

УДК 004.7:519.87(043.3):621.391
DOI: 10.15587/1729-4061.2019.164087

Розробка методу контролю перевантаження телекомунікаційної мережі на основі нейронної моделі

М. А. Віноградов, М. М. Степанов, Я. І. Торошанко, В. М. Черевик,
А. С. Савченко, В. М. Гладких, О. С. Торошанко, Т. В. Уварова

Розглянута схема контролю перевантажень з використанням зворотного зв'язку по знаку функції чутливості продуктивності телекомунікаційної мережі. Для визначення даної функції запропоновано використання простої нейронної мережної моделі динамічної системи. Контроль наявності або загрози перевантаження здійснюється на основі аналізу довжини черги на стороні приймача інформації. Для аналізу системи визначена функція вартості як цільова функція наявності перевантаження. Запропонований алгоритм оптимального управління забезпечує формування керуючого сигналу таким чином, щоб вихід системи відповідав як можна ближче заздалегідь встановленим характеристикам – ключовим показникам ефективності мережі. Розроблена схема контролю перевантаження зі зворотним зв'язком по знаку чутливості функції продуктивності системи. Знак чутливості продуктивності надає оптимальний напрям для налаштування швидкості джерела даних.

Запропонована нейронна модель для багатокрокового передбачення стану черги на стороні приймача телекомунікаційної мережі. Якщо нейронна мережа налаштована на відслідковування динаміки системи і показує, що квадратична помилка є незначною, вважається, що виконаний крок відповідає наперед передбаченому виходу системи.

Запропонований алгоритм адитивного збільшення / множинного зменшення, який визначає зміну швидкості джерела даних в залежності від знаку функції чутливості показника продуктивності. Даний алгоритм є альтернативою системи прогнозування перевантаження і керування потоком, заснованої на пороговому заповненні черги.

Проведено порівняльний аналіз ефективності керуючих схем виявлення перевантаження на основі черг і на основі функції чутливості продуктивності телекомунікаційної мережі. Показано, що величина черги і коливання швидкості джерела менші для схеми на основі чутливості, ніж для схеми на основі черги.

Результати моделювання продуктивності запропонованої схеми показують, що схема на основі функції чутливості має кращі ключові показники ефективності в порівнянні зі звичайною схемою вибору порога черги

Ключові слова: телекомунікаційна мережа, функція чутливості, нейронна мережа, динамічна система, управління чергою

1. Вступ

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.

Телекомунікаційна мережа являє собою сукупність ресурсів, що використовуються конкуруючими користувачами [1–3]. Ресурси таких мереж можна поділити на два основні класи:

- *головні*: буферна пам'ять, смуга пропускання, процесорний час;
- *побічні*: простір імен, записи таблиць, логічні канали і т.д.

Сукупність ресурсів 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 [3–6]. 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 [7–9]: "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 проведення досліджень в цьому напрямку.

2. Аналіз літературних даних і постановка проблеми

В роботах [10, 11] розглянуто метод RED боротьби з перевантаженнями, який використовується в протоколі TCP (Transmission Control Protocol) мережі Інтернет. Запропонована розгорнута класифікація способів і алгоритмів боротьби з перевантаженнями, які використовуються на сьогоднішній день.

Описані характеристики алгоритмів, визначені переваги та недоліки їх застосування в певних умовах роботи мереж. Однак згадані роботи [10, 11] носять, певною мірою, оглядовий характер. Не приводяться кількісні порівняльні оцінки ефективності/складності методу RED та інших методів, зокрема, Tail Drop, WRED (weighted random early detection) тощо. Крім того, не наведені аналітичні матеріали стосовно розділення трафіку за типами пакетів.

У роботах [8, 12] досліджено алгоритм TCP Veno. Алгоритм є досить поширеним для боротьби з перевантаженнями і здатен ефективно працювати як в проводових, так і в безпроводових мережах. Відомо, що алгоритм TCP Veno створений для безпроводових мереж з високою долею втрачених пакетів. Він намагається виділяти втрати, не пов'язані з перевантаженням, щоб не включати алгоритми боротьби з перевантаженням там, де це не потрібно. Між тим, сам метод розпізнавання характеру втрат за алгоритмом TCP Veno, по суті, є доволі тривіальним (лінійний класифікатор). Про це в роботах [8, 12] не згадується, очевидно, із-за неможливості детального аналізу внаслідок відсутності статистики достатньо великого об'єму.

В роботі [4] досліджений спосіб запобігання перевантаженням шляхом збільшення об'єму пам'яті вхідних буферів. Однак в даній роботі не проаналізовано проблему 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 пропускну 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 привести до збільшення негативних наслідків, пов'язаних з перевантаженням.

Важливу роль в системах управління потоками даних і прогнозуванні перевантаження, а також в оцінюванні і оптимізації ключових показників ефективності телекомунікаційних мереж відіграють показники чутливості складних систем. У фундаментальній роботі [13] визначена категорія чутливості складної системи як математичного показника та запропоновані логарифмічні функції чутливості. Однак при високому ступені абстракції та строгості математичного апарату в роботі [13] не наводяться підходи до використання функцій чутливості як інструментарію прикладного аналізу технічних систем.

В роботах [14, 15] обґрунтована доцільність використання математичного апарату аналізу чутливості складних систем в задачах управління перевантаженнями в комп'ютерних телекомунікаційних мережах. Результати, приведені в публікаціях [14, 15], 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.

В [16] 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 [14, 15] не конкретизовані визначення функцій чутливості телекомунікаційної мережі та методи їх ідентифікації у мережних системах управління з явним або непрямим зворотним зв'язком.

У роботі [16] 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 [17], 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. У даній роботі зроблено спробу розв'язання цих проблем.

3. Мета і задачі дослідження

Метою дослідження є вдосконалення і розробка способу контролю і прогнозування перевантаження в телекомунікаційній мережі з використанням явного зворотного зв'язку по знаку функції чутливості продуктивності мережі.

Для досягнення поставленої мети вирішувались наступні задачі:

- провести аналіз способів виявлення перевантажень на основі контролю кількісного значення довжини черги на стороні приймача, визначити їх переваги і недоліки;
- розробити і обґрунтувати схему багатокрокового передбачення стану черги і загрози перевантаження на основі нейронної моделі;
- розробити алгоритм навчання нейронної мережі і формування признаку перевантаження мережі.

4. Контроль перевантаження та багатокрокового передбачення стану телекомунікаційної мережі

4. 1. Контроль перевантаження за показником довжини черги

Найбільш ефективні способи і алгоритми контролю перевантаження реалізуються за принципом явного зворотного зв'язку [1, 2, 16].

Одним із поширених способів управління потоком даних за принципом зворотного зв'язку є двійковий бітовий контроль [18, 19]. 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 (контроль із кінця в кінець).

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 (темپ 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. 1). При одночасній реалізації S інформаційних процесів (дейтаграмний спосіб передачі) пакети від S джерел $D_1 \dots D_S$ надходять у відповідні черги на обслуговування, з яких через схеми регулювання видачі пакетів $РВП_1 \dots РВП_S$ 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 "підтвержене перевантаження".

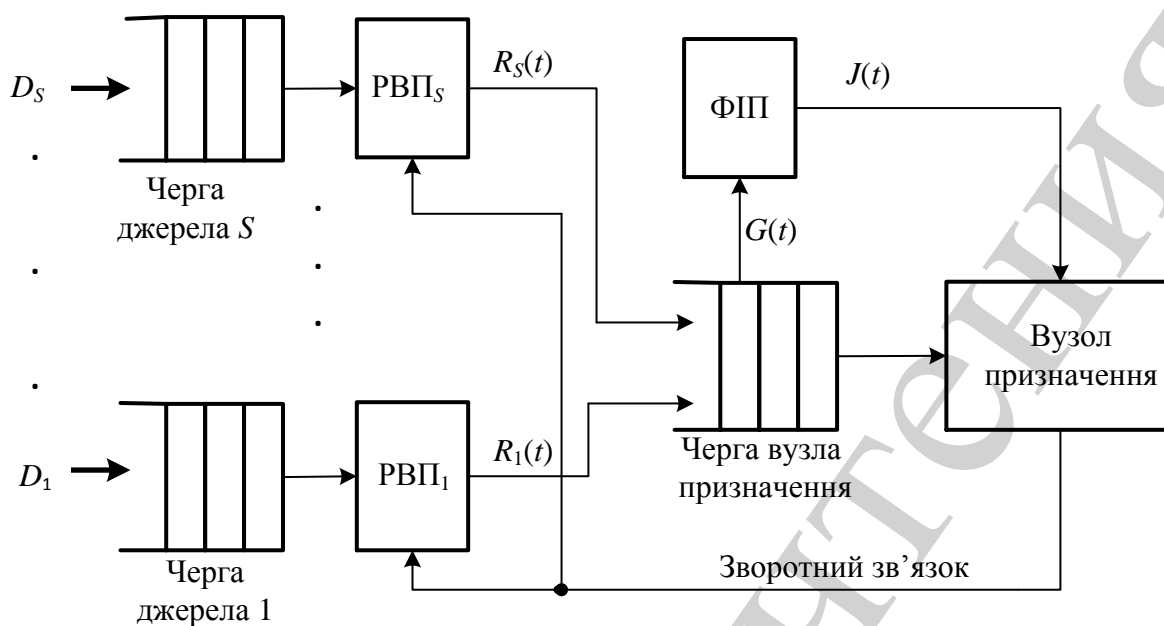


Рис. 1. Встановлення біта перевантаження на основі довжини черги

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. 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:

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

де R – швидкість заповнення черги на вході вузла призначення; $F^+ > 0$ – фактор адитивного збільшення; $0 < F^- < 1$ – фактор мультиплікативного зниження; $E(t-t) = Q - G(t-t)$ – функція помилки; t_d – час затримки передачі стану черги затору до джерела.

4. 2. Нейронна мережна модель визначення перевантаження

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 [15]. Запропонована схема для визначення функції чутливості використовує просту нейронну мережу модель динамічної системи.

Нейронні мережі представляють собою математичні структури, здатні до самонавчання на основі отримуваної ззовні інформації [20].

Нехай динаміка системи обслуговування виражена наступним рівнянням введення-виведення:

$$G(t+1) = f[G(t), G(t-\tau), \dots, G(t-l\tau), R(t), R(t-\tau), \dots, R(t-m\tau)], \quad (2)$$

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 ; $f[\cdot]$ – 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, $Q(t) = Q = \text{const}$). Послідовно-паралельна нейронна модель [16] невідомої системи (2) може бути представлена як

$$\hat{G}^i(t+i\tau) = \hat{f}^i[G(t), \dots, G(t-l\tau), R(t), \dots, R(t-m\tau)], \quad i = 1, 2, \dots, L, \quad (3)$$

де $\hat{G}^i(t+i\tau)$ – вихід нейронної мережі; \hat{f}^i – оцінка функції f ; L – горизонт передбачення.

На рис. 2 показана схема багатокрокового передбачення стану черги на основі нейронної системи.

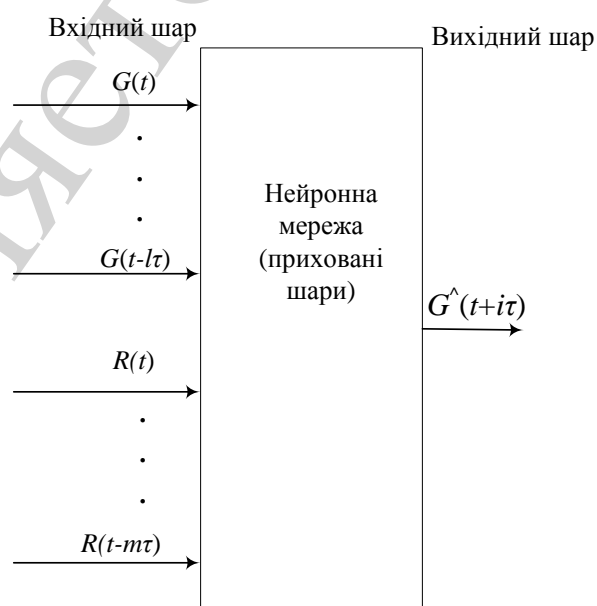


Рис. 2. Нейронна модель передбачення стану черги

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

$$[G(t+i\tau) - G^\wedge(t+i\tau)]^2 = \varepsilon \quad (4)$$

є незначною, вважається, що i -й крок відповідає наперед передбаченому виходу системи (2). Як результат, керуючий сигнал може бути обраний таким чином, що $G^\wedge(t+i\tau)$ знаходиться якомога ближче до Q .

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

$$J = \frac{1}{2} e^2(t+i\tau) = \frac{1}{2} [Q - G^\wedge(t+i\tau)]^2, \quad i=1, 2, \dots, L. \quad (5)$$

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:

$$R(t+\tau) = R(t) + \Delta R(t) = R(t) - \eta \frac{\partial J}{\partial R(t)}, \quad (6)$$

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 [8, 22].

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

Диференціюючи (5) по функції $R(t)$, отримуємо:

$$\Delta J(t) = \frac{\partial J}{\partial R(t)} = -e(t+i\tau) \frac{\partial G^\wedge(t+i\tau)}{\partial R(t)}, \quad i=1, 2, \dots, L, \quad (7)$$

де вираз $\partial G^\wedge(t+i\tau)/\partial R(t)$ відомий, як чутливість чи градієнт системи [7, 15].

Підставляючи (7) в (6), отримуємо:

$$R(t+\tau) = R(t) + \eta e(t+i\tau) \frac{\partial G^\wedge(t+i\tau)}{\partial R(t)}, \quad i=1, 2, \dots, L. \quad (8)$$

Гradient системи може бути аналітично оцінено, використовуючи відому структуру нейронної мережі [20].

Вираз (6) може бути представлений у вигляді

$$R(t + \tau) = R(t) - \eta \operatorname{sgn}[\Delta J(t)], \quad (9)$$

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

Таким чином, можемо зробити висновок, що (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 :

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

Значення ΔJ , обчисленого за допомогою формули (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.

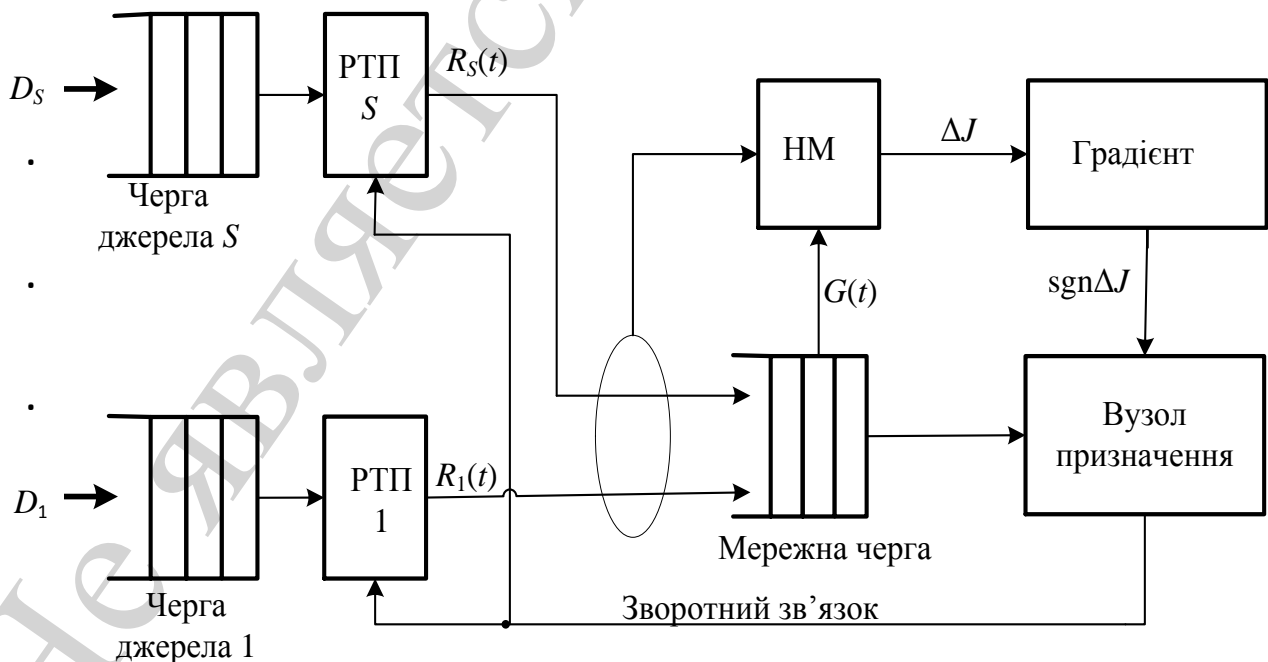


Рис. 3. Оптимізаційна схема управління перевантаженням

Поточні значення швидкостей регульованих вхідних потоків $R_1(t) \dots R_s(t)$ поступають на вхідний шар нейронної мережі НМ. На основі аналізу значень $R_1(t) \dots R_s(t)$ 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 $\text{sgn} \Delta J$ враховується вузлом призначення при формуванні сигналу зворотного зв'язку.

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

$$R(t + \tau) = \begin{cases} R(t) + F^+, & \Delta J(n - \tau) < 0, \\ R(t) - F^-, & \Delta J(n - \tau) \geq 0. \end{cases} \quad (10)$$

4. 3. 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 [22].

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:

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

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

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

Крок 4: Використовуючи вираз

$$[G^{\wedge}(t) - G(t)] w_{ij}^{(p)} \Leftarrow w_{ij}^p + \Delta w_{ij}^p(t)$$

для оновлення ваг нейронної мережі, виконується обчислення:

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

для

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

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 (\bar{f}) досягнення розміру черги і швидкості джерела:

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

– 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/одиночку часу, одиниця часу $\tau=0,25$ ms, мінімальна швидкість $R_{\min}=100$ пакетів/одиночку 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 – 6 τ . В дослідженні передбачається два вузьких місця:

- синусоїдальне $\text{integer}[35(1+\sin(2(\pi\tau)))]=10$ пакетів/одиночку часу;
- випадкове (максимальне значення 80 пакетів/одиночку 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. 1, де G_{av} і R_{av} позначають середній за часом розмір черги і швидкість джерела.

Таблиця 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
Гradient: 1-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
Гradient: 3-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. Окремі аспекти цієї задачі розглядалася в деяких попередніх роботах авторів [4, 15].

5. Обговорення результатів дослідження способу контролю і прогнозування перевантаження мережі

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 ве-

ктору 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.

6. Висновки

1. 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.

2. В результаті проведених досліджень розроблено новий метод оптимального управління перевантаженнями телекомунікаційної мережі з використанням нейронної мережі як системи моніторингу та керування. Запропонована і обґрунтована схема 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 каналний та обчислювальний ресурси.

3. Розроблений алгоритм навчання нейронної мережі і формування признаку перевантаження мережі. Принциповою відмінністю отриманих результатів є розробка багатосарової динамічної нейронної мережі 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. У результаті перевірки теоретичних результатів шляхом комп'ютерного моделювання отримані кількісні порівняльні оцінки ефективності (точності та потрібного обчислювального ресурсу) розробленого методу та методів, що існували раніше. Можна стверджувати, що з довершеною архітектурою нейромережі, придатної для моделювання динаміки системи, можна отримати уповні задовільну продуктивність телекомунікаційної системи як об'єкту управління.

Література

1. Tanenbaum A. S., Wetherall D. J. Computer Networks. 5th ed. PrenticeHall, 2010. 960 p.
2. Stallings W. Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud. Pearson Education, Inc., OldTappan, New Jersey, 2016. 544 p.
3. Mao G. Connectivity of Communication Networks. Springer, 2017. 435 p. doi: <https://doi.org/10.1007/978-3-319-52989-9>
4. Анализ нагрузки на сети передачи данных в системах критичного применения / Виноградов Н. А., Дровозов В. И., Лесная Н. Н., Зембицкая А. С. // Зв'язок. 2006. № 1 (61). С. 9–12.
5. Снарский А. А., Ландэ Д. В. Моделирование сложных сетей: учеб. пос. Киев: Инжиниринг, 2015. 212 с.
6. Bonaventure O. Computer Networking: Principles, Protocols and Practices. 2018. 272 p. URL: <https://www.computer-networking.info/2nd/cnp3bis.pdf>
7. Keshav S. Congestion Control in Computer Networks. University of California, 1991.
8. Kurose J. F., Ross K. W. Computer Networking: A Top-Down Approach. 7th ed. Pearson Education, Inc., 2017. 864 p.
9. Göransson P., Black C., Culver T. Software Defined Networks: A Comprehensive Approach. 2nd ed. Morgan Kaufmann, US, 2017. 436 p.

10. Королькова А. В., Кулябов Д. С., Черноиванов А. И. К вопросу о классификации алгоритмов RED // Вестник Российского университета дружбы народов. Серия: Математика, информатика, физика. 2009. № 3. С. 34–46.
11. Максимов В. В., Чмихун С. О. Класифікація алгоритмів боротьби з перевантаженнями // Наукові записки Українського науково-дослідного інституту зв'язку. 2014. № 5 (33). С. 73–79.
12. Максимов В. В., Чмихун С. О. Дослідження алгоритму боротьби з перевантаженнями TCP VENO // Телекомунікаційні та інформаційні технології. 2015. № 4. С. 30–36.
13. Томович Р., Вукобратович М. Общая теория чувствительности. Москва: Советскоерадио, 1972. 240 с.
14. Shooman M. L. Reliability of Computer Systems and Networks – Fault Tolerance, Analysis and Design. John Wiley & Sons, 2002. 546 p.
15. Торошанко Я. І. Управління надійністю телекомунікаційної мережі на основі аналізу чутливості складних систем // Телекомунікаційні та інформаційні технології. 2016. № 3. С. 31–36.
16. Торошанко Я. І. Аналіз чутливості систем масового обслуговування на основі моделі адаптації і регулювання зовнішнього трафіка // Вісник Хмельницького національного університету. 2016. № 6 (243). С. 171–175.
17. Overload Control for Signaling Congestion of Machine Type Communications in 3GPP Networks / Lu Z., Pan Q., Wang L., Wen X. // PLOS ONE. 2016. Vol. 11, Issue 12. P. e0167380. doi: <https://doi.org/10.1371/journal.pone.0167380>
18. Мельников Д. А. Информационные процессы в компьютерных сетях. Москва, 1999. 256 с.
19. Гольдштейн Б. С., Соколов Н. А., Яновский Г. Г. Сети связи. Санкт-Петербург: «БХВ – Петербург», 2014. 400 с.
20. Тархов Д. А. Нейросетевые модели и алгоритмы. Москва: Издательство «Радиотехника», 2014. 352 с.
21. Стеклов В. К., Беркман Л. Н., Кільчицький Є. В. Оптимізація та моделювання пристроїв і систем зв'язку. Київ: Техніка, 2004. 576 с.
22. Галушкин А. И. Нейронные сети: основы теории. Москва: Горячая линия – Телеком, 2010. 496 с.