постійна заміна обладнання на більш продуктивне. При цьому значна частина проблем функціонування мережі криється зовсім не у вичерпанні ліміту пропускної здатності, а в проблемах взаємодії апаратури, конфігурації, організації мережі і роботи користувачів.

Інший недолік – використання адміністраторами мережі однієї-двох навмання вибраних частинних методик діагностики та моніторингу і, відповідно, необ'єктивне оцінювання стану мережі. Суть проведеного дослідження полягає саме в реалізації системного підходу.

Методика безперервної діагностики мережі полягає в розбитті процесу на наступні взаємопов'язані етапи:

– *визначення* частинних параметрів;

– *отримання* узагальнених параметрів на основі частинних параметрів;

– *визначення* на основі відповідної обробки попередніх еталонів узагальнених мережних показників;

– *вибір* та обгрунтування показників якості.

Процес виникнення аномалій унаслідок несанкціонованої мережної активності суб'єкта і процес виявлення розглядаються як Марковські процеси [15]

з наступними станами:

– штатний режим s_1 без перевантажень та аномалій; (1.1)

– аномальний режим s_2 роботи на грані фізичної відмови або перевантаження. (1.2)

У свою чергу, режим s_2 підрозділяється на гілки:

– виявлення аномалії s_{21}; (1.3)

– розпізнавання аномалії s_{22}; (1.4)

– прогноз аномалії s_{23}; (1.5)

– управління аномалією s_{24} (1.6)

У всіх випадках процеси мають дискретний характер, переходи з довільного стану j в будь-який інший стан k, $j,k=\overline{1,6}$, відбуваються стрибком, а імовірність станів міняється залежно від наявності апріорної інформації і інформації, що знов поступає (апостеріорної).

Таким чином, процеси виду (1.1) – (1.6), по суті, є дискретними напівмарківськими процесами з довільним розподілом часу переходу в новий стан. Перевага моделей дискретних напівмарковських процесів (або вкладених ланцюгів Маркова) полягає саме в тому, що можна не цікавитися розподілами t_i.

При заданому початковому стані $\{s_{0i}\}$, $i=\overline{1,6}$ розвиток процесу повністю визначається матрицею імовірності переходу $\{\rho_{jk}\}$, $j,k=\overline{1,6}$, і матрицею функцій розподілів $\{F_{jk}(t_i)\}$.

Однак, в ряді випадків самі вихідні припущення, що послужили визначенням найпростішого потоку і вивчені в багатьох публікаіях, не випливають з розгляду фізичної картини явища. І дійсно, в деяких практичних завданнях спостерігаються відхилення дійсних потоків від потоків найпростішого типу [16]. Здавалося б, у телекомунікаційній мережі з різнорідним трафіком Triple Play/Quadruple Play статистика трафіку відхиляється від найпростішої, оскільки реальний трафік є різнорідним (фрактальним). В силу величезного розмаїття умов протікання реальних явищ такі відхилення повинні бути правилом, а не винятком. Однак

виявляється, що великі розбіжності спостерігаються значно рідше, ніж це можна було б очікувати, виходячи з апріорних міркувань.

Таким чином, поряд із завданням з'ясування причин, в силу яких можуть з'являтися потоки, відмінні від найпростіших, виникає і прямо протилежне завдання: пояснити, чому так часто найпростіший потік добре узгоджується з плином реальних потоків. Вихідною ідеєю зазначених досліджень було припущення, що спостережені потоки представляють собою суми великого числа незалежних потоків малої інтенсивності, кожен з яких передбачається ординарним і стаціонарним. Щодо відсутності післядії ніяких гіпотез не робилося. Потік пакетів (повідомлень, кадрів) в мережі передачі даних є сумою елементарних потоків від різних відправників. Потік пакетів від відправника до одного одержувача, що розділяється (з міркувань прискорення доставки) на кілька парціальних потоків, які проходять за різними маршрутами, на вході приймального обладнання одержувача також є сумою (парціальних) потоків .

Можна припускати, що при досить широких умовах щодо вихідних потоків сумарні потоки будуть близькі до Пуасонівських, в тому числі і найпростіших. Викладене відповідає ідеї, близької до тієї, яка майже двісті років є керівною в багатьох застосуваннях теорії ймовірностей: в теорії похибок спостережень, молекулярній фізиці, теорії стрільби та багатьох інших. А саме, вплив, що спостерігається, розглядається як сума елементарних впливів, кожний з яких є випадковою величиною, незалежною від інших; при цьому кожний з доданків надає в деякому сенсі малий вплив на суму.

Однак для теоретичної або хоча б експериментальної перевірки та обгрунтування цих припущень необхідно побудувати більш менш коректні моделі еволюції параметрів і стану процесів виявлення аномалій, прогнозу та управління навантаженням на мережу, надійністю її елементів тощо. Не вдаючись у докладні дискусії щодо порівняльного аналізу придатності до застосування тих чи інших математичних моделей, розглянемо, на нашу думку, найбільш популярні та такі моделі, за допомогою яких можна отримати асимптотичні результати у замкненій формі.

Розглянемо мережні комутаційні вузли, що потерпають від перевантажень або фізичних відмов. Припустимо, що інтервали між послідовними подіями взаємно незалежні і однаково розподілені з загальною щільністю ймовірності $w(t)$. Якщо після виявлення перевантаження маршрут перевантаженого вузла негайно заміняється резервним (вузол, що відмовив, заміняється справним), отримуємо послідовність подій, яку логічно назвати процесом відновлення.

Нехай в початковий момент часу $(t_0 = 0)$ система знаходиться в стані s_{jk}. У момент часу t_i система переходить в стан s_{mn} з імовірністю ρ_{mn}. Тоді із застосуванням теорем множення і складання імовірності знаходимо безумовну функцію розподілу повного часу знаходження системи в стані ϕ_j:

$$F_j(t) = P\{t_j < t\} = \sum_{k=1}^{5} \rho_{jk} F_{jk}(t), \qquad j,k = \overline{1,5}. \tag{1.7}$$

Імовірність станів обчислюється таким чином. Нехай $\Xi_{ij}(t)$ є умовна (інтервально-перехідна) імовірність того, що у момент часу t система знаходиться в стані ϕ_j, якщо у момент часу $t_0 = 0$ вона була в стані ϕ_i. Система, що знаходиться в початковому стані ϕ_i, може попасти в стан ϕ_j у момент часу t різними шляхами. Якщо $\phi_i = \phi_j$, вона може залишатися в стані ϕ_i протягом всього проміжку часу або, вийшовши із стану ϕ_i, щонайменше, одного разу, вона повернеться в стан $\phi_j = \phi_i$ до моменту часу t. Ці події є несумісними, імовірність цих подій буде складатися. Таким чином, приходимо до рівняння

$$\Xi_{ij}(t) = \delta_{ij} G_i(t) + \sum_{k=1}^{5} \rho_{ik} \int_0^t w_{ik}(t) \Xi_{kj}(t-\tau) d\tau, \quad 1 \le i, j \le 5, \tag{1.8}$$

де δ_{ij} – символ Кронекера;

$G_i(t)$ – імовірність того, що система не покине стан ϕ_i до моменту часу t.

Отже, першим доданком враховується імовірність події $\phi_i = \phi_j$.

Другим доданком описується послідовність переходів з ϕ_i в ϕ_k (включаючи і перехід з ϕ_i в ϕ_i, тобто в себе) до моменту τ і переходів після цього із стану ϕ_k

в стан ϕ_j за час $t-\tau$, що залишився. Імовірність проміжних переходів підсумовується по всіх відповідних проміжних станах ϕ_k, в які можливі переходи з початкового стану ϕ_i, і інтегрується по всіляких моментах часу переходу τ, $0<\tau\le t$.

Для спрощення вирішення рівнянь вигляду (1.7) використаємо метод перетворення Лапласа і отримаємо зображення функції $\Xi_{ij}(t)$ у вигляді

$$\Xi_{ij}^*(s)=\delta_{ij}G_i^*(s)+\sum_{k=1}^{5}\rho_{ik}w_{ik}^*(s)\Xi_{kj}^*(s),\quad 1\le i,\ j\le 5\quad s=\alpha+j\omega, \tag{1.9}$$

де
$$G_i^*(s)=\frac{1}{s}\left[1-w_i^*(s)\right]. \tag{1.10}$$

Отримана система алгебраїчних рівнянь (1.9) і (1.10) зв'язує перетворення Лапласа від інтервально-перехідної імовірності з основними характеристиками процесу, які задаються апріорі і коректуються в процесі спостереження. Спостережуваний напівмарковський процес характеризується фінальними (при $t\to\infty$) імовірностями станів p_j, які не залежать від початкового стану і тому є безумовними. (Відповідно, і спостережуваний процес є ергодичним.)

Ці фінальні імовірності є розв'язком системи алгебраїчних рівнянь вигляду:

$$p_j=\sum_{i=1}^{5}p_i\rho_{ij,}\quad \sum_{i=1}^{5}p_i=1,\quad j=\overline{1,5}. \tag{1.11}$$

З урахуванням умови нормування фінальної імовірності можна записати:

$$\Xi_{ij}=p_j\left\langle T_j\right\rangle\left[\sum_{i=1}^{5}p_j\left\langle T_j\right\rangle\right]^{-1}=\Xi_j, \tag{1.12}$$

де $\left\langle T_j\right\rangle$ – середні безумовні інтервали чекання в кожному із станів (1.1)-(1.5).

Застосовуючи той же метод рішення, який був використаний для здобуття рівняння (1.8), можна отримати вираз для оцінки числа переходів системи із стану j в стан k на інтервалі спостереження $(0,t)$.

Таким чином, процес є марковським лише в моменти переходу. Проте в більшості практично цікавих завдань можна ігнорувати випадковий характер часу

чекання і цікавитися лише моментами переходу, оскільки самі значення станів дають вичерпну інформацію про функціонування системи.

Як вказано у [15], оптимальною процедурою виявлення аномалій з врахуванням результатів попереднього аналізу в найзагальнішому випадку є обчислення деякого функціонала ефективності

$$\Psi\left(\mathbf{H}_A, \mathbf{H}_V\right) \underset{A,V}{\rightarrow} \max, \tag{1.13}$$

де $\mathbf{H}_V = \mathbf{V}\mathbf{V}^T$ – матриця перехідних імовірностей;

$\mathbf{H}_A$ – матриця станів системи;

$\mathbf{V}$ – вектор статистичних показників мережного вузла (кількість вхідних та вихідних пакетів в одиницю часу, середній час отримання пакетів тощо).

Матриці $\mathbf{H}_A$ і $\mathbf{H}_V$ мають розмірність $K \times K$ і $N \times N$ відповідно, тому для конкретизації функціонала (1.13) необхідно вибрати узагальнені параметри матриць і деяку універсальну міру об'єднання множини цих параметрів. Зокрема, в якості такої міри можна вибирати нормалізовані коефіцієнти варіації кожного показника.

Інформація про стан мережі періодично знімається з датчиків системи виявлення аномалій. Якщо перевищений пороговий рівень β_{0I}, приймається рішення про виявлення аномальної поведінки – перевантаження або фізичної відмови.

Наступними етапами є розпізнавання типу аномалії, побудова прогнозу розвитку аномального стану, оцінка міри загрози, вибір адекватних заходів локалізації і захисту.

Ефективність використання мережі в значній мірі визначається якістю управління в умовах перевантаження. Поки мережа завантажена незначно кількість оброблюваних пакетів дорівнює числу прийнятих. Однак, коли в мережу надходить занадто великий обсяг даних, може виникнути перевантаження, і робочі характеристики погіршуються. При надмірних завантаженнях пропускна здатність каналу або мережі може стати нульовою. Така ситуація призводить до колапсу

мережі.

Почасти це може бути пов'язано з нестачею пам'яті для вхідних буферів, але навіть якщо маршрутизатор має нескінченну пам'ять, ефект перевантаження може виявитися ще більш важким. Це пов'язано з часом очікування обробки. Якщо воно перевищує тривалість часу очікування, з'являються повторно передані пакети, що призводить до зниження корисної пропускної здатності мережі. Причиною перевантаження може бути повільний процесор або "вузьке горло" – низька пропускна здатність окремої ділянки мережі. Просте підвищення швидкодії процесора або інтерфейсу не завжди вирішує проблему – вузьке місце, як правило, переноситься в інший фрагмент мережі.

Перевантаження породжує лавинні процеси: переповнення буфера призводить до втрати пакетів, які доведеться передавати повторно або навіть кілька разів. Процесор сторони, яка передає, отримує додаткове паразитне завантаження. Все це свідчить про те, що контроль перевантаження є вкрай важливим процесом.

Слід розрізняти контроль потоку і контроль перевантаження [1]. Під контролем потоку мається на увазі балансування потоку відправника і можливості прийому і обробки одержувача. При цьому виді контролю передбачається наявність зворотного зв'язку між одержувачем і відправником. У процесі беруть участь, як правило, тільки два партнера. Перевантаження – більш загальне явище, що відноситься до мережі в цілому або до її сегменту.

Одним з поширених методів боротьби з перевантаженнями є управління зі зворотним зв'язком. Механізм управління зі зворотним зв'язком може поліпшити продуктивність мережі, скорочуючи втрати пакетів, і запобігти поширенню перевантаження.

В принципі можна послати повідомлення про перевантаження відправнику, проте при цьому перевантажений ділянку мережі навантажується ще більше. Тому завдання управління вирішується на транспортному рівні засобами протоколу TCP. При виявленні перевантаження швидкість передачі знижується шляхом зменшення розміру ковзного вікна.

Як висновок можна зауважити наступне.

У мережах із затримками сигнальної та управляючої інформації, по суті, має місце управління з запізнілим зворотним зв'язком. При неправильному урахуванні характеристик запізнювання система може втратити стійкість і перейти в незатухаючий коливальний режим, або коригування інтенсивності потоку буде здійснюватися занадто пізно. Це призводить до погіршення продуктивності мережі, особливо для додатків реального часу. Компенсація затримки зворотного зв'язку може виконуватися методами прогнозу, наприклад, з використанням моделі авторегресії і ковзного середнього (АРКС) або шляхом усереднення параметрів вікна. Другий варіант простіший, але, природно, забезпечує значно більш низьку якість сервісу.

При дуже короткому періоді реакції мережного вузла система управління буде отримувати послідовність суперечливих інформаційних сигналів. Система буде перебувати в стані незатухаючих коливань і не прийде в стабільний стан. З іншого боку, якщо період реакції буде занадто довгим, механізм управління станом реагуватиме занадто повільно, щоб взагалі принести якусь справжню користь. Щоб мати відповідну якість процесу управління, потрібно застосовувати певний спосіб адаптації, але правильний вибір постійних часу – це нетривіальне питання, яке планується розглянути в майбутньому. Позитивного результату можна досягти шляхом варіації значень тайм-аутів, зміни політики повторної передачі пакетів. У деяких випадках результат може бути отриманий зміною схеми буферизації.

Управління зі зворотним зв'язком широко використовується в архітектурі інтегрованих служб (Integrated Service Architecture – ISA) для підтримки служб з різними рівнями якості сервісу (Quality of Service – QoS) в Інтернет і в інших об'єднаних мережах [1, 4, 10].

2 РЕГУЛЮВАННЯ ТРАФІКА І КОНТРОЛЬ ПЕРЕВАНТАЖЕННЯ НА ОСНОВІ ПРОГНОЗУЮЧИХ НЕЙРОННИХ СИСТЕМ

2.1 Методи регулювання і формування трафіка

Трафік – це послідовність або потік пакетів (кадрів, комірок, повідомлень) з випадковими характеристиками, тобто як дискретний випадковий процес. Саме статистичні характеристики потоку пакетів як заявок на обслуговування в мережному або термінальному вузлі становлять інтерес для рішення задач управління та оптимізації мереж.

QoS (англ. Quality of Service — якість обслуговування) — цим терміном в області комп‘ютерних мереж називають ймовірність, що мережа зв‘язку відповідає заданій угоді про трафік або ж у ряді випадків, неформальне позначення ймовірності того, що пакет пройде між двома точками мережі.

Для більшості випадків якість зв‘язку визначається чотирма параметрами.

1) Смуга пропускання (Bandwidth), описує номінальну пропускну здатність середовища середи передачі інформації, визначає ширину каналу. Вимірюється в Bit/s (Bps), Kbit/s (Kbps), Mbit/s (Mbps).

2) Затримка при передачі пакету (Delay), вимірюється в мілісекундах.

3) Коливання (тремтіння) затримки при передачі пакетів — джиттер (Jitter).

4) Втрата пакетів (Packet loss). Визначає кількість пакетів, втрачених згубити в мережі під час передачі.

При передачі даних в комп’ютерних мережах виділяють наступні типи сервісів:

– *Негарантована доставка (best effort service),* яка полягає в забезпеченні зв’язності вузлів мережі без гарантії доставки пакету адресату; при цьому відкидання пакету може відбутись при переповненні буферу вхідної або вихідної черги будь-якого комунікаційного вузла.

– *Диференціальне обслуговування (differentiated service),* яке передбачає розподілення трафіку на класи згідно до вимог якості обслуговування в мережі, що доцільно використовувати в мережах з інтенсивним трафіком.

– *Гарантоване обслуговування (garanted service),* яке передбачає резервування мережевих ресурсів по всьому маршруту передачі даних для обслуговування різних потоків з необхідними параметрами.

В свою чергу в диференціальному обслуговуванні для забезпечення QoS виділяють такі технології:

– інтегровані послуги (IntServ);

– диференційовані послуги (DiffServ).

IntServ використовується для обслуговування одиночних потоків, яким може надаватися 3 види послуг: гарантовані, з керованою завантаженістю та послуги з максимальними зусиллями, з яких перші два – є обов'язковими:

– гарантовані послуги пов'язані з визначенням максимального порушення синхронізації та гарантуванням певної ширини полоси пропускання;

– кероване навантаження, при якому потоку, що обслуговується, надається постійний рівень послуг;

– послуги з максимальними зусиллями майже не використовуються в сучасних мережах і призначені для додатків з текстовим інтерфейсом.

До переваг протоколу IntServ можна віднести гарантовану величину затримки: значення часу затримки пакетів в мережі може бути задано абсолютною верхньою границею, тобто при використанні одиночних потоків відсутні затрати на налаштування QoS, що дозволяє збільшити швидкість реакції та скоротити об'єм додаткового трафіку. Суттєвим недоліком IntServ є низька масштабованість. Продуктивність IntServ залежить від кількості потоків, що обробляються, тобто таку технологію неможливо реалізовувати в мережах з великою кількістю користувачів.

Послуги DiffServ надаються в області мережі, яку називають доменом. DS-домен (Differentiated Services Domain) – це цілісна множина маршрутизаторів, які підтримують архітектуру DiffServ та забезпечують необхідний QoS.

При використанні технології DiffServ процес передачі даних включає такі ключові етапи:

– визначення рівня сервісу, який буде надаватися пакетам даного потоку;

– встановлення DS-байта (DiffServ— DS), значення якого залежить від належності певному потоку та від рівня сервісу.

Перевагою даної технології у порівнянні з попередньою є забезпечення масштабованості обслуговування в мережі без необхідності запам’ятовувати стани кожного потоку.

У випадку DiffServ відправник і отримувач не обмінюються інформацією про вимоги до якості обслуговування, що зменшує часові затрати на визначення шляху передачі. Враховуючи всі переваги моделі DiffServ, можна сказати, що її доцільно використовувати в магістральних та високошвидкісних областях комп’ютерних мереж.

В таблиці 2.1 наведена порівняльна характеристика цих двох технологій.

Таблиця 2.1 Порівняльна характеристика технологій IntServ та DiffServ

Параметри	Досліджуванні технології	
	IntServ	DiffServ
Метод забезпечення QoS	Резервування	Пріоритизація
Число класів, що обслуговуються	3	3
Необхідність використання додаткових протоколів	+ (протокол RSVP)*	-
Вимоги до продуктивності маршрутизаторів	Високі	Низькі
Кількість потоків, що обробляються	Обмежена	Необмежена
Гарантованість забезпечення якості	Висока	Середня

***** RSVP – протокол резервування мережевих ресурсів (Resource ReSerVation Protocol)

Боротьба з перевантаженням – забезпечення транспорту потоку трафіка відповідно до заявленої якості сервісу *QoS*, тобто без збільшення затримок, числа помилково переданих біт, числа повторних передач і т.д. Задача зважується на всіх ділянках мережі, термінальних і мережних комутаційних вузлів. Аналізуються й коректуються процеси зберігання й пересилання, ураховуються інші фактори, через які знижується пропускна здатність мережі в цілому або окремому сегменті.

Управління потоком – задача передачі даних між двома термінальними вузлами – відправником і одержувачем. Вона полягає в узгодженні швидкості передачі відправника зі швидкістю, з якою одержувач у стані приймати та обробляти потік трафіка.

Управління потоком і боротьба з перевантаженням – пов'язані задачі [1, 4]. В обох випадках використовується зворотний зв'язок (ЗЗ) між учасниками процесу. Різниця між цими поняттями полягає в наступному. При управлінні потоком є ЗЗ між відправником й одержувачем, при боротьбі з перевантаженням – ЗЗ у вигляді спеціальних повідомлень, що посилають різним відправникам з метою регулювання швидкості передачі щоб уникнути заторів. Методи боротьби з перевантаженням діляться на наступні класи (рис. 2.1).

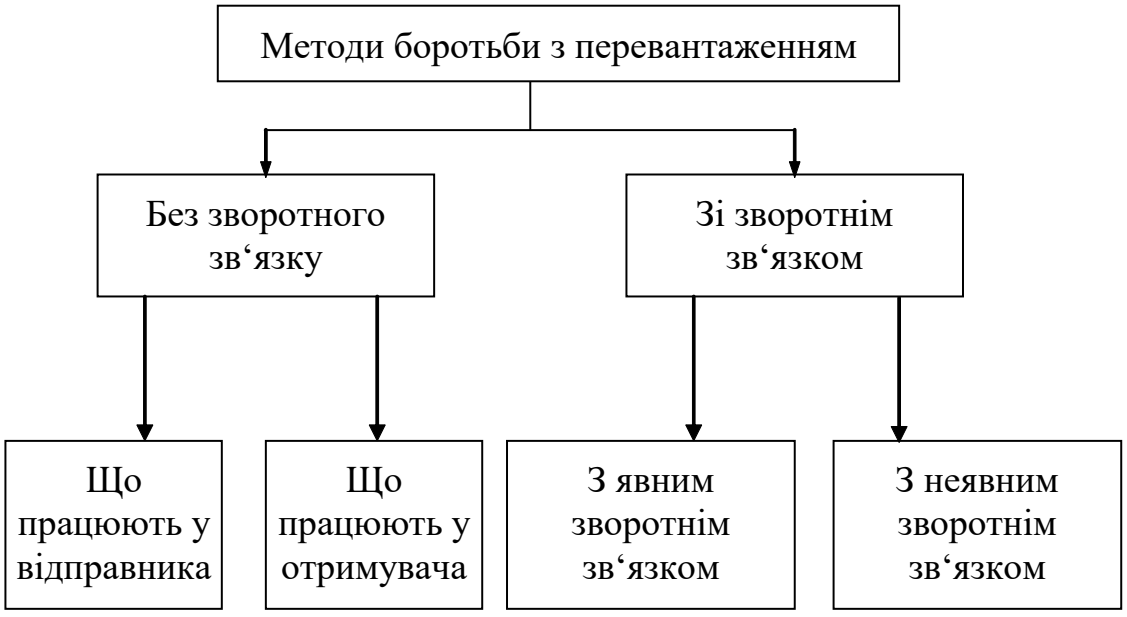

Рисунок 2.1 – Методи боротьби з перевантаженням

У методах з явним зворотнім зв‘язком від точки виникнення перевантаження у зворотному напрямку посилають повідомлення – попередження про затор. У методах з неявним ЗЗ джерело дійде висновку про затор самостійно, ґрунтуючись на результатах локального аналізу (наприклад, по зростанню затримок одержання квитанцій про доставку або по зростанню числа вимог про повторну передачу пакетів).

Наявність перевантаження означає, що інтенсивність потоку перевищує ресурси даної ділянки маршруту або сегмента мережі. Очевидно, із цієї ситуації є тільки два виходи: знизити навантаження або збільшити ресурси.

У комп'ютерній мережі є набір ресурсів, які можуть використовуватися конкуруючими користувачами (повідомленнями). Ресурси таких мереж підрозділяються на два основних класи:

– основні: розмір буфера, пропускна здатність мережі, час обробки даних.
– другорядні: простір для імені, таблиці входів, логічні канали й т.д.

Набір ресурсів завжди обмежений, що є причиною конфлікту між користувачами системи. Конфлікти можуть призвести до зниження продуктивності системи, внаслідок чого виникають певні труднощі в роботі, і пропускна здатність буде падати, що є типовою поведінкою “конкуруючих” систем. Навіть при наявності найкращої процедури маршрутизації під час перевантаження в мережі неминуче зниження пропускної здатності (продуктивності). Мережі не обслуговують трафік зі стрибкоподібними змінами інтенсивності без попереднього контролю. Існують правила, за допомогою яких відбувається управління прийомом трафіка ззовні і координація потоку в мережі.

Більшість мереж добре працюють при невеликому навантаженні, а проблеми виникають при зростанні завантаження мережі [17]. Перевантаження може викликати втрату інформації, затримки при передачі даних і дисперсію затримок. Тому запобігання перевантажень і управління ними є основною проблемою при розробці та експлуатації мереж.

2.2 Сучасний стан управління перевантаженнями в комп'ютерних мережах

Розглянемо сучасний стан управління перевантаженнями в комп'ютерних мережах (без установлення з'єднання та орієнтованих на з'єднання).

У результаті ранніх досліджень у комп'ютерних мережах були розроблені мережі передачі даних із проміжним нагромадженням без резервування ресурсів. Запропоновано модель обміну даними без попередньої установки прямого з'єднання. Ці мережі піддаються перевантаженням, тому що ні кількість користувачів, ні їхні робочі навантаження не регулюються.

Власне кажучи, ефективність при передачі зі статистичним мультиплексуванням і спільним використанням ресурсів мережі зменшується через можливість виникнення перевантажень. Ця проблема була виявлена досить швидко, і був запропонований ряд схем керування перевантаженням [1, 3, 7].

Останнім часом інтерес до питання контролю перевантажень знову зріс. Перевантаження – це втрата ефективності у зв'язку зі збільшенням мережного завантаження. Отже, управління перевантаженням є набором механізмів, які запобігають або зменшують таке погіршення. Тобто, для управління перевантаженням кожному користувачеві надаються механізми для визначення ефективності роботи мережі.

У строгій схемі із прогнозуванням механізм управління перевантаженнями повинен резервувати ресурси мережі так, щоб доступність ресурсу виразно гарантувала передачу даних. В активній схемі джерела даних повинні контролювати й реагувати на зміни в стані мережі, щоб запобігти перевантаженню. В обох типів управління є свої плюси й мінуси. З управлінням і прогнозуванням користувачам можна гарантувати, що втрати ефективності не будуть мати місця. З іншого боку, щоб гарантувати це, кількість користувачів повинна бути обмежена або керована, і якщо механізм управління джерелами даних не розроблений належним чином, це може призвести до зниження ефективності використання й іноді до перевантаження мережі. Активне управління із прогнозуванням дозволяє більш гнучко розподіляти ресурси. Тому що користувачам звичайно не дають

гарантій, рівень ефективності ресурсів мережі може бути динамічно розподілений. Однак завжди є шанс, що через сплеск інтенсивності трафіка відбудеться перевантаження мережі, що призведе до погіршення робочих характеристик.

У більшості мереж співіснують управління із прогнозуванням і пасивне управління зі зворотним зв'язком. Мережа може підтримувати, наприклад, два типи користувачів: гарантуємий (або високий пріоритет) і сервіс типу "*Best effort*" з максимальною ефективністю (або низький пріоритет) [1, 3]. Гарантійні користувальницькі служби дають гарантію якості обслуговування і забезпечення ресурсів, зарезервованих для них.

Користувачі, яким надається сервіс типу "*Best effort*", не мають установлених гарантій й одержують ресурси, які залишаються від привілейованих користувачів. Під час перевантаження мережі користувачі змагаються за ресурси, збережені в загальному фонді. Крім того, мережа може резервувати деяку мінімальну величину ресурсів для кожного користувача із сервісу типу "*Best effort*".

Найбільш істотними причинами, виникнення перевантаження в обчислювальних мережах є наступні.

1. Обмежений об'єм каналу передачі або добутки затримки на смугу: швидкість лінії зв'язку, помножена на час розповсюдження пакета до адресата і назад, визначає обсяг даних, при якому мережа використовується повністю. Час розповсюдження пакета до адресата і назад обмежується швидкістю розповсюдження через мережу. Отже, від вихідної пропускної здатності й швидкості з'єднання залежить кількість втрат даних для кожного підключення. Якщо воно перевищує розмір буфера, можлива втрата інформації, що, відповідно, приводить до втрати ефективності. Таким чином, більші затримки і недостатня пропускна здатність викликають проблеми при активному контролі й можуть призвести до перевантаження.

2. Невідповідність швидкодії: якщо комутатор з'єднує високошвидкісну лінію з низькошвидкістною, то при пульсуючому трафіку і відправленні даних на максимальній швидкості ресурс буфера може бути вичерпаний, що веде до втрат

пакетів при передачі. Це створює перевантаження для додатків, чутливих до затримок й/або втратам пакетів.

3. Недоліки топології: якщо кілька вхідних ліній одночасно відправляють дані через комутатор до єдиного шляху відправлення, лінія на виході може бути перевантажена, що приводить до великої черги, затримкам, і можливому перевантаженню чутливого до затримки трафіка. Це є особливим випадком проблеми невідповідності швидкості.

4. Неправильне функціонування. Перевантаження може бути викликане неконтрольованими джерелами, які посилають потоки пакетів з нерегульованою інтенсивністю в мережу. Комп'ютерна мережа повинна здійснювати захист від такої неправомірної поведінки.

5. Відсутність контролю динаміки: зі збільшенням швидкодії мережі також змінюється динаміка завантаженості мережі. Якщо черги зростають швидше, перевантаження буде виникати все частіше. Наслідки можуть бути катастрофічними, оскільки мережа повністю зупиняється, а відновлення працездатності йде набагато повільніше, ніж падіння.

У комп'ютерних мережах, де джерела не резервують ресурси, управління там, де це необхідно, повинне бути активним. Активна схема управління перевантаженнями може бути реалізована у двох точках: у комутаторах, де відбувається перевантаження, і в джерелах, які управляють відправленням пакетів у мережу. Як правило, комутатор використовує деякий показник (наприклад, переповнення буферів), щоб визначити початок перевантаження, і неявно або явно повідомляє цю проблему джерелам, які зменшують вхідний трафік. Сценарії управління включають наступні очевидні задачі.

1. Виявлення/запобігання перевантаження: вибір оптимального методу за критеріями швидкодії та вірогідності, які, очевидно, є суперечливими.

2. Повідомлення про перевантаження: передача інформації джерелам з мінімальними затримками.

3. Боротьба з перевантаженням: вибір алгоритмів роботи мережних вузлів.

4. Виявлення джерел перевантаження.

5. Управління потоком даних джерел.

6. Примус, якщо деякі джерела ігнорують сигнали управління.

Залежно від зробленого вибору при відповіді на кожне питання застосовується та або інша схема управління перевантаженнями.

У літературі [2, 3] наведені численні схеми управління перевантаженнями в мережах без установлення з'єднання та у мережах, орієнтованих на з'єднання.

2.3. Методи регулювання і формування трафіка

Термінами “регулювання” і “формування” (в англомовній літературі “policing” й “shaping” відповідно [2]) визначаються наступні операції над потоком трафіка.

Регулювання трафіка (рис. 2.2) полягає в тім, що пакети, при надходженні яких перевищується межа C_{max} пропускної здатності мережі, відкидаються або позначаються як надлишкові (штрих-пунктирна лінія). Їм присвоюється нижчий пріоритет.

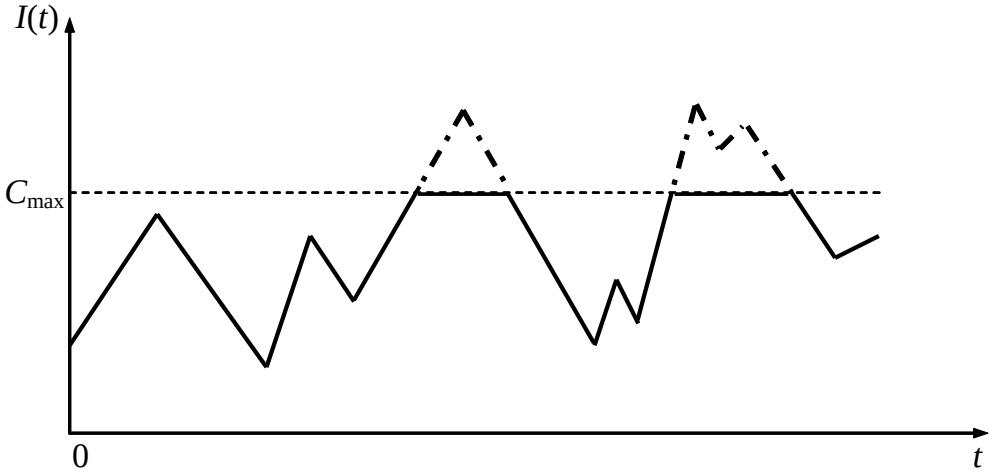

Рисунок 2.2 – Регулювання трафіка

Формування трафіка (рис. 2.3) полягає в тому, що згладжуються сплески інтенсивності потоку пакетів, тобто усереднюється період проходження пакетів у потоці (потік зі згладженою інтенсивністю відокремлений лінією нижче пунктирної C_{max}).

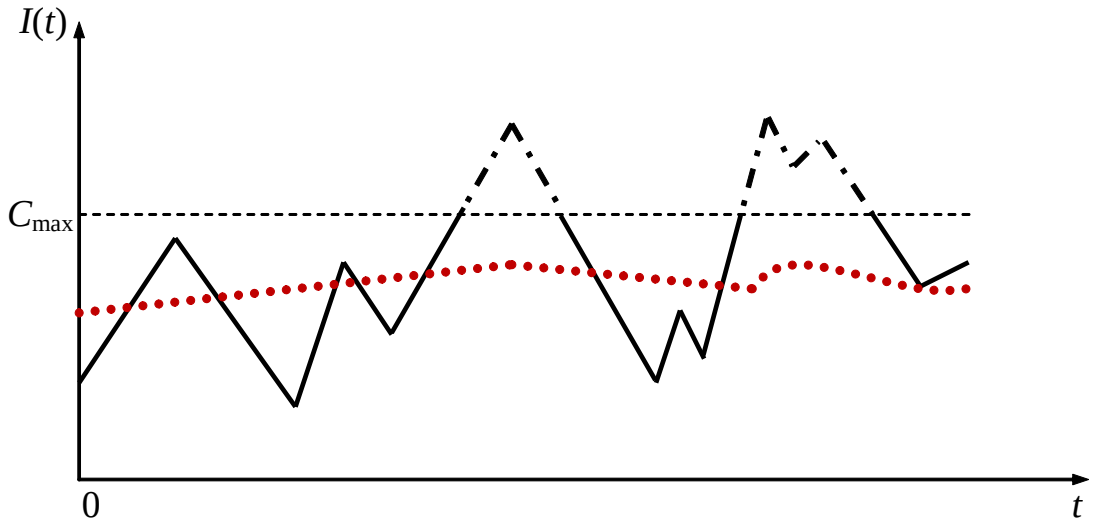

Рисунок 2.3 – Формування трафіка

У принципі навіть якщо комп'ютер посилає пакети з випадковим періодом, можна регулювати період потоку безпосередньо на вході в мережу доступу. Для формування трафіка застосовують відомі алгоритми так називаного дірявого або маркерного відра [11, 16].

Джерелами сплесків інтенсивності трафіка є підсумовування потоків від різних джерел, різнорідність трафіка (мова, відео, дані), переданого по тому самому маршруту, широкомовні шторми, групові розсилання [1, 2] і багато інших факторів. Зокрема, через вплив деяких факторів статистика трафіка набуває самоподібний характер [16, 17]. Для того, щоб обмежити інтенсивність трафіка, що входить у мережу або виходить з неї, регулювання трафіка часто реалізується на границі мережі [3, 6, 10].

В ідеалі, звичайно, бажано було б мати потік з постійною інтенсивністю, тобто з постійним періодом проходження пакетів. У термінах теорії масового обслуговування такий потік називається регулярним. Мережа являє собою систему масового обслуговування з різними розподілами часу обслуговування, але з регулярним вхідним потоком. Відповідно й довжина черги в буфері комутаційного вузла буде рости повільніше в порівнянні з пуасоновським вхідним потоком або в порівнянні самоподібним вхідним потоком.

Реально при згладжуванні однаково мають місце залишкові коливання інтенсивності трафіка. Для зменшення цих залишкових коливань застосовують різні комбіновані методи багатоступінчастого формування регулярного потоку. Наприклад, регулювання і формування можуть застосовуватися спільно. Пропонується використовувати послідовно маркерне і діряве відра [11]. Однак у тому ж джерелі відзначається, управління такою схемою виявляється досить складним і громіздким. Тому необхідно шукати інші методи підвищення ефективності формування трафіка.

Одна із проблем мереж – це ефективний контроль трафіка. Спостереження за потоком трафіка називається *політикою трафіка* [1, 4]. Методи рішення проблеми контролю трафіка ґрунтуються на застосуванні політики, що дозволяє гармонійно розподілити інформаційний потік. Провідні розроблювачі мережного обладнання намагаються втілити ідею використання політики контролю трафіка в технічні рішення. Як правило, реалізуються методи QoS (Quality of Service) і CoS (Class of Service), які досить тісно пов'язані між собою.

У тих випадках, коли об'єм трафіка або затримки в мережі можуть зростати в межах 10–30%, розумно буде задіяти QoS/CoS-політику як спосіб контролю трафіка. Якщо організація планує використовувати відеоконференцію або систему групової роботи, у мережі обов'язково потрібно закріпити необхідний ресурс за певними додатками. Розроблювачі пропонують схеми, що дозволяють ефективно застосувати політику QoS/CoS у мережі.

У результаті застосування політики пріоритетів створюються марковані потоки трафіка. Алгоритм політики, наприклад, може визначити, що пакети, які згенеровані додатками, чутливими до затримок, мають перевагу по доставці перед пакетами інших типів – обміну файлів або електронної пошти. Якщо ж виникне ситуація, коли частина пакетів у результаті перевантаження буде відкинута, то в першу чергу це станеться з пакетами з низьким пріоритетом. Поряд із забезпеченням достатньої смуги пропущення схема пріоритетів дозволяє здійснити доставку необхідного трафіка в пункт призначення.

Одним з раціональних шляхів управління трафіком (що вимагає значно менших витрат, ніж реалізація політики QoS) може служити динамічне управління ресурсами мережі – розділення смуги пропущення на кілька частин для конкретних потреб. У випадку, коли зростає навантаження, пакети починають розміщатися в буферній пам'яті мережних вузлів, що приводить до затримок. Якщо буфер маршрутизатора (комутатора) переповняється, деякі пакети взагалі можуть бути загублені. Однак очевидне, на перший погляд, рішення – збільшення об'єму буферної пам'яті – не приводить до поліпшення ситуації з перевантаженнями, а іноді може навіть погіршити її. По-перше, “нетерплячі” пакети з малим часом таймаута просто залишають чергу. По-друге, “терплячі” пакети можуть вийти із черги із затримкою, що також перевищила час таймаута. І в тому, в іншому випадку відбудуться повторні передачі пакетів, що дасть додаткове навантаження на мережу.

Тому подальшим розвитком політики трафіка є застосування диференційованих та інтегрованих послуг [2] – динамічний розподіл розташовуваної смуги пропущення каналу передачі між користувачами з урахуванням їх пріоритетів замість направлення всіх потоків по одному або декількох локально-оптимальних маршрутах. При такому підході вдається уникати періодичних місцевих перевантажень маршрутів, які були визначені як оптимальні на поточному етапі, і по яких всі користувачі направили свій трафік. З урахуванням цієї обставини алгоритм вибору маршрутів за критерієм рівномірного або зваженого завантаження ліній передачі може давати глобальний оптимум на більш тривалому етапі функціонування мережі або її сегмента.

На рис. 2.4 представлена поетапна схема реалізації системи QoS в телекомунікаційних мережах та її сегментах

Таким чином, для реалізації політики трафіка з диференційованими послугами необхідно вирішувати задачу розподілу потоків між локально-оптимальними або квазіоптимальними маршрутами в реальному масштабі часу. Однак спочатку доцільно проаналізувати процес формування сумарних потоків у парціальних каналах загальної мережі передачі даних.

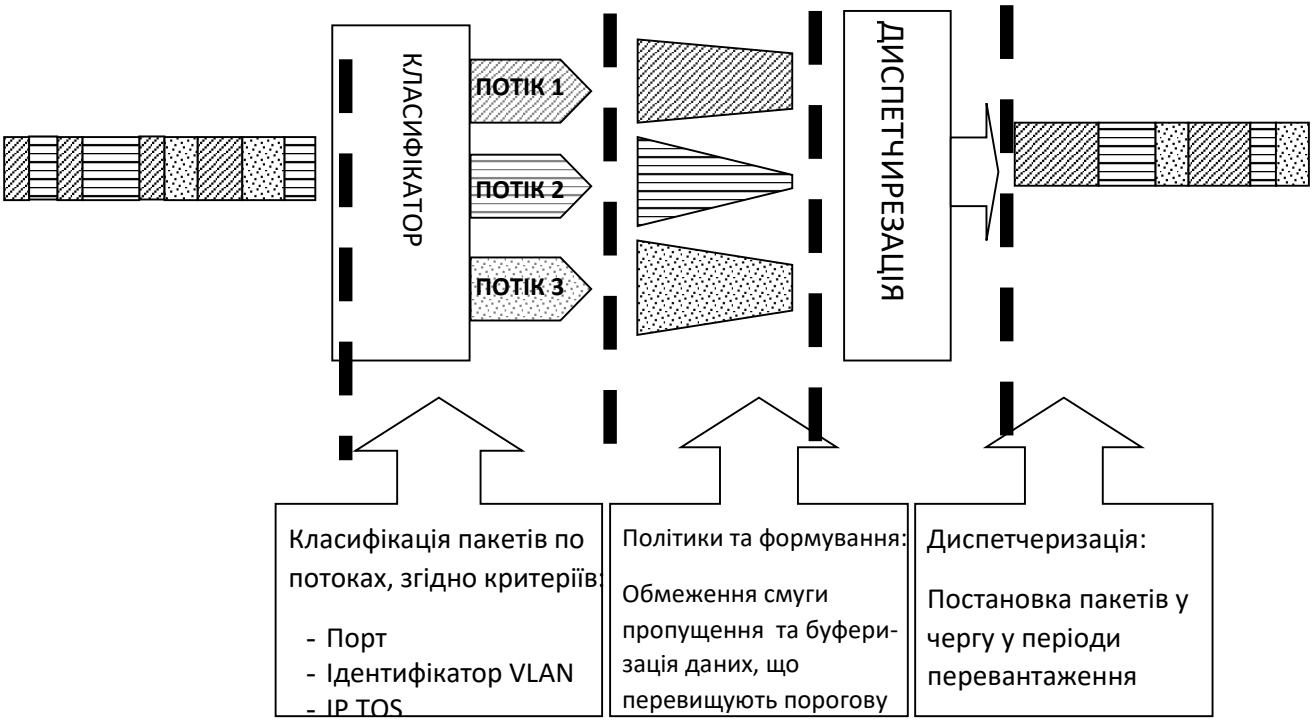

Рисунок 2.4 – Спрощена схема етапів QoS

Підсумовуючи, зауважимо наступне.

Задачі управління мережею і методи боротьби з перевантаженнями – дві частини загальної проблеми якості сервісу і функціонування мережі із заданими експлуатаційно-технічними характеристиками. Вони нерозривно зв'язані один з одним і вирішуються в комплексі.

Одним з ефективних методів боротьби з перевантаженнями мереж або їхніх окремих сегментів є формування й/або регулювання (упорядкування) трафіка – приведення потоків даних, по можливості, до регулярного виду. Якщо трафік, що входить через формувач у сегмент мережі, має довільний розподіл (пуасонівський, Ерланга, самоподібний), то трафік на виході формувача повинен мати приблизно постійний період проходження пакетів. Для рішення цієї задачі застосовуються формувачі типу «діряве відро» або «маркерне відро». Показана доцільність застосовувати адаптивних методів (зокрема, нейронних систем) управління параметрами формувачів трафіка.

2.4 Контроль перевантаження на основі прогнозуючих нейронних систем

Управління надійністю мережі в основному базується на аналізі чутливості систем передачі інформації до змін ключових параметрів ефективності мереж: продуктивності, пропускної здатності, затримки інформації, характеристик якості послуг

Вирішення задач адаптивного управління телекомунікаційними мережами, як складними системами, пов’язане із значною складністю обчислювального процесу, що призводить до суттєвих обмежень в точності управління внаслідок допустимих затримок управляючої інформації. Одним із шляхів вирішення цієї задачі є використання нейронних систем обробки інформації.

Найповніше визначення і уявлення про нейронні системи і мережі приведене в [13]. Нейронна система – це розподілений паралельний процесор, що складається з елементарних одиниць обробки інформації – нейронів, які накопичують експериментальні знання і надають їх для подальшої обробки. Знання поступають в нейронну систему з навколишнього середовища і використовуються в процесі навчання, а для накопичення знань використовуються зв’язки між нейронами.

Нейронні системи мають багато цікавих можливостей, такі як здатність оброблювати шум і неповні дані, висока відмовостійкість, яка дозволяє мережам працювати нормально з декількома пошкодженими нейронами чи зваженими з'єднаннями в мережі, також вони здатні функціонувати в режимі реального часу в зв'язку з властивим їм паралелізмом. З цими корисними можливостями нейронні системи набули дуже широкого застосування в розпізнаванні образів, обробці зображень, розпізнаванні голосу, адаптивному управлінні.

В багатьох випадках процес навчання нейронної системи закінчується з досягненням конвергентних критеріїв. Після завершення певного етапу навчання система використовується для отримання і зберігання інформації з допомогою процесу повторного запиту. Отримані в розглянутому процесі результати використовуються при розробці нових алгоритмів навчання нейронних систем і

мережевих архітектур, а також з метою розширення області їх застосувань. Здатність багатошарових прогнозуючих нейронних систем до приблизного невідомого довільного відображення була добре досліджена в літературі. В [14] було показано, що багатошарові прогнозуючі системи як з декількома так і з одним прихованим шаром і, відповідно, неявним прихованим шаром функції активації здатні наближувати з довільною точністю довільну функцію і її похідні.

Успішно навчена нейронна система направляє вхідний вектор $\overline{X}$ з *n*-вимірним простором до вихідного вектору $\overline{Y}$ в *m*-вимірному просторі. Цей процес можна описати так:

$$\overline{Y} = f(\overline{X}), \tag{2.1}$$

де $\overline{Y} = (y_1, y_2, ..., y_m)$ і $\overline{X} = (x_1, x_2, ..., x_n)$.

Перша похідна для будь-якої пари "функція-аргумент" в (1) $\partial y_j / \partial x_i$ показує швидкість зміни y_j по відношенню до зміни x_i. Отже, вона надає інформацію про те, наскільки чутливий вихід y_j по відношенню до входу x_i. Іншими словами, чим більше абсолютне значення $\partial y_j / \partial x_i$, тим важливішим є x_i по відношенню до y_j. Тоді вхідні змінні x_i, $i = 1, 2, ..., n$ можна рангувати в порядку важливості, відповідно до значень $\left[\partial y_j / \partial x_i\right]$. Похідну $\partial y_j / \partial x_i$ можна розвивати з допомогою структури прогнозуючої нейронної системи, що навчається.

Розглянемо деякі способи для оцінювання похідних виду $\partial y_j / \partial x_i$, використовуючи нейронні структури. Методи обчислення чутливості з використанням нейронних структур розподіляють, як правило, на два класи:

1. Прямі методи, в яких похідні першого порядку (і будь-які похідні вищих порядків, за бажанням) обчислюються безпосередньо в нейронній системі.

2. Методи пертурбації, які полягають у введенні незначних завад (збурень) на кожному вході нейронної системи як в одному так і в іншому напрямках з подальшим усередненням результуючих відхилень на кожному виході і в обох напрямках.

Розглянемо прямий метод для обчислення вихідних чутливостей в залежності від змін на входах для багатошарових прогнозуючих нейронних систем з диференційними нелінійними функціями активації. Ці чутливості можна використати як основу для визначення взаємозв'язків між входами і виходами складної системи. Крім того, модель обробки прогнозуючої нейронної системи може легко забезпечити ефективні впровадження методів на основі градієнту для оптимізації обробки.

Важливою рисою розглянутого способу є те, що чутливості обчислюються однократно в кожному шарі починаючи з вихідного шару і продовжуючи у зворотному напрямку до вхідного шару багатошарової нейронної системи. Для спрощення аналізу введемо допущення, що всі активаційні функції одного виду і диференційовані. Також припустимо, що не існує обхідних з'єднань в нейронній системі (тобто, прямі зв'язки можуть існувати тільки від одного шару до наступного найближчого). Зауважимо також, якщо припущення про диференційованість активаційної функції є фундаментальним, ти інші припущення – ні. Однак, розглянутий тут метод можна легко модифікувати для систем без вказаних обмежень.

Розглянемо більш детально структуру прогнозуючої багатошарової нейронної системи, яка складається з вхідного шару, N прихованих шарів і вихідного шару (рис. 2.5).

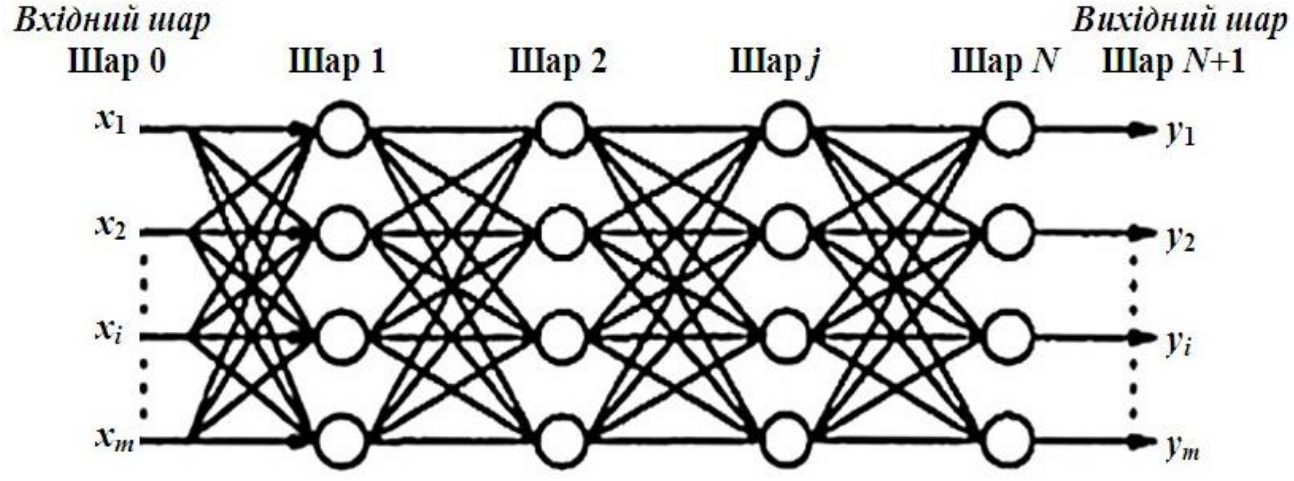

Рисунок 2.5 – Багатошарова нейронна система прямого поширення

Представлений вхідним шаром (шар 0) вектор входу передається до першого прихованого шару (шар 1). Тоді обчислюється зважена сума вхідного вектора на кожному нейроні в першому прихованому шарі на основі ваг зв'язків між вхідним шаром і першим прихованим шаром. Потім ця сума використовується для обчислення виходу кожного нейрона, застосувавши сигмовидну активаційну функцію, н-д, (2), (5) або (7). Далі нейрони в першому прихованому шарі передають свої значення до наступних шарів, і так до вихідного шару (шар N+1). Значення функцій на виходах нейронів в прихованих шарах і вихідному шарі обчислюються таким же чином, як і в першому прихованому шарі, з використанням ваг зв'язків між шарами і логістичної сигмовидної активаційної функції.

Вихідні чутливості першого порядку обчислюються застосуванням правила простої зворотної ланцюгової часткової диференціації. По-перше, вихідні чутливості залежать від змін в значеннях нейронів шару N, обчислюються, потім здійснюється зворотне ланцюгове обчислення вихідних чутливостей для різних значень змінних на вході нейронної системи. Це робиться з використанням активаційної функції, наступним чином:

– для нейронів в шарі N:

$$\frac{\partial y_k}{\partial h_i^N} = \frac{\partial y_k}{\partial net_k^{N+1}} \cdot \frac{\partial net_k^{N+1}}{\partial h_i^N} = \alpha y_k (1 - y_k) \cdot w_{ik}^N, \forall i,k; \tag{2.2}$$

– для нейронів в решті прихованих шарів (шари j, $j = N–1, \ldots, 1$):

$$\frac{\partial y_k}{\partial h_i^j} = \sum_l \frac{\partial y_k}{\partial h_l^{j+1}} \cdot \frac{\partial h_l^{j+1}}{\partial net_l^{j+1}} \cdot \frac{\partial net_l^{j+1}}{\partial h_i^j} = \sum_l \frac{\partial y_k}{\partial h_l^{j+1}} \cdot \alpha h_l^{j+1} (1 - h_l^{j+1}) \cdot w_{il}^j, \forall i,k; \tag{2.3}$$

– для вхідного шару (шар 0):

$$\frac{\partial y_k}{\partial x_i} = \sum_l \frac{\partial y_k}{\partial h_l^1} \cdot \frac{\partial h_l^1}{\partial net_l^1} \cdot \frac{\partial net_l^1}{\partial x_i} = \sum_l \frac{\partial y_k}{\partial h_l^1} \cdot \alpha h_l^1 (1 - h_l^1) \cdot w_{il}^0, \forall i,k, \tag{2.4}$$

де y_k – вихід k-го нейрону у вихідному шарі (шар N+1);

h_k^j – вихід k-го нейрону в шарі j, j=1, ... N; x_i є i-м входом нейронної системи;

$net_k^j = \sum_{i=1}^{m} w_{ik}^{j-1} h_i^{j-1}$ – зважена сума на вході i-го нейрону в j-му шарі, j = 1, …N + 1;

w_{ik}^j – ваги зв'язку між i-м нейроном в шарі j і k-м нейроном в шарі j+1 , j = 0...N.

$g(net_k^j) = \dfrac{1}{\left(1 + e^{-\alpha(net_k^j)}\right)}$ – активіаійна функція *k*-го нейрону в *j*-му шарі.

На рис. 2.6 показана структура нейрону (*k*-й нейрон *j*-го шару).

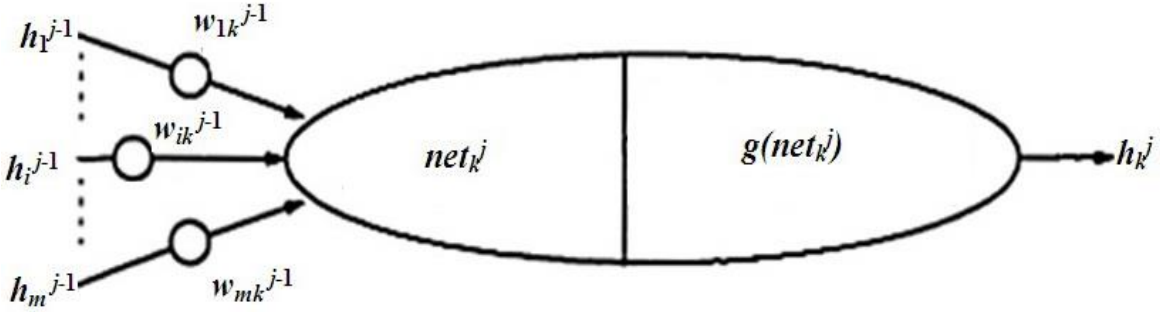

Рисунок 2.6 – Структура нейрону

Похідні чутливості другого і навіть вищих порядків, функції виходів нейронної системи можуть бути обчислені, використовуючи правило зворотного ланцюгування, подібно до похідних чутливості першого порядку.

Зазначені вище вирази (2.2), (2.3) і (2.4) показують, що часткова похідна $\partial y_k / \partial x_i$ залежить не тільки від вивченої нейронною системою інформації, яка зберігається розподілено в з'єднаннях w_{ik}^j, але також від активаційних функцій нейронів.

Запропонована модель обробки на основі прогнозуючої нейронної системи забезпечує ефективну реалізацію градієнтних методів оптимізації систем.

3 КОНТРОЛЬ ПЕРЕВАНТАЖЕННЯ МЕРЕЖІ НА ОСНОВІ НЕЙРОННОЇ МОДЕЛІ

3.1 Контроль перевантаження за показником довжини черги

В ряді робіт [21, 22] для запобігання перевантаженням в комп'ютерних мережах досліджуються способи, основані на збільшенні об’єму пам'яті вхідних буферів маршрутизаторів. Однак, при такому підході виникає проблема розбухання буферу (Bufferbloat). Зі зростанням об’єму буферної пам’яті збільшується кількість необроблених пакетів, а головне – збільшується час очікування їх обробки. Це може привести до перевищення допустимих норми тривалості тайм-ауту, що призводить до подальшого зниження корисної пропускної здатності мережі. Зазвичай, це може спричинити виникнення лавинного процесу: переповнення буфера призводить до втрати пакетів, які доведеться передавати повторно або навіть кілька разів. Таким чином, обчислювальний вузол маршрутизатора-відправника отримує надлишкове паразитне завантаження, що може призвести до збільшення негативних наслідків, пов’язаних з перевантаженням [23].

Найбільш ефективні способи і алгоритми контролю перевантаження реалізуються за принципом явного зворотного зв’язку [1, 4, 18.

Одним із поширених способів управління потоком даних за принципом зворотного зв’язку є двійковий бітовий контроль [19, 20]. Зворотний зв'язок реалізується між вузлами, через які здійснюється кожен окремий сеанс інформаційного обміну. Зауважимо, що схема зворотного зв'язку може бути реалізована між двома суміжними вузлами, або між двома кінцевими вузлами, між якими впродовж маршруту інформаційного обміну є декілька транзитних вузлів, тобто, контроль із кінця в кінець [19, 20]..

Мережа надає бінарну індикацію про те, чи зустрілись перевантаження чи їх загрози впродовж з’єднання. Значення "0" індикатора перевантаження встановлюється в заголовку пакета джерелом даних, тобто на стороні вузла-

відправника. Будь-який вузол вздовж шляху може встановити (або підтвердити) значення біта перевантаження в пакеті в "1", щоб показати наявність або загрозу перевантаження в цьому вузлі. Кінцева система (вузол-отримувач) контролює біти перевантаження отриманих пакетів і повертає повідомлення зворотного зв'язку до джерела з інформацією про наявність перевантажень. Якщо повідомлення зворотного зв'язку повідомляє про відсутність перевантаження, тоді джерело інформації збільшує вхідний трафік (темп надходження пакетів, потік даних) – адитивний зворотний зв'язок. Якщо повідомлення зворотного зв'язку повідомляє про наявність чи загрозу перевантаження, вхідний трафік від джерела інформації зменшується – множинний (мультиплікативний) зворотний зв'язок.

Виявлення і індикація перевантаження на основі довжини черги є найпростішою і зазвичай використовуваною схемою управління інформаційним потоком на основі зворотного зв'язку (рис. 3.1). При одночасній реалізації S інформаційних процесів (дейтаграмний спосіб передачі) пакети від S джерел $D_1 \ldots D_S$ надходять у відповідні черги на обслуговування, з яких через схеми регулювання видачі пакетів $РВП_1$ ...$РВП_S$ поступають в чергу вузла призначення. Формувач індикатора перевантаження(ФІП) на основі інформації про стан черги $N(t)$ надає дані про наявність чи загрозу перевантаження $J(t)$ до кінцевого користувача чи транзитного вузла.

Стан перевантаження традиційно визначається довжиною черги $G(t)$ в транзитному вузлі зв'язку чи пункті призначення кінцевого отримувача даних в момент часу t. Коли довжина черги досягає попередньо визначеного граничного значення Q, пакети, що проходять крізь чергу, матимуть біт індикації, встановлений в стан "підтверджене перевантаження".

Основна перевага схеми на основі черг є її низька складність, тому що абсолютну довжину черги можна контролювати за допомогою одного лічильника. Проте, цей метод не є ефективним при способі контролю сегментів мережі із транзитними вузлами передачі інформації. Використання одного біта перевантаження для всього сегменту не дає можливості локалізувати місце

перевантаження з точністю до вузла. Даний метод може створювати великі черги в мережних вузлах і викликати великий час затримки інформації зворотного зв'язку.

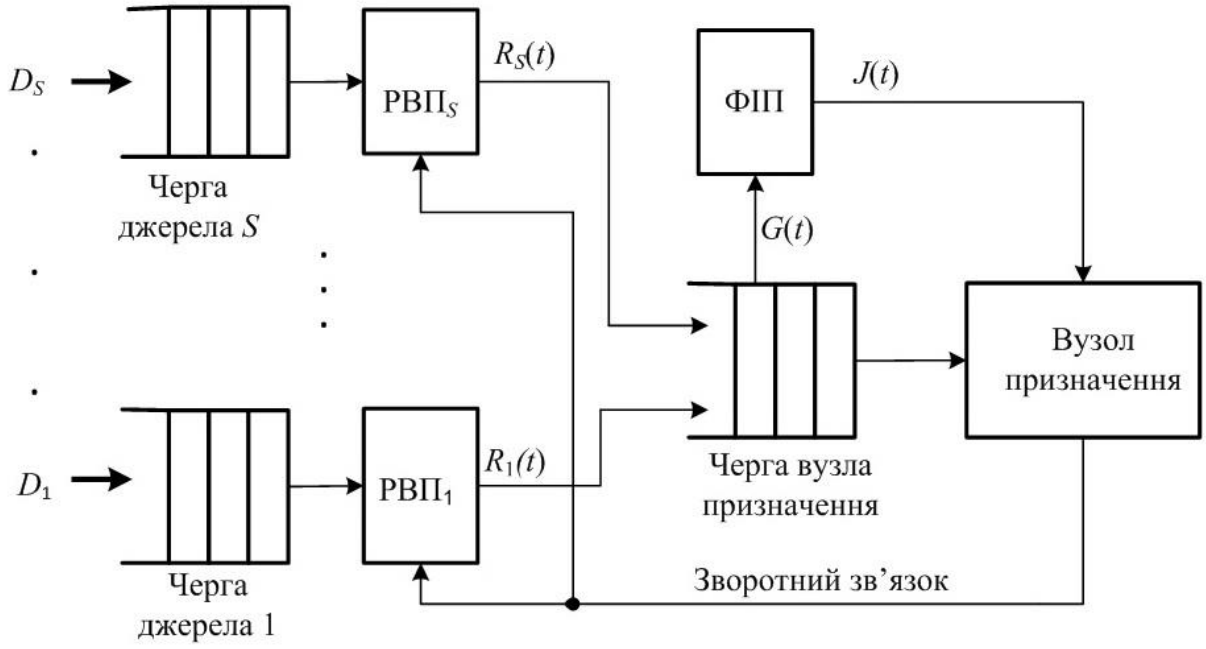

Рисунок 3.1 – Встановлення біта перевантаження на основі довжини черги

Виявлення створеного перевантаження повинно затриматися на певний час, необхідний для створення черги. Аналогічно, виявлення рішення перевантаження також затримується на певний час, необхідний для оброблення черги. Для управління "адитивним збільшенням / множинним зниженням" швидкості джерела маємо:

$$R(t+1)=\begin{cases} R(t)+F^{+},\ G(t-t_d)<Q \quad \text{or} \quad E\left(t-t_d\right)>0 \\ F^{-}R(t),\ G(t-t_d)\geq Q \quad \text{or} \quad E(t-t_d)\leq 0 \end{cases}, \tag{3.1}$$

де R – швидкість заповнення черги на вході вузла призначення;

$F^{+}> 0$ – фактор адитивного збільшення;

$0 < F^{-}< 1$ – фактор мультиплікативного зниження;

$E(t–\tau) = Q – G(t–\tau)$ – функція помилки;

t_d – час затримки передачі стану черги затору до джерела.

В [18, 25] розглянуті способи і методи визначення чутливості вихідних характеристик телекомунікаційних мереж як систем масового обслуговування. В основу цих способів покладені моделі управління чергою для адаптації

регульованого доступу зовнішнього трафіку в систему з метою отримання очікуваної межі продуктивності. В роботах не конкретизовані визначення функцій чутливості телекомунікаційної мережі та методи їх ідентифікації у мережних системах управління з явним або непрямим зворотним зв’язком.

У роботі [18] моделі й методи управління чергами базуються на підґрунті функцій чутливості вихідних характеристик телекомунікаційних мереж як систем масового обслуговування. Однак асимптотичні характеристики функцій чутливості остаточно не визначені. Крім того, не отримані вирази у замкненій формі для функціонального або статистичного зв’язку параметрів функцій чутливості та відповідних параметрів системи управління чергами.

У роботі [26], присвяченій контролю над перевантаженнями телекомунікаційної мережі, запропонований алгоритм активного розподілу спроб одночасного доступу до слоту з рівномірним часовим розподілом. Такий підхід є справедливим для найменш переважного розподілу потоку заявок (Least Favourable Distribution, LFD). В сучасних телекомунікаційних мережах розподіли потоків заявок далекі від рівномірного, тому асимптотичні оцінки, що отримані, приводитимуть до невиправдано оптимістичних висновків. Реалістичні оцінки можна отримати при застосуванні нейромережних моделей, які мають адаптуватися під стрибки навантаження та варіації імовірнісних розподілів потоків заявок.

3.2 Контроль перевантаження на основі застосування нейромережних моделей

В роботі [25] обґрунтована доцільність використання математичного апарату аналізу чутливості складних систем в задачах управління перевантаженнями в комп'ютерних мережах. Результати, приведені в публікаціях [15, 16], відіграють роль рекомендацій загального характеру. Користуючись ними, звичайно, можна розробляти конкретні методи боротьби з перевантаженнями, але це потребує нових

нетрадиційних підходів до розв'язання проблеми в цілому. Одним із таких підходів є застосування методів штучного інтелекту. В першу чергу, це нейронні мережі удосконаленої архітектури із явними та прихованими шарами та поточною оптимізацією параметрів за результатами навчання шляхом зворотного розповсюдження похибки.

Розглянемо схему контролю перевантаження із зворотним зв’язком по знаку чутливості функції продуктивності системи. Знак чутливості продуктивності дії надає оптимальний напрям для налаштування швидкості джерела даних [16]. Запропонована схема для визначення функції чутливості використовує просту нейронну мережну модель динамічної системи.

Нейронні мережі представляють собою математичні структури, здатні до самонавчання на основі отримуваної ззовні інформації [27].

Нехай динаміка системи обслуговування виражена наступним рівнянням введення-виведення:

$$G(t+1) = f[G(t), G(t-\tau),...,G(t-l\tau),\ R(t), R(t-\tau),...,R(t-m\tau)], \qquad (2.2)$$

де $G(t)$ – скалярний вихід: довжина черги або затримки обслуговування в момент часу t;

$R(t)$ – скалярний вхід: миттєва швидкість черги на вході в момент часу t;

$f[.]$ – невідома функція, оцінювана за допомогою нейронної мережі;

l, m – відповідно, порядки $\{N(t), R(t)\}$;

τ – період часових відліків, період тактової частоти системи.

Метою алгоритму оптимального управління є обрання керуючого сигналу $R(t)$ таким чином, щоб вихід системи $G(t)$ відповідав як можна ближче заздалегідь встановленим характеристикам $Q(t)$ (як правило, $Q(t)$=Q=const). Послідовно-паралельна нейронна модель [18] невідомої системи (2.2) може бути представлена як

$$G^{\wedge}(t+i\tau) = f^{\wedge}[G(t),...,G(t-l\tau),R(t),...R(t-m\tau)], \quad i=1,2,...,L, \qquad (2.3)$$

де $G^{\wedge}(t+i\tau)$ – вихід нейронної мережі;

$f^{\wedge}$ – оцінка функції f;

L – горизонт передбачення.

На рис. 3.2 показана схема передбачення стану черги на основі нейронної системи.

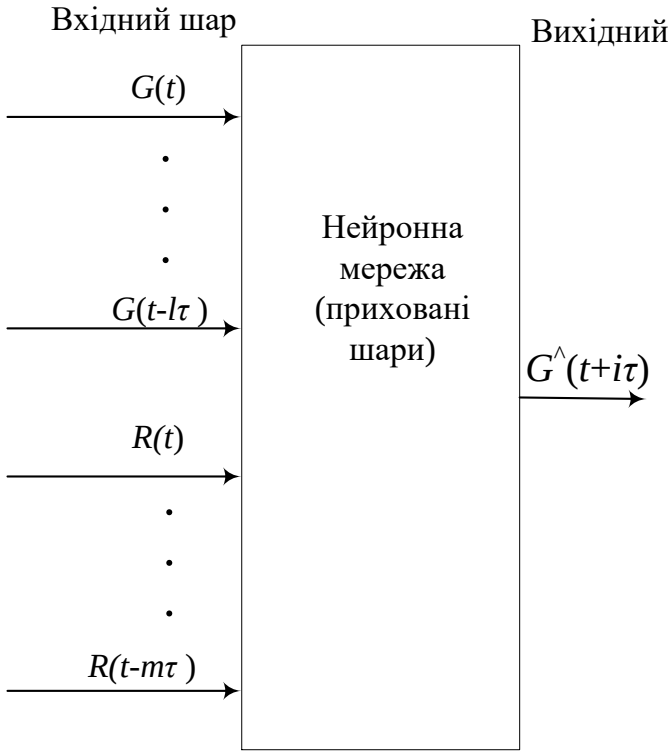

Рисунок 3.2 – Нейронна модель передбачення стану черги

Якщо нейронна мережа налаштована на відслідковування динаміки системи і показує, що квадратична помилка

$$[G(t+i\tau)-\hat{G}(t+i\tau)]^2=\varepsilon \tag{3.4}$$

є незначною, вважається, що i-й крок відповідає наперед передбаченому виходу системи (3.2). Як результат, керуючий сигнал може бути обраний таким чином, що $\hat{G}(t+i\tau)$ знаходиться якомога ближче до Q.

Визначимо функцію вартості (cost function), як цільову функцію наявності перевантаження, наступним чином:

$$J=\frac{1}{2}e^2\left(t+i\tau\right)=\frac{1}{2}\left[Q-\hat{G}\left(t+i\tau\right)\right]^2,\quad i=1,2,...,L. \tag{3.5}$$

Керуючий сигнал $R(t)$ (тобто, швидкість джерела даних) повинен бути обраний так щоб мінімізувати J. В дискретному випадку керуюча змінна оновлюється відповідно до наступного правила градієнтного спуску:

$$R(t+\tau) = R(t) + \Delta R(t) = R(t) - \eta \frac{\partial J}{\partial R(t)}, \tag{3.6}$$

де η є розміром кроку управління.

Очевидно, найбільш слабким припущенням для характеристики функції втрат, що дає можливість використовувати ці методи, є унімодальність функції втрат. Для реальних телекомунікаційних систем та мереж це припущення є досить реалістичним [28, 29].

Можна побачити, що для визначення відповідного керуючого сигналу процес мінімізації (6) спирається на наближення, зроблене нейронною мережею. Тому необхідно, щоб $G^{\wedge}(t+i\tau)$ наближалося до виходу дійсної системи $G(t+i\tau)$ асимптотично. Це може бути досягнуто шляхом підтримки навчання нейронної мережі он лайн.

Диференціюючи (3.5) по функції $R(t)$, отримаємо:

$$\Delta J(t) = \frac{\partial J}{\partial R(t)} = -e(t+i\tau)\frac{\partial G^{\wedge}(t+i\tau)}{\partial R(t)}, \quad i = 1,2,...,L, \tag{3.7}$$

де вираз $\partial G^{\wedge}(t+i\tau)/\partial R(t)$ відомий, як чутливість чи градієнт системи [25, 26].

Підставляючи (3.7) в (3.6), отримуємо:

$$R(t+\tau) = R(t) + \eta e(t+i\tau)\frac{\partial G^{\wedge}(t+i\tau)}{\partial R(t)}, \quad i = 1,2,...,L. \tag{3.8}$$

Градієнт системи може бути аналітично оцінено, використовуючи відому структуру нейронної мережі [28].

Вираз (3.6) може бути представлений у вигляді

$$R(t+\tau) = R(t) - \eta \,\mathrm{sgn}[\Delta J(t)], \tag{3.9}$$

де $\mathrm{sgn}[\Delta J(t)]$ позначає знак $\Delta J(t)$ (який може бути додатним чи від'ємним).

Таким чином, можемо зробити висновок, що (3.7) реалізує правило регулювання адитивного збільшення / множинного зменшення швидкості джерела. Тому привабливою альтернативою схеми формування індикатора перевантаження,

заснованій на пороговому заповненні черги, є алгоритм адитивного збільшення / множинного зменшення. Алгоритм визначає зміну в швидкості джерела даних $R(t)$ в залежності від знаку чутливості показника продуктивності $\Delta J(t–t_d)$. Тобто, індикатор перевантаження $B(t)$ формується в залежності від градієнта системи $\Delta J(t)$ в момент часу t:

$$\begin{cases} B(t)=0 \quad \text{if} \quad \Delta J(n)<0, \\ B(t)=1 \quad \text{if} \quad \Delta J(n)\geq 0. \end{cases}$$

Значення ΔJ, обчисленого за допомогою формули (3.7) дає оптимальний напрямок для регулювання швидкості джерела. Коротше кажучи, тільки знак, а не величина ΔJ має значення в цьому випадку.

Схема регулювання вхідного потоку даних (швидкості надходження пакетів) з використанням нейронної мережі для аналізу чутливості показана на рис. 3.3.

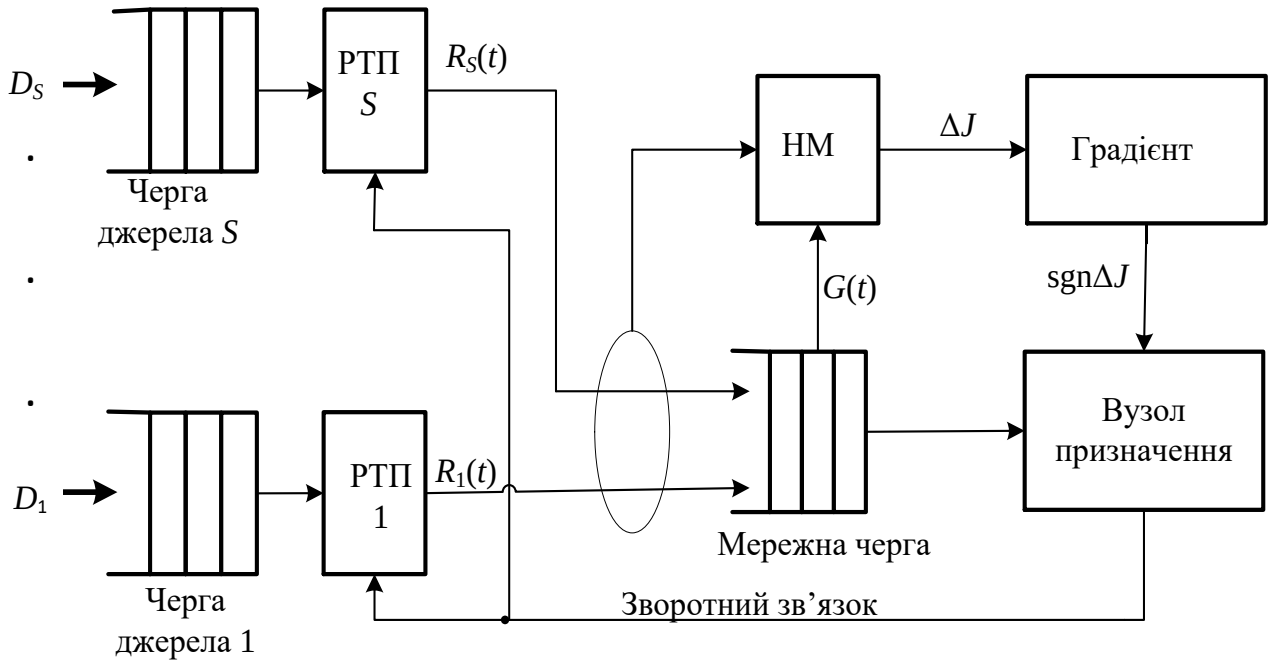

Рисунок 3.3 – Оптимізаційна схема управління перевантаженням

Поточні значення швидкостей регульованих вхідних потоків $R_1(t)…R_S(t)$ поступають на вхідний шар нейронної мережі НМ. На основі аналізу значень $R_1(t)…R_S(t)$ та поточного значення довжини мережної черги $G(t)$ вихідний шар нейронної мережі відслідковує величину відхилення ΔJ цільової контрольованої

функції J (наявності чи загрози перевантаження). Знак величини відхилення $\mathrm{sgn}\Delta J$ враховується вузлом призначення при формуванні сигналу зворотного зв’язку.

Для алгоритму керування адитивним збільшенням / множинним зменшенням швидкості джерела маємо:

$$R(t+\tau)=\begin{cases} R(t)+F^{+}, & \Delta J(n-\tau)<0, \\ R(t)-F^{-}, & \Delta J(n-\tau)\geq 0. \end{cases} \tag{3.10}$$

3.3 Багатокрокова модель прогнозування та виявлення перевантаження комп'ютерної мережі

Точність передбачення перевантаження та управління потоком даних на основі зворотного зв'язку у високошвидкісних комп’ютерних мережах в значній мірі залежить від затримок передачі даних між мережевими вузлами зв'язку. Внаслідок цього відповіді команд управління вступлять в силу в межах мережі після деякої затримки, а керуюча інформація, отримана в джерелах даних чи мережевих точках доступу, може виявитись застарілою [29].

Тому розробка ефективного механізму управління перевантаженнями є досить важливою і складною задачею, що обумовлює актуальність досліджень в даному напрямку.

Як покано вище, ефективним способом керування потоком даних і зменшення затримки інформації зворотного зв'язку є використання технології передбачення на основі нейронних мереж та функції чутливості ключових показників ефективності.

Підхід керування виглядає наступним чином: рівень завантаження мережі контролюється через шляхом аналізу певних порогів вхідних і вихідних черг з об’єктом управління деякого порогу Q. На основі різниці між поточним (в момент часу t) значенням довжини черги $G(t)$ і Q, датчик перевантаження на основі результатів розрахунків періодично коригує швидкість прийому інформації вузлом-отримувачем. В свою чергу, джерела трафіку зменшують швидкість передачі даних до рівня, дозволеного контролером. Вибір встановленого значення

порогу Q відображає компроміс між значенням затримки пакетів, втрати пакетів, втрати пропускної здатності джерел трафіку.

В нашому дослідженні пропонується розвиток багатокрокового методу передбачення, що використовується для керування станом мережевої черги, в якій передбачення робиться, з допомогою нейронних мереж. Аналізується можливість використання нейронних мереж і прогностичних методів для вирішення проблем управління перевантаженням на основі зворотного зв’язку.

Більшість схем керування перевантаженням на основі нейронних мереж, запропоновані в літературі, не враховують ефект затримки контуру управління в розробці алгоритмів керування, так як моделі стають занадто громіздкими і не піддатливими до розуміння. Часто такі моделі супроводжуються необхідністю вирішенням задач підвищеної обчислювальної складності, потребують суттєвих часових затрат, що призводить до значних, а часом і неприйнятних, затримок управляючої інформації [27, 28, 30].

Розглянемо структури управління перевантаженнями при обслуговуванні трафіка з низьким пріоритетом на послуги в комп’ютерній мережі. Трафік з низьким пріоритетом буде використовувати доступну смугу пропускання, яка існує незалежно від смуги пропускання трафіка більш високого пріоритету. Для трафіка вищого пріоритету частина смуги пропускання використовується миттєво без очікування. Трафік з низьким пріоритетом має доступ до повної смуги пропускання тільки тоді, коли трафік більш високого пріоритету не чекає передачі. Такий підхід дозволяє покращити ефективність використання каналу, не впливаючи на якість обслуговування трафіку високого пріоритету. Джерела з низьким пріоритетом повинні постійно пристосовувати свої швидкості для зміни в часі пропускної здатності. Така схема управління перевантаженням забезпечує шляхи для динамічного оцінювання рівня доступних ресурсів, залишених невикористаними в мережі при наявності трафіку високого пріоритету. Функції оцінки та розподілу (або спільного використання ресурсів) для обчислення швидкості кожного окремого джерела засновані на використанні нейронної мережі для контролю рівня заповнення буферної пам'яті та черги трафіку з низьким пріоритетом. Буфер

мережевого вузла, виділений для трафіка з низьким пріоритетом обслуговування, використовується для того, щоб отримати кращі статистичні дані обробки даних в мережі.

Розглянемо модель управління, де до мережевого вузла з різних напрямків надходить трафік з різними вимогами на обслуговування. Нехай ці вимоги відображені в наступних класах пріоритетності: P_0, P_1 і P_2, з яких клас P_0 має найвищий пріоритет, P_2 – найнижчий. Для цього прикладу, на мережевому вузлі організовані три черги, призначені вказаним трьом типам трафіка.

На рис. 3.4 показана модель системи керуванням перевантаженням зі спільною обробкою трафіку з різною пріоритетністю. Вертикальна пунктирна лінія відображає спільне керування швидкістю надходження пакетів з урахуванням узагальненого значення порогу Q_Σ для всіх видів трафіку з класами пріоритетності: P_0, P_1 і P_2.

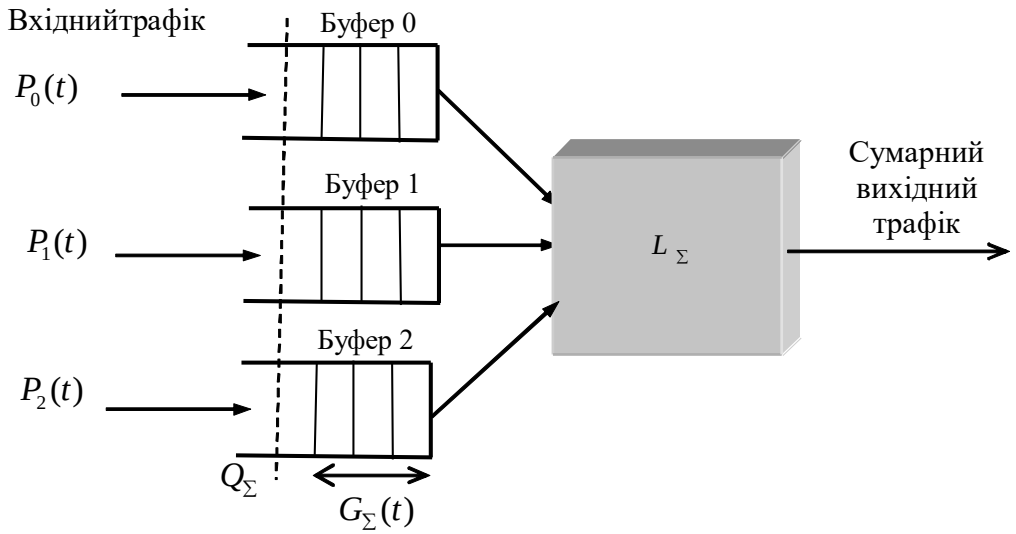

Рисунок 3.4 – Модель системи керуванням перевантаженням зі спільною обробкою трафіку з різною пріоритетністю

Для такої моделі сумарний вихідний трафік формується із умови аналізу і врахування загальної ємності L_Σ вихідної ланки вузла.

На рис. 3.5 показана модель системи керуванням перевантаженням з виокремленою. обробкою трафіку з нижчою пріоритетністю.

Сервер направляє пакети в буфери (черги) відповідно до їх пріоритетів P_0, P_1 і P_2. Алгоритм управління швидкістю повинен обчислити доступну пропускну спроможність, залишену трафіками P_0 і P_1 (тобто, трафіками вищого пріоритету).

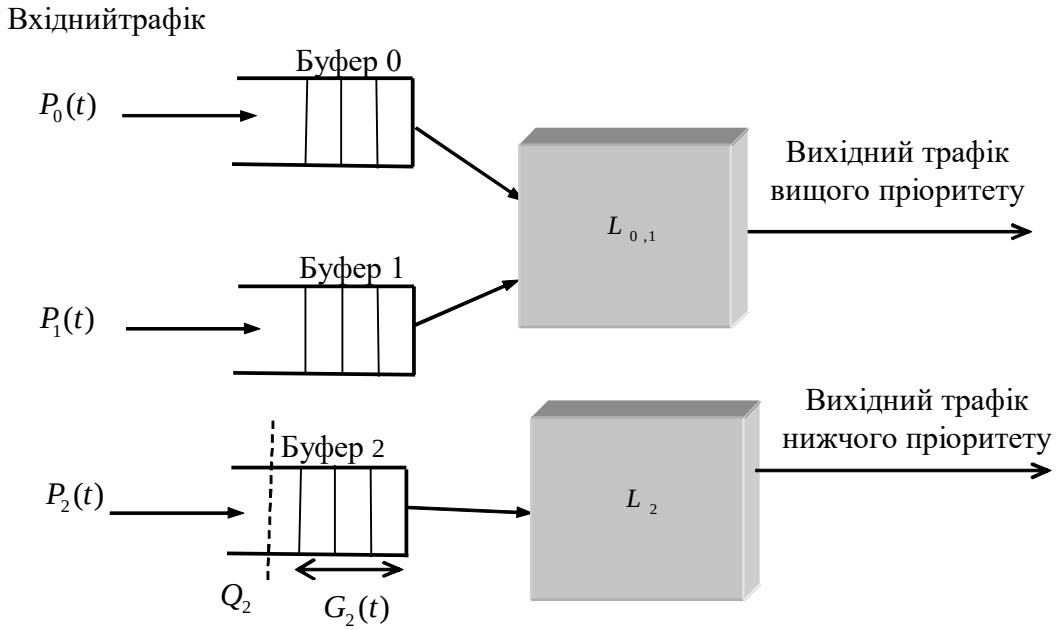

Рисунок 3.5 – Модель системи керуванням перевантаженням з виокремленою обробкою трафіку з нижчою пріоритетністю

Для управління перевантаженням даного мережевого вузла використовуються наступні параметри і змінні черги з пріоритетом P_2:

$G_2(t)$ – кількість пакетів з пріоритетом, очікуваних на видачу в час t;

$Q_2(t)$ – поріг черги в час t, зазвичай $Q_2(t) = Q_2 = \text{const}$;

$N_2(t)$ – кількість з'єднань низького пріоритету (P_2), встановлених через вузол в час t;

L_Σ – загальна ємність вихідної ланки вузла.

Вихідний трафік вищих класів пріоритетності використовує пропускну смугу вузла поза чергою, використовуючи ємність вихідної черги $L_{0,1}$. Управління і

контроль перевантаженням здійснюється тільки для черги з нижчим пріоритетом (буфер 2) на основі аналізу порогу Q_2 і довжини черги $G_2(t)$ найнижчого пріоритету.

На рис. 3.6 представлена структурна схема системи прогнозування і виявлення перевантаження, побудована на основі нейронної мережі з використанням функції чутливості ключових параметрів ефективності – продуктивності (пропускної здатності) телекомунікаційної мережі.

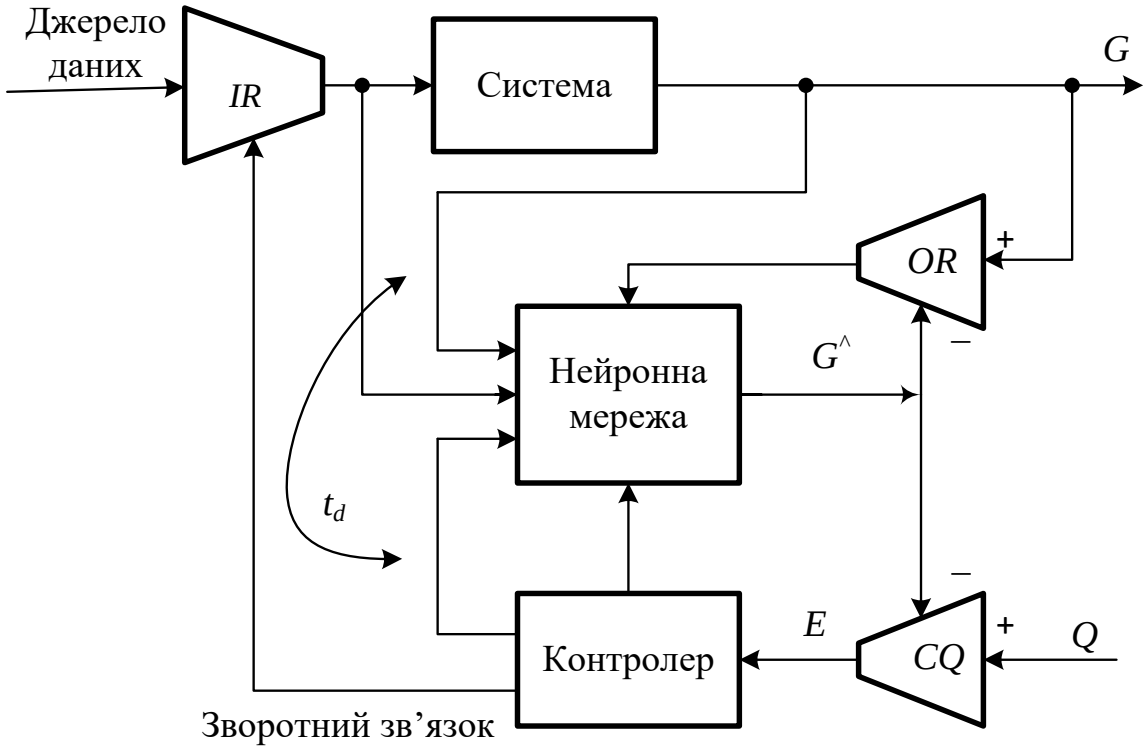

Рисунок 3.6 – Схема керування перевантаженням із замкненим круговим контуром

Система побудована по схемі із замкненим круговим контуром зворотного зв'язку [4, 29]. Регулювання швидкості вхідного потоку системи (швидкості заповнення вхідної черги) здійснюється на пристрої *IR (Input Regulator)*. Регулювання і контроль довжини черги вихідного потоку здійснюється на пристрої *OR (Output Regulator)* із врахуванням поточної G і прогнозованої $G^{\wedge}$ довжини вихідної черги.

Прогнозоване значення вихідної черги, як ключового показника ефективності, формується прогнозуючою схемою управління (рис. 3.7), яка реалізована на основі нейронної мережі і контролера із врахуванням поточного значення величини черги. Пристрій *CQ* (Control Queue) відслідковує величину *E* – відхилення поточного значення величини черги від її порогового (граничного) значення *Q*. Вхідний шар нейронної мережі здійснює обробку параметрів вхідних і вихідних трафіків поточних і попередніх потоків, передбачуваного (прогнозованого) вхідного потоку з врахуванням відхиленння від порогового значення *Q* вихідної черги.

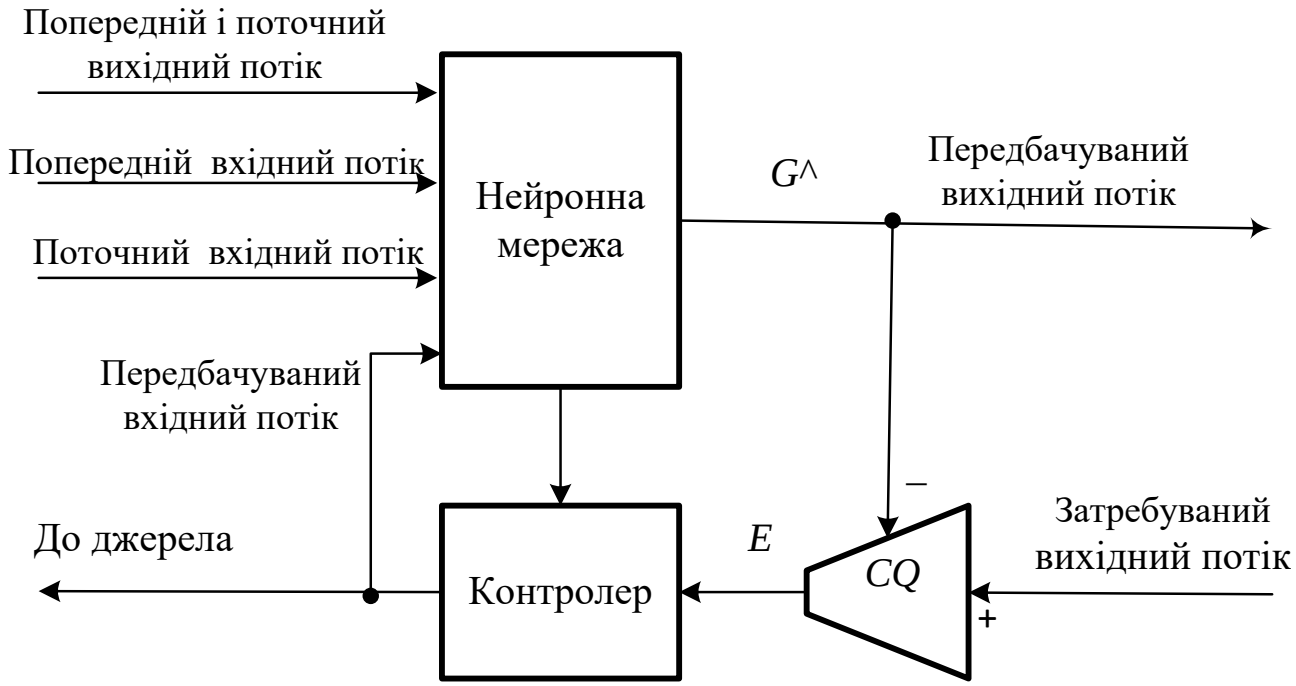

Рисунок 3.7 – Прогнозуюча схема управління

Для створення математичної моделі, яка б дозволяла представити цільову задачу прогнозування перевантаження в необхідній формі для застосування методів керування, спочатку розглянемо одне джерело передачі даних до мережевого буфера і фіксованої кругової затримки.

Керування перевантаженням можна розділити на 3 процеси:

– вимірювання загроз перевантаження в мережі на основі використання функції чутливості і передбачення нейронної мережі;

– обмін інформацією між вузлами (використовуючи пакети управління);

– заборона передачі пакетів в мережі шляхом регулювання швидкості джерела на основі допустимих значень довжини черг.

Прогнозуючі контролери, як правило, мають чотири головні характеристики:

– моделлю процесу необхідно управляти. В нашій задачі управління ця модель використовується, щоб передбачати вихідні дані процесу через горизонт передбачення;

– визначення критерія мінімізації для отримання оптимальної сукупності даних шляхом передбачення горизонту. Зазвичай, це квадратичний критерій, який відстежує помилки і іноді використовує вихідні дані контролера.

– наявність еталонної (розрахункової) траєкторії виходу процесу, тобто, передбачувані значення вихідних даних процесу.

– виконання процедури мінімізації.

Контролери з заданими характеристиками можуть бути розроблені з допомогою відповідного вибору цих характеристик. Позначимо через $U(t)$ керуючу змінну, яка забезпечує швидкість, з якою джерело повинно передавати дані. Індекс продуктивності в загальному вигляді визначається як [30]

$$J = \int_{t_0}^{f} C(\underline{x}(t), \underline{x^e}(t), u(t))dt \ , \tag{3.11}$$

де $\underline{x}(t)$ – вектор змінних стану контрольованого процесу і $\underline{x^e}(t)$ – бажаний стан.

Величина C позначає функцію витрат, і може бути визначена як відстань між дійсним станом системи $\underline{x}(t)$ і бажаним станом $\underline{x^e}(t)$.

Мінімізація функції витрат C здійснюється на основі використання оптимального керування $U^*(t)$, $t_0 \le t \le t_f$. Так як контрольований процес стохастичний, функцію C потрібно визначати як очікуване значення. Пряма процедура мінімізації C потребує багато обчислень в кожній ітерації. Звичайний шлях зменшити кількість обчислень за ітерацію – засновувати оновлення змінних управління на миттєвому індексі продуктивності

$$J = C(\underline{x}(t), \underline{x^e}(t), u(t)) \ . \tag{3.12}$$

Добре відомий приклад мінімізації процедури і алгоритму градієнтного спуску, для дискретного випадку, керуюча змінна оновлюється за наступним правилом:

$$U(t+1) = U(t) - \eta \frac{\partial J}{\partial U(t)} \ , \tag{3.13}$$

де η>0 – розмір кроку управління.

Чутливість індексу продуктивності $\Delta J(t) = \partial J / \partial U(t)$ аналітично оцінюються, використовуючи відому структуру нейронної мережі [18, 28].

Вирази (3.11), (3.12), (3.13) можуть бути використані при створенні систем керування перевантаженням на основі нейронних мереж з використанням функції чутливості продуктивності та інших ключових показників ефективності телекомунікаційних мереж.

В роботі для прогнозування і виявлення перевантаження телекомунікаційної мережі використано нейронну систему, яка реалізує і відслідковує поточний характер функції чутливості ключових параметрів мережі. Використання оптимальних алгоритмів налаштування вагових коефіцієнтів нейронної мережі дає можливість підвищити точність визначення керуючих сигналів, зменшити затримки зворотного зв’язку і, як результат, мінімізувати середні витрати ресурсу.

Можна показати, що для системи прогнозування перевантаження алгоритм і процедура градієнтного спуску по вектору вагових коефіцієнтів *i*-го шару нейронної мережі представляється у вигляді прямокутної матриці. У той же час відомо, що найбільш простий вид градієнт приймає в разі квадратної матриці вагових коефіцієнтів. Асимптотичні оцінки обчислювальної складності запропонованого методу при реалізації прогнозуючого контролера можуть бути отримання розрахунковим шляхом або шляхом комп’ютерного моделювання. Враховуючи, комп’ютерне моделювання потребує значних обчислювальних потужностей, цю проблему можна розв’язати у доволі потужному обчислювальному центрі, наприклад, в обчислювальному центрі провідного оператора зв’язку.

Використовувані рішення щодо керування перевантаженням телекомунікаційної мережі базуються на відомих положеннях теорії систем і адаптивного управління. Це дозволяє стверджувати, що підходи, основані на моделюванні нейронної мережі забезпечують концептуальні основи і процедури для нелінійного передбачення і управління систем передбачення перевантажень.

3.4 Алгоритм навчання нейронної мережі і формування признаку перевантаження

Модель навчання для нейронної мережі може змінюватися з часом, тому в керуючій схемі прогнозування перевантаження на основі функції чутливості необхідне он лайн навчання. Оскільки використовується явний зворотний зв'язок за функцією чутливості, у даній роботі обраний метод навчання нейронної мережі на основі зворотного розповсюдження похибки як найбільш точний та статистично стійкий [28].

Алгоритм навчання нейронної мережі і формування признаку перевантаження полягає в наступному.

Початковий стан: Момент часу *t.*

Крок 1: Отримати виміри системи на вході і виході: $R(t)$ і $G(t)$.

Доступна історія системи:

$$R(t-\tau),...,R(t-m\tau);\quad G(t-\tau),...,G(t-l\tau).$$

Історія нейронної мережі:

$$G^{\wedge}(t),\ \delta_i^{(p)}(t-\tau),\ x_i^{(p)}(t-\tau)\ ,$$

для всіх шарів p і всіх вузлів i.

Крок 2: Обчислення нейронною мережею прогнозованого значення

$$G^{\wedge}(t-i\tau).$$

Обчислення нейронною мережею значення зворотного зв'язку

$$\frac{\partial G^{\wedge}(t-i\tau)}{\partial R(t)}.$$

Крок 3: Обчислення функції градієнта

$$\Delta J(t) = \frac{\partial J}{\partial R(t)}.$$

Формування індикатора перевантаження

$$B(t) = \begin{cases} 1 & \text{if} \quad \Delta J \geq 0, \\ 0 & \text{else.} \end{cases}$$

Крок 4: Використовуючи вираз

$$\left[G^{\wedge}(t) - G(t) \right]$$

для оновлення ваг нейронної мережі, виконується обчислення:

$$\Delta w_{ij}^{(p)} = -\left[G^{\wedge}(t) - G(t) \right] \delta_i^{p-1}(t-\tau) x_j^{p-1}(t-\tau) + \gamma \Delta w_{ij}^{p}(t-\tau),$$

$$w_{ij}^{(p)} \Leftarrow w_{ij}^{p} + \Delta w_{ij}^{p}(t)$$

для

$$p = p_s,\ p_{s-1}, ..., 1; \quad i = 1, 2, ..., m_{(p)}; \quad j = 1, 2, ..., m_{p-1}.$$

β – швидкість навчання нейромережі;

γ – постійний імпульс нейромережі.

Крок 5: Наступний інтервал: $t \leftarrow t+\tau$. Перейти до кроку 1.

Примітка: Кроки 3 і 4 можуть виконуватись паралельно.

Визначимо наступні показники ефективності для виконання моделювання, де T – час виконання моделювання:

– $G_{\max} = \max\{G(t): 0 \leq t \leq T\}$. Максимальне значення $G(t)$ відображає розмір буфера, який потрібний у вузькому місці для уникнення втрати пакету;

– середній час (f ˜)досягнення розміру черги і швидкості джерела:

$$f^{\square} = \frac{1}{T} \int_0^T f(t) dt;$$

– дисперсія $\sigma^2(f)$ розміру черги і швидкості джерела:

$$\sigma^2(f) = \frac{1}{T} \int_0^T [f(t) - f]^2 dt.$$

Проведемо порівняльний аналіз ефективності керуючих схем виявлення перевантаження на основі функції чутливості продуктивності телекомунікаційної

мережі. Розглянемо одне з'єднання з наступним набором параметрів: пікова швидкість джерела R_{max}=100 пакетів/одиницю часу, одиниця часу τ=0,25 ms, мінімальна швидкість R_{min}=100 пакетів/одиницю часу, коефіцієнт адитивного збільшення $F^{+}=R_{max}/16$, коефіцієнт множинного зменшення $F^{-}=15/16$. Поріг перевантаження встановлений при Q = 500 пакетів.

Розглянемо затримку округлення з 6 одиниць часу – 6τ. В дослідженні передбачається два вузьких місця:

– синусоїдальне integer[35(1+sin(2(πτ))=10] пакетів/одиницю часу;

– випадкове (максимальне значення 80 пакетів/одиницю часу).

Архітектура нейронної мережі така: 3-шарова нейронна мережа (8 входів, 8 вхідних нейронів, 8 прихованих нейронів, 1 вихідний нейрон); порядок довжини черги l=3; порядок швидкості заповнення черги m=3. Для визначення стану вузького місця черги в даному дослідженні розглядаються два різних горизонти передбачення: 1-й і 3-й кроки передбачення.

Короткий перелік результатів моделювання наведений в табл. 3.1, де G_{av} і R_{av} позначають середній за часом розмір черги і швидкість джерела.

Таблиця 3.1 Перелік результатів моделювання

Підхід	Тип обслуговування	G_{max}	G_{av}	$\sigma^2(G)$	Індекс G_{max}	R_{av}	$\sigma^2(R)$
На основі черги	Випадковий	1560	724.5	187569.3	757	45.1	1370.1
	Синусоїдальний	1556	713.8	179305.5	281	45.7	1211.4
Градієнт: 1-й крок	Випадковий	1293	666.2	108986.4	480, 481, 482	44.4	878.3
	Синусоїдальний	1428	653.3	86115.5	48	44.4	1066.9
Градієнт: 3-й крок	Випадковий	1307	622.9	88526.7	800	44.2	929.9
	Синусоїдальний	1443	628.7	89936.5	49	44.5	1139.7

Результати моделювання, приведені в табл. 1, показують, що величина черги і коливання швидкості джерела менші для схеми, керованої на основі функції чутливості, ніж для схеми, керованої на основі аналізу черги. Схема на основі чутливості з трьохкроковим передбаченням стану дає кращу продуктивність, ніж відповідна схема з однокроковим передбаченням. Це відбувається тому, що для трьохкрокового передбачення, затримка керуючих сигналів зворотного зв’язку (тобто, індикаторів перевантаження), отриманих в джерелі даних з мережної черги, несуттєва. Тобто, індикатор перевантаження ближче відображає мережні умови, які передбачають процес. В схемах керування перевантаженням на основі зворотного зв’язку зі значними затримками поширення управляючі сигнали, отримані в джерелах, можуть бути застарілими. Дія управління зворотного зв'язку вступить в силу в межах мережі тільки після деякої затримки. Це означає, що більш точно передбачити стан черги можна буде із значною затримкою.

Таким чином, недоліком передбачення стану черги є те, що чим з більшою затримкою ми будемо робити передбачення, тим складніше отримати передбачення з припустимими помилками.

Схема на основі черги більш чутлива до змін в швидкості обслуговування черги, ніж для схеми на основі чутливості. Слід зазначити, що в схемі на основі черги при значному зниженні швидкості обслуговування черги за певний проміжок часу розмір черги зростає екстенсивно за межі раніше спостережуваних значень. Збільшення розміру черги більш стабільне при тих же умовах для схем на основі чутливості. Значення розбіжності дає уявлення про спостережувані величини.

Для мінімізації похибок прогнозу стану мережі треба дотримуватися очевидного правила: затримки доставки службової та користувальницької інформації повинні бути одного порядку з часом реакції комутаційного вузла. Задача точного визначення затримок та часу реакції, розробка методів регулювання затримок в залежності від стану завантаження мережі та часу реакції комутаційних вузлів аж ніяк не є тривіальною. Її повний розв'язок, зокрема, з отриманням кількісних оцінок, потребує ретельних досліджень теоретичного та практичного характеру.

У даній роботі для контролю перевантаження телекомунікаційної мережі як складної системи з випадковими затримками службової (управляючої) та користувальницької інформації застосовано нейронну мережу. На відміну від традиційних систем контролю перевантаження за змінами стану та параметрів черг, запропонована система працює за оптимальними алгоритмами налаштування вагових параметрів. Завдяки цьому підвищується точність визначення керуючих сигналів, зменшується вплив їх затримки і, як результат, мінімізуються середні витрати ресурсу.

Розглядаючи нейронну мережу як систему керування телекомунікаційною мережею, за замовченням робиться припущення про належність використовуваної нейронної мережі до класу динамічних нейронних мереж. По суті, це статична нейронна мережа, до складу якої вводиться зворотний зв’язок через елемент затримки на один такт. Це припущення для пакетних телекомунікаційних мереж є вельми логічним. До того ж, градієнт функціоналу мінімізується по вектору вагових коефіцієнтів нейронної мережі. Цей градієнт розглядається як набір градієнтів за матрицями вагових коефіцієнтів окремих шарів. При цьому виявляється, що для обчислення градієнта по матриці ваг нижчого шару можуть бути істотно використані результати розрахунку по матриці ваг шару, що лежить вище. Завдяки запропонованій архітектурі динамічної нейронної мережі, по-перше, значно прискорюється поточний процес підбору оптимальних вагових коефіцієнтів, а, по-друге, спрощується процедура самонавчання мережі методом зворотного розповсюдження похибки.

Для побудови оптимальних алгоритмів настройки параметрів нейронної мережі потрібно обчислювати окремо як градієнт векторного виходу нейронної мережі, так і градієнт функціоналу по вектору нев'язки. Обчислення по вектору нев'язки не представляє особливих труднощів. Обчислення градієнта векторного виходу нейронної мережі можна виконати за допомогою методу, близького до методу зворотного поширення помилки. Однак при цьому необхідно ретельно контролювати значення коренів характеристичного поліному динамічної нейронної мережі, що потребує додаткового обчислювального ресурсу.

Можна показати, що вираз для градієнта виходу мережі по вектору вагових коефіцієнтів i-го шару представляється у вигляді прямокутної матриці. У той же час відомо, що найбільш простий вид градієнт приймає в разі квадратної матриці вагових коефіцієнтів. Для асимптотичного оцінювання тенденції зростання обчислювальної складності при відхиленні форми матриці від квадратної пропонується розраховувати порівняльні оцінки потрібного обсягу обчислень при обертанні квадратної $N \times N$ матриці та при псевдообертанні прямокутної $N \times L$ матриці.

Отримання (розрахунковим шляхом або шляхом комп'ютерного моделювання) асимптотичних оцінок обчислювальної складності запропонованого методу дасть можливість отримання потенціальних характеристик системи керування. Це обіцяє гарні перспективи системного рішення проблеми оптимального керування телекомунікаційною мережею. При цьому треба враховувати, що виведення виразів у замкненій формі та відповідних розрахункових виразів може бути пов'язане з громіздкими, хоча і досить простими викладками. В свою чергу, комп'ютерне моделювання потребує значних обчислювальних потужностей. Представляється, що цю проблему можна розв'язати у доволі потужному обчислювальному центрі, наприклад, в обчислювальному центрі провідного оператора зв'язку.

ВИСНОВКИ

Традиційні підходи до вирішення задачі керування перевантаженнями базуються на основі контролю кількісного значення довжини черги вхідного і вихідного потоків мережі. Ідентифікатор перевантаження може вказувати тільки на наявність перевантаження по з’єднанню, але не місце розташування чи причини виникнення перевантаження. Просте збільшення ємності буферів призводить до збільшення кількості пакетів, призначених для повторної передачі, та неприпустимого зростання затримок обслуговування.

Проблема виявлення перевантажень на основі контролю кількісного значення довжини черги полягає в тому, що ідентифікатор перевантаження може вказувати тільки на наявність перевантаження по з’єднанню, але не місце розташування чи причини виникнення перевантаження. Крім того, спроба боротьби з перевантаженнями шляхом простого збільшення ємності буферу не веде до розв'язання проблеми, а навпаки, веде до розбухання буферу та неприпустимого зростання затримок обслуговування.

В посібнику розглянуто метод оптимального управління перевантаженнями мережі з використанням нейронної мережі як системи моніторингу та керування.

В схемі багатокрокового передбачення стану черги для передбачення та завчасного виявлення перевантаження використаний апарат загальної теорії чутливості з непрямим зворотним зв’язком та керуванням активності джерел повідомлень. Результати цієї теорії використані для побудови системи керування з непрямим зворотним зв’язком, що дозволяє економити канальний та обчислювальний ресурси.

Представлено алгоритм навчання нейронної мережі і формування признаку перевантаження мережі. Принциповою відмінністю є багатошарова динамічна нейронна мережа з комбінацією явних та прихованих шарів. Завдяки вибору такої архітектури, по-перше, спрощується процедура пошуку оптимальних вагових

коефіцієнтів, а по-друге – прискорюється процес навчання нейронної мережі класичним методом зворотного розповсюдження похибки.

З довершеною архітектурою нейромережі, придатної для моделювання динаміки системи, можна отримати уповні задовільну продуктивність системи як об’єкту управління.

Розглянуті рішення ґрунтуються на використанні розвинутої архітектури нейронної мережі, тому важливим є вивчення алгоритмів навчання нейронної мережі для динамічних процесів передачі даних.

СПИСОК ЛІТЕРАТУРИ

1. Tanenbaum A.S., Computer Networks, 5th Ed./ Andrew S. Tanenbaum, David J. Wetherall. – Prentice Hall, Cloth, 2011. – 960 p.

2. Толубко В. Б. Методи оптимізації / В. Б. Толубко, Л. Н. Беркман. – Київ : ДУТ, 2016. – 442 с

3. Бакланов И. Г. NGN: принципы построения и организации / И. Г. Бакланов ; под ред. Ю. Н. Ченышова. – Москва: Эко-Трендз, 2008. – 400 с.

4. Stallings W. Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud / W. Stallings. – Pearson Education, Inc., Old Tappan, New Jersey, 2016. - 538 pp.

5. Лесная Н. Н. Сравнительный анализ методов оценки характеристик интеллектуальной сети // Наукові записки Українського науково-дослідного інституту зв'язку. – 2009. – №2(10). – С. 97-102.

6. Виноградов Н. А. Анализ потенциальных характеристик устройств коммутации и управления сетями новых поколений / Н. А. Виноградов // Зв'язок. – 2004. – №4. – С. 10-17.

7. Королькова А. В. К вопросу о классификации алгоритмов RED / А. В. Королькова, Д. С. Кулябов, А. И. Черноиванов // Вестник Российского университета дружбы народов.– 2009. – №3. – С. 34-46.

8 Максимов В. В. Класифікація алгоритмів боротьби з перевантаженнями / В. В. Максимов, С. О. Чмихун // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – № 5(33). – С. 73-79.

9. Максимов В. В. Дослідження алгоритму боротьби з перевантаженнями TCP VENO / В. В. Максимов, С. О. Чмихун // Телекомунікаційні та інформаційні технології. – 2015. – №4. – С. 30-36.

10. Vinogradov N. A. Comparative analysis of the functionals of optimal control corporate computer network / Nick A. Vinogradov, Alina S. Savchenko. – Journal of Qafqaz University (Mathematics and Computer Science). – 2013, Vol. 1, Nr. 2. – PP. 156-167.

11. Козелкова Е. С. Управление потоками данных в цифровых телекоммуникационных сетях с разнородным трафиком / Е. С. Козелкова, Я. И. Торошанко, Л. А. Харлай // Вісник Національного університету «Львівська політехніка». Серія «Радіоелектроніка та телекомунікації». – 2016. – №819.

12. Торошанко Я. І., Хобта Б. М., Хобта П. М. Засоби управління перевантаженнями в комп'ютерних мережах // Телекомунікаційні та інформаційні технології.– 2016. – №4. – С.100-106.

13. Саймон Х. Нейронные сети: полный курс / Х. Саймон. – 2-е изд., испр.: Пер. с англ. – Москва : ООО “И.Д. Вильямс”, 2006. – 1104 с.

14. Нейронные сети. STATISTICA Neural Networks : Методология и технологии современного анализа данных / [Под ред. В. П. Боровикова]. –2-е изд. – Москва : Горячая линия - Телеком, 2008. – 392 с.

15. Торошанко Я. І., Якимчук Н. М. Статистичні моделі управління телекомунікаційними мережами та методи боротьби з перевантаженнями // Телекомунікаційні та інформаційні технології. – 2017. – №3(56). – С. 111-118.

16. Торошанко Я. І., Якимчук Н. М. Аналіз і моделювання різнорідного самоподібного трафіку комп’ютерних мереж // Телекомунікаційні та інформаційні технології. – 2017. – №4(57). – С. 42-51.

17. Чанг Шу. Математические модели алгоритмов регулирования и формирования трафика / Чанг Шу // Проблеми інформатизації та управління. – 2007. № 1(19). – С. 154-162.

18. Торошанко Я. І. Аналіз чутливості систем масового обслуговування на основі моделі адаптації і регулювання зовнішнього трафіка / Я. І. Торошанко // Вісник Хмельницького національного університету. – 2016. – №6(243). – С. 171-175.

19. Мельников Д. А. Информационные процессы в компьютерных сетях / Д. А. Мельников. – Москва. – 1999. – 256 с.

20. Гольдштейн Б. С. Сети связи / Б. С. Гольдштейн, Н. А. Соколов, Г. Г. Яновский. – Санкт-Петербург: «БХВ – Петербург», 2014. – 400 с.

21. Mao G. Connectivity of Communication Networks / G. Mao.– Springer International Publishing AG, 2017.– 435 p.

22. Виноградов Н. А. Анализ нагрузки на сети передачи данных в системах критичного применения / Н. А. Виноградов, В. И. Дрововозов, Н. Н. Лесная, А. С. Зембицкая // Зв'язок. – 2006. – № 1 (61). – С. 9-12.

24. Томович Р. Общая теория чувствительности / Р. Томович Р., М. Вукобратович. – Москва: Советское радио, 1972. – 240 с.

25. Shooman M.L. Reliability of Computer Systems and Networks – Fault Tolerance, Analysisand Design / M.L. Shooman. – JohnWiley&Sons, Inc., NewYork, 2002. – 546 p.

26. Lu Z. Overload Control for Signaling Congestion of Machine Type Communications in 3GPP Networks / Zhaoming Lu, Qi Pan, Luhan Wang, Xiangming Wen // PLOS ONE. – December 9, 2016. – 11 p. DOI:10.1371/journal.pone.0167380.

27. Тархов Д. А. Нейросетевые модели и алгоритмы / Д. А. Тархов. – Москва: Издательство «Радиотехника», 2014. – 352 с.

28. Галушкин А. И. Нейронные сети: основы теории / А. И. Галушкин. Москва: Горячая линия – Телеком, 2010. – 496 с.

29. Kurose J. F. Computer Networking: A Top-Down Approach, 7th Ed / James F. Kurose, Keith W. Ross. – Pearson Education, Inc., 2017.– 864 p.

30. Göransson P. Software Defined Networks: A Comprehensive Approach, 2nd ed. / Paul Göransson, Chuck Black, Timothy Culver.– Morgan Kaufmann, US, 2017. –