

МЕТОД ОПТИМІЗАЦІЇ ПРОЦЕСУ ПЕРЕДАЧІ ДАНИХ ПІД ЧАС СЕАНСІВ ЗВ'ЯЗКУ TCP-ПРОТОКОЛУ

Национальный авиационный университет

Проведено аналіз існуючих версій протоколу TCP, а також перспективних рішень в області управління передачею пакетів на транспортному рівні еталонної моделі взаємодії відкритих систем (EMBBS). Виявлено ряд недоліків математичних моделей і методів, покладених в основу даного протоколу, усунення яких пов'язане з оптимізацією відповідних мережових параметрів. Запропоновано підхід до рішення задачі оптимізації, заснований на динамічній моделі TCP-сеансів і ряд умов-обмежень, направлених на забезпечення заданих показників якості і стійкості TCP-сеансів.

Актуальність теми

Сучасні мультисервісні телекомунікаційні мережі (ТКС) розвиваються в напрямку впровадження концепції мереж наступного покоління NGN (NextGeneration Network) [1, 2]. Ефективність ТКС в основному залежить від результативності постановки та вирішення завдань, пов'язаних з управлінням мережевими ресурсами. В основу NGN відповідно до рекомендації Міжнародного союзу електрозв'язку серії Y 2000 закладається стек протоколів TCP/IP [3], в рамках якого важливу роль відіграють функції, що виконуються транспортним рівнем EMBBS. Зокрема, основними завданнями протоколу TCP (Transmission Control Protocol) є гарантія успішної доставки даних та забезпечення обміну даними між процесами, що виконуються на кінцевих системах (пристроях, вузлах), за допомогою служби обміну даними, що надається протоколом мережевого рівня [4]. Протокол TCP є ефективним засобом управління потоками даних, мережі збалансованого навантаження, боротьби з перевантаженнями, забезпечення наскрізних показників якості та ін.. Однак його ефективність багато у чому залежить від коректності вибору тих чи інших параметрів як самого TCP протоколу, так і інших взаємодіючих з ним засобів боротьби з перевантаженнями в ТКС.

Дослідження моделей і методів, покладених в основу базового протоколу TCP і його множинних версій (TCP-Reno, TCP Vegas, TCP Tahoe, TCP Westwood, ScalableTCP, ModbusTCP і т.д.) [3, 4], показав ряд властивих їм недоліків, з метою мінімізації яких сформовані наступні вимоги, що ставляться до перспективних рішень:

1) облік динамічності процесів передачі даних на транспортному рівні ТКС;

2) врахування особливостей передачі одночасно декількох TCP-потоків, що властиво для реальних мультисервісних ТКС, що використовують для передачі даних протокол TCP (багато протоколів прикладного рівня, наприклад, FTP, Telnet, HTTP, POP3, SMTP та ін. одночасно працюють з цим протоколом);

3) облік зміни режимів передачі даних в рамках процесу функціонування протоколу TCP;

4) забезпечення узгодженого використання механізмів боротьби з перевантаженнями і протоколу транспортного рівня TCP, що дозволяє оперативно реагувати на ймовірні перевантаження в мережі і відповідно, уникнути необґрунтованих втрат пакетів;

5) орієнтація на різні версії TCP, що повинно надати гнучкість у використанні тієї чи іншої існуючої реалізації даного протоколу, а також орієнтувати на використання в перспективних рішеннях у цій галузі;

6) можливість вибору різних моделей блокування (відкидання) пакетів в механізмах боротьби з перевантаженнями, що дозволить адаптуватися під структурні та функціональні особливості мережі та характеристики переданих трафіків, своєчасно реагувати на зміни цих параметрів;

7) забезпечення стійкості TCP-сеансів за рахунок формулювання відповідних умов в вигляді додаткових обмежень в рамках існуючих моделей і методів мережевого управління або як деякої цільової функції.

Постановка задачі

Актуальним завданням є оптимізація TCP-сеансів у мультисервісних ТКС, що пов'язано з необхідністю обґрунтованого вибору чисельних значень параметрів протоколу TCP і параметрів взаємодіючих з ним засобів боротьби з перевантаженнями, таких як RED (Random Early

Detection), WRED (Weighted Random Early Detection) і т.д.

Шляхи вирішення

Для математичного опису ТКС на сьогоднішній день існує безліч підходів, в основу яких покладено використання можливостей різних математичних апаратів [5]. Серед них можна виділити моделі прийняття оптимальних рішень, моделі керуючих автоматів, імовірнісні та агрегативної моделі складних систем. Кожен з математичних апаратів визначає свій підхід до формалізації задач, пов'язаних з процесами маршрутизації, управління мережевими параметрами і процесами інформаційного обміну, які, в кінцевому рахунку, формулюються як оптимізаційні задачі.

Поряд з перерахованими математичними підходами, заслуговує на увагу апарат диференціально-різницевого рівняння стану мережі, який досить точно відповідає пред'явленим вимогам. В рамках моделі простору станів задовольняються вимоги з обліку динамічності процесів (в тому числі і TCP-сеансу), що протікають в ТКС, зміна мережових параметрів (структурних і функціональних), а також обліку можливості одночасного існування декількох потоків в мережі.

Розглянемо процес інформаційного обміну в IP-мережі, де при установці і підтримці TCP-сеансу з метою запобігання можливих перенавантажень на проміжних вузлах зв'язку і подальших втрат пакетів, використовуються додаткові алгоритми RED/WRED.

Дослідження різних версій даного протоколу показало, що в процесі роботи TCP можна виокремити кілька стандартних режимів (фаз) передачі даних (Slow Start, Congestion Avoidance, Fast Recovery) [3, 4]. Перехід в той чи інший режим, починаючи з встановлення з'єднання, залежить від ступеню завантаженості каналів зв'язку, буферів проміжних і приймального пристроїв, кількості помилок і спотворень даних. Тоді модель TCP-сеансів у просторі станів можна представити як систему неавтономних диференціальних рівнянь, що відображає динаміку передачі даних між вузлом-відправником і вузлом-одержувачем при наявності в каналі декількох TCP-потоків:

$$\frac{d\lambda_i(t)}{dt} = \begin{cases} (1 - P(t)) \cdot \frac{MSS}{RTT} \cdot \lambda_i(t) - P(t) \cdot (\lambda_i(t))^2 + \\ + P(t) \cdot MSS \cdot \lambda_i(t), \\ (1 - P(t)) \cdot \left(\frac{MSS}{8 \cdot RTT} \cdot \lambda_i(t) + \frac{MSS^2}{RTT^2} \right) - \\ - P(t) \cdot (\lambda_i(t))^2 + P(t) \cdot MSS \cdot \lambda_i(t) \end{cases} \quad (1)$$

де $\lambda_i(t)$ – інтенсивність TCP-потoku, що передається в i -му сеансі, $i = \overline{1, M}$; MSS – максимальний розмір сегмента (Maximum Segment Size); RTT – період кругового звернення пакета (Round Trip Time); $P(t)$ – ймовірність блокування (відкидання) пакетів.

В рамках запропонованої моделі ймовірність блокування (відкидання) пакетів $P(t)$ може представлятися довільною (але абгрунтованою) функцією. В окремому випадку, при використанні в якості механізму обмеження довжини черги алгоритму RED, ймовірність блокування визначається згідно рівності [6]:

$$P(t) = \frac{1}{m} \cdot \frac{N(t) - N_{min}}{N_{max} - N_{min}} \quad (2)$$

де m – знаменник граничної ймовірності; $N(t)$ – поточний середній розмір черги; N_{min} , N_{max} – відповідно мінімальний і максимальний середній розмір черги.

Черга на мережевому вузлі (маршрутизаторі) включає пакети сумарного потоку, що дозволяє врахувати розподіл пропускної спроможності між TCP-сеансами в межах одного каналу. У свою чергу, система (1) дає можливість застосування до різних TCP-потоків пріоритетизації на основі виділення різної пропускної здатності.

У механізмі WRED, передбачається використання окремих значень граничної ймовірності та ваг для різних IP-пріоритетів, дозволяючи забезпечувати різні рівні якості обслуговування для різних типів трафіку в періоди переповнення черг мережових вузлів і каналів.

У моделі передбачається, що середній розмір черги на мережевому вузлі можна розрахувати згідно положенням теорії масового обслуговування, виходячи з формули Літтла [7, 8]:

$$N(t) = \frac{\lambda(t)}{B - \lambda(t)} \cdot \frac{\lambda(t)}{B} \quad (3)$$

де B – пропускна здатність каналу зв'язку.

Метод оптимізації процесу передачі даних в рамках TCP-сеансів

В основу запропонованого методу, пов'язаного з оптимізацією вибору параметрів протоколу TCP і алгоритмів RED/WRED, покладено рішення завдання оптимізації, в ході якого здійснюється мінімізація (максимізація) деякого цільового функціоналу втрат. З точки зору пошуку компромісу між необхідністю врахування фізики процесів передачі даних в рамках TCP-сеансів, з одного боку, і можливістю отримання чисельних результатів розрахунку в

рамках відомих оптимізаційних процедур, з іншого боку, скористаємося критерієм, що визначає мінімум функціоналу:

$$J = \min \left[\sum_{i=1}^M \alpha_i \cdot \left(\lambda_i^{\text{внм}} - \frac{1}{T} \int_0^T \lambda_i(t) dt \right) \right] \quad (4)$$

де α_i – пріоритет i -го TCP -потoku; T – часовий інтервал оптимізації, який, як правило, відповідає середній тривалості TCP -сеансу в мережі.

Використання критерію (4) гарантує мінімізацію зважених щодо пріоритету відмов в обслуговуванні на транспортному рівні.

Після перетворень вираз (4) можна привести до вигляду:

$$J = \min \left[\sum_{i=1}^M \alpha_i \cdot \left(\lambda_i^{\text{внм}} - \frac{1}{T} \int_0^T \lambda_i(t) dt \right) \right] \quad (5)$$

Фізичний зміст функціоналу (5) тепер уже полягає в максимізації середньої сумарно зваженої щодо пріоритету інтенсивності TCP -потоків на інтервалі T .

При вирішенні задачі оптимізації в якості керуючих параметрів виступають параметри протоколу TCP (Величина MSS , інтервал RTT , розмір вікна перевищення значення порогу, режимів передачі), а також параметри алгоритмів боротьби з перевантаженнями $RED/WRED$ (величини N_{\min} , N_{\max} , m). На практиці можна оптимізувати тільки частину з перерахованих параметрів, при цьому інші характеристики фіксувати на деяких значеннях.

Враховуючи мультисервісний характер сучасних ТКС в ході вирішення завдання оптимізації, крім виконання умови, пов'язаної з динамікою TCP -сеансів, необхідно враховувати вимоги до показників якості обслуговування (*Quality of Service, QoS*). Кількісна оцінка показників *QoS* може бути визначена, наприклад, на підставі рекомендацій *Y 1540, Y 1541* Міжнародного союзу електрозв'язку.

З метою формулювання умов забезпечення гарантованої якості обслуговування введені обмеження за тимчасовими показниками *QoS*:

$$\sum_{j=1}^K \tau_i^j(t) \leq \tau_i^{\text{внм}} \quad (6) \quad \sum_{j=1}^K \sigma_i^j(t) \leq \sigma_i^{\text{внм}} \quad (7)$$

і за показниками надійності:

$$\prod_{j=1}^K P_i^j(t) \leq P_i^{\text{внм}} \quad (8)$$

де параметр K визначає кількість ділянок мережі між вузлом-джерелом і вузлом-одержувачем; $\tau_i^j(t)$, $\sigma_i^j(t)$, $P_i^j(t)$ – поточні значення середньої затримки, джиттера і ймовірності втрат пакетів

в j -му каналі зв'язку TCP -з'єднання для i -го TCP -потoku; $\tau_i^{\text{внм}}$, $\sigma_i^{\text{внм}}$, $P_i^{\text{внм}}$ – вимоги, що пред'являються до міжкінцевих значень тих же показників *QoS*.

При цьому у виразі (6) формалізовано вимогу до міжкінцевої (сумарної по всіх ділянках мережі між вузлом-джерелом і вузлом-одержувачем) середньої затримки пакетів в мережі, в виразі (7) – до джиттеру, а співвідношення (8) відображає вимога щодо величини ймовірності відкидання (блокування) пакетів.

Дослідження і подальше забезпечення стійкості процесів інформаційного обміну та управління мережевими ресурсами, протікають в реальних ТКС, істотно впливає на ефективність їх роботи. Це дозволяє виявити і попередити раптове переповнення черг мережеских вузлів і каналів зв'язку, розриви сеансів зв'язку та інші фактори, що призводять до нестійкості TCP -сеансів при незначних змінах структурних та функціональних параметрів. Дослідження відомих підходів до аналізу стійкості динамічних систем, до яких відноситься TCP -сеанс, що описуються нелінійними диференціальними рівняннями, дозволив зробити висновок на користь застосування можливостей теорії бифуркацій [9, 10].

Таким чином, реалізація методу оптимізації процесу передачі даних в рамках TCP -сеансів припускає наступну послідовність дій:

1. Моніторинг та збір вихідних даних про поточні значення TCP -з'єднання, режим передачі в протоколі TCP , параметрів алгоритмів запобігання перевантаження ($RED/WRED$) на вузлах мережі.

2. На основі аналізу зібраних вихідних даних про стан мережі встановлюється їх аналітичний взаємозв'язок шляхом аналізу аналітичної моделі (1).

3. Формалізація умов-обмежень по тимчасовими показниками *QoS* (6) і (7), по показникам надійності (8), а також умов забезпечення стійкості TCP -сеансів.

4. Постановка оптимізаційної задачі (5), в ході вирішення якої забезпечується мінімізація відмов в обслуговуванні трафіків користувачів і узгоджено розраховуються основні параметри протоколу TCP і алгоритми боротьби з перевантаженням, наприклад, алгоритмів $RED/WRED$.

5. Параметри протоколу TCP і алгоритмів боротьби з перевантаженням, наприклад, алгоритмів $RED/WRED$ в автоматичному режимі

підлягають коригуванню, забезпечуючи, тим самим, оптимізацію *TCP*-сеансів в цілому.

6. По закінченню інтервалу *RTT* здійснюється повторна постановка і вирішення оптимізаційної задачі (5), тобто здійснюється перехід до п.1.

У деяких випадках періодичний характер розрахунків може бути порушений, наприклад, при «нештатній» зміні стану ТКС і її *TCP*-з'єднань, викликаній зміною структури мережі, стрибкоподібній зміні інтенсивності надходження в мережу трафіку або числа *TCP*-потоків.

Висновки

Таким чином, з метою оптимізації параметрів протоколу *TCP* і взаємодіючих з ним механізмів боротьби з перевантаженням запропонований відповідний метод, пов'язаний з представленням *TCP*-сеансів системою нелінійних диференціальних рівнянь. В основу методу покладено вирішення сформульованої на виході нелінійної диференціальної моделі *TCP*-сеансів оптимізаційного завдання, де в якості критерію (5) виступав мінімум зважених щодо пріоритету відмов у обслуговуванні на транспортному рівні ТКС. Сама ж оптимізація здійснюється з урахуванням динамічних обмежень (1), а також обмежень по тимчасовим показникам *QoS* (6) і (7), по показникам надійності (8) шляхом чисельного розрахунку таких важливих параметрів як величини *MSS*, інтервалу *RTT*, режимів передачі в *TCP*, а також параметрів алгоритму боротьби з перевантаженнями.

В ході постановки оптимізаційної задачі, формулювання умов обмежень (1), (6) - (8) вироблено задоволення таких важливих вимог до протоколу *TCP*, як облік динамічності, багатопотоковості і мультирежимності процесу передачі даних; забезпечення узгодженості з

роботою інших засобів боротьби з перевантаженням з можливістю вибору тієї чи іншої моделі блокування пакетів.

Список літератури

1. Бакланов И.Г. NGN: принципы построения и организации / И.Г. Бакланов; под ред. Ю.Н. Чернышова. – М.: Эко-Трендз, 2008. – 400 с.
2. Семенов Ю.В. Проектирование сетей связи следующего поколения / Ю.В. Семенов. – СПб.: Наука и Техника, 2005. – 240 с.
3. Стивенс У.Р. Протоколы *TCP/IP*. Практическое руководство / У.Р. Стивенс; пер. с англ. и комментарии А.Ю. Глебовского. – СПб.: «Невский диалект»-«БХВ-Петербург», 2003. – 672 с.
4. Куроуз Джеймс Ф. Компьютерные сети: многоуровневая архитектура Интернета / Джеймс Ф. Куроуз, Кит В. Росс. – 2-е изд. – СПб.: Питер, 2004. – 765 с.
5. Математичні основи теорії телекомунікаційних систем / Під заг. ред. В.В. Поповського. – Х.: ТОВ "Компанія СМІТ", 2006. – 564 с.
6. Вегешна Ш. Качество обслуживания в сетях *IP*: пер. с англ. / Ш. Вегешна. – М.: Изд. дом «Вильямс», 2003. – 368 с.
7. Вишневский В.М. Теоретические основы проектирования компьютерных сетей / В.М. Вишневский. – М.: Техносфера, 2003. – 512 с.
8. Крылов В.В. Теория телетрафика и ее приложения / В.В. Крылов, С.С. Самохвалова. – СПб.: БХВ-Петербург, 2005. – 288 с.
9. Анищенко В.С. Устойчивость, бифуркации, катастрофы / В.С. Анищенко // Соросовский образовательный журнал. – 2000. – № 6. – С. 105-109.
10. Томпсон Дж. М.Т. Неустойчивости и катастрофы в науке и технике: пер. с англ. / Дж. М.Т. Томпсон. – М.: Мир, 1985. – 254 с.