

АНАЛІТИЧНА МОДЕЛЬ WEB-КЕШУВАННЯ ЛОКАЛЬНОЇ МЕРЕЖІ НА ОСНОВІ ПРОТОКОЛУ WCCP

© Кузьмін О., 2008

Запропоновано аналітичну модель для дослідження процесів WEB-кешування на основі протоколу WCCP. Наведені результати моделювання клієнтської частини WEB-системи з використанням мережевого кеш-процесора та за його відсутності.

In this work analytical model for process analysis of WEB-caching based on WCCP protocol is suggested. The results of simulation of WEB-system's client part in the presence of network cache-processor and in its absence are given.

Вступ

Зростання послуг, що надаються через Internet, призводить до зростання відповідного трафіка інформації. Це призводить, своєю чергою, до зростання часу відповіді на запити до WEB-систем. Тому розробляються нові технології доступу, які забезпечують зменшення часу відповіді. Однією з таких технологій є застосування мережевого кешування за протоколом WCCP. У роботі запропоновано аналітичну модель, яка дає змогу оцінити вплив на час відповіді до WEB-системи застосування кешування за протоколом WCCP.

1. Постановка задачі

Протокол маршрутизації кешування WCCP (Web Cache Control Protocol) локалізує мережевий трафік і забезпечує інтелектуальний розподіл навантаження між декількома мережевими кешами, що дає змогу досягти максимального завантаження вмістимого WEB-сторінок. Кеш-процесорна система складається з протоколу WCCP та одного або декількох кеш-процесорів Cisco, які зберігають дані в локальній мережі (рис. 1).

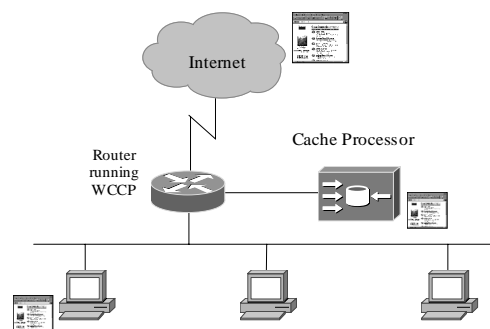


Рис. 1. Кеш-процесорна система локальної мережі

За протоколом WCCP маршрутизатор скеровує WEB-запити від робочих станцій локальної мережі кеш-процесору, а не на відповідний WEB-сервер. Кеш-процесор виконує прозоре кешування так [1].

1. Користувач запитує з броузера WEB-сторінку.
2. Маршрутизатор аналізує запит і перескерує його кеш-процесору.
3. Якщо у кеш-процесора немає відповідної сторінки, він створює з'єднання з потрібним сервером, звідки отримує необхідну WEB-сторінку і зберігає її у себе.

4. Кеш-процесор пересилає WEB-сторінку користувачу. Наступні звертання до того ж вмістимого кеш-механізму виконує прозоро, використовуючи дані з локального сховища.

Функціонування такої конфігурації локальної мережі математично можна описати за допомогою замкненої мережі систем масового обслуговування (СМО). Час відповіді на WEB-запит залежить від багатьох факторів, а саме:

- продуктивності платформи клієнта;
- продуктивності локальної мережі, яка з'єднує клієнта з маршрутизатором;
- швидкодії мережі, яка з'єднує маршрутизатор і постачальника послуг Internet;
- параметрів продуктивності маршрутизатора;
- параметрів продуктивності кеш-процесора;
- затримок, які викликаються системою передачі даних Internet;
- затримок на формування WEB-сторінки віддаленим WEB-сервером;
- затримок, пов'язаних з постачальником послуг;
- параметрів навантаження, яке створюється клієнтами локальної мережі.

Ставиться задача: побудувати аналітичну модель функціонування розглянутої конфігурації локальної мережі і дослідити залежність часу відповіді на WEB-запит від навантаження, яке створюється клієнтами локальної мережі за наявності та відсутності WEB-кешування.

2. Опис моделі

Мережева модель СМО, яка відповідає розглянутій локальній мережі, зображена на рис.2.

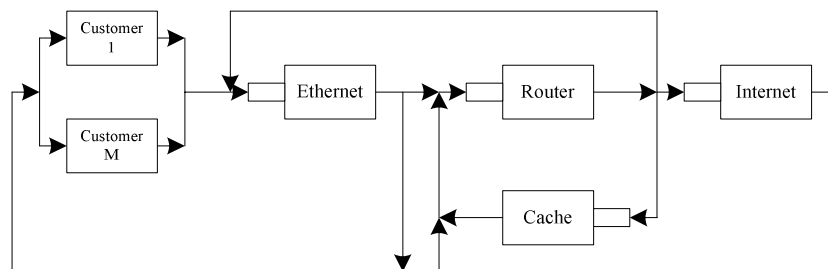


Рис. 2. Мережева модель СМО з WEB-кешуванням

WEB-запити формуються клієнтами $Customer_i$ ($i=1,M$), які представлені в моделі окремою СМО без черги. Час, який проводять запити в цій системі, відповідає часу обмірковування, витраченого клієнтом від початку отримання документа до формування нового запиту на наступний документ. СМО Ethernet моделює затримки в локальній мережі. Маршрутизатор представлений окремою СМО і враховує час, який втрачає запит на маршрутизаторі. СМО Internet враховує сумарні затримки в часі, пов'язані з затримками у вхідних і вихідних лініях зв'язку з постачальником Internet, самого Internet та віддалених WEB-серверів. СМО Cache визначає затримки у часі на обслуговування WEB-запитів. Визначимо основні параметри, за допомогою яких обчислюватимемо середній час обслуговування СМО мережі:

- V_L – пропускна здатність локальної мережі, Мбіт/с;
- T_R – час затримки пакета на маршрутизаторі, мс;
- V_P – пропускна здатність з'єднання з постачальником послуг Internet, Мбіт/с;
- T_{RTT} – середній період звертання в Internet, мс;
- V_I – середня швидкість передачі даних в Internet, Кб/с;
- T_C – середній час обмірковування клієнтом;
- M – кількість клієнтських робочих станцій;
- N_Q – середній розмір WEB-запиту, який браузер скеровує до сервера в байтах;
- N_R – середній розмір документа для всіх запитів, Кб, які сформовані клієнтом.

Згідно з визначеним в п.1 алгоритмом WEB-кешування можливі дві послідовності етапів обслуговування WEB-запитів:

1. Customer – Ethernet – Router – Cache – Router – Ethernet – Customer.
2. Customer – Ethernet – Router – Cache – Router – Internet – Router – Cache – Router – Ethernet – Customer.

Якщо прийняти, що частка результативних запитів під час звертання до Cache становить 0.2 [3], то середня кількість звертань до СМО мережі, яке припадає на обслуговування будь-якого одного WEB-запиту, становитиме: $\alpha_{customer}=1$, $\alpha_{ethernet}=2$, $\alpha_{router}=3.6$, $\alpha_{cache}=1.8$, $\alpha_{internet}=0.8$. Відповідно до цього матриця ймовірностей передач P для розглянутої моделі замкненої мережі СМО матиме вигляд:

$$P = \begin{pmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0.28 & 0 & 0.5 & 0.22 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

де $i,j=1$ відповідає СМО Ethernet; $i,j=2$ – Router; $i,j=3$ – Cache; $i,j=4$ – Internet; $i,j=5$ – Customer.

У випадку відсутності кеш-процесора відповідна мережева модель СМО представлена на рис.3, а матриця ймовірностей передач P має вигляд:

$$P = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

де $i,j=1$ відповідає СМО Ethernet; $i,j=2$ – Router; $i,j=3$ – Internet; $i,j=4$ – Customer.

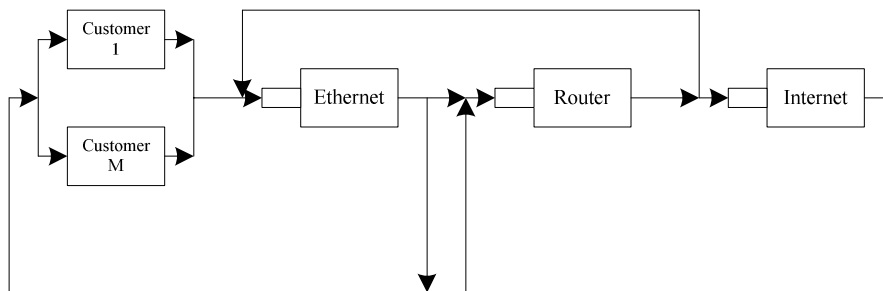


Рис. 3. Мережева модель СМО без WEB-кешування

Середній час обслуговування СМО мережі D_i визначатимемо так [2]:

$$D_{Customer} = T_C;$$

$$D_{Ethernet} = 0.5(N_Q/V_L + N_R/V_L);$$

$$D_{Router} = (0.5(N_Q + 1024N_R)/1460 + 6)T_R,$$

де 1460 – максимальний розмір сегмента TCP, константа 6 враховує розмір двох сегментів синхронізації, необхідних для встановлення TCP-з'єднання, одного TCP-сегмента даних, необхідних для передачі WEB-запиту, і трьох сегментів, які використовуються для закриття з'єднання;

$$D_{Cache} = 0.25 \cdot 10^{-3} + 6N_R,$$

де 0.25 – час, мс, необхідний кеш-процесору для обробки одного запиту, 6 – час, мс зчитування 1 Кб даних з диску;

$$D_{Internet} = 2 \cdot T_{RTT} + N_R/V_I + (N_Q + 3 \cdot 100)/V_P + (1024 \cdot N_R + 2 \cdot 100)/V_P,$$

де 100 – загальна кількість службової інформації в байтах, яка припадає на одну дейтаграму.

Для розглянутої моделі шукатимемо середній час відповіді на WEB-запит за наявності WEB-кешування та його відсутності, а також коефіцієнти завантаження СМО.

3. Аналіз результатів

Запропоновану модель досліджували для таких вхідних даних: $V_L = 10$ Мбіт/с, $T_R = 50$ мкс/пакет, $V_P = 56$ Кбіт/с (та для лінії T1 - 1.544 Мбіт/с), $T_{RTT} = 100$ мс, $V_I = 20$ Кб/с, $T_C = 3.333$ с, $M = 1 \div 10$, $N_Q = 290$ байт, $N_R = 22.23$ Кб. Для цих вхідних даних значення D_i будуть такими: $D_{Customer} = 3333$ мс, $D_{Ethernet} = 8.79$ мс, $D_{Router} = 0.695$ мс, $D_{Cache} = 133.63$ мс, $D_{Internet} = 4600$ мс (та для лінії T1 – 1433.54 мс). Залежність часу відповіді на WEB-запит від кількості активних клієнтів за наявності WEB-кешування та його відсутності наведено на рис. 4.

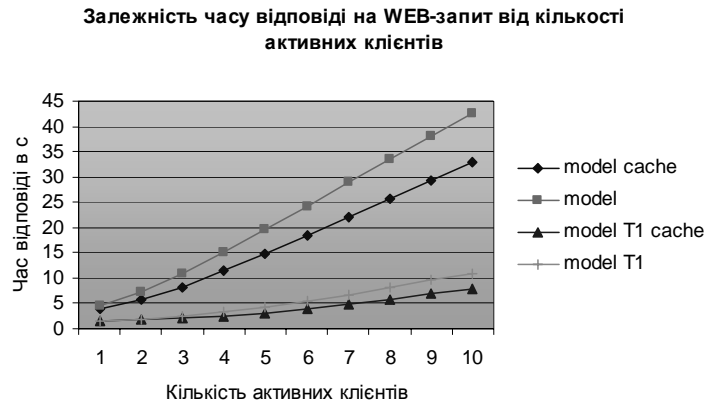


Рис. 4. Залежність часу відповіді на WEB-запит від кількості активних клієнтів

Як видно з функціональних залежностей, використання WEB-кешування зменшує час відповіді на WEB-запити. Так, при роботі 10 клієнтів вираш у часі при швидкості лінії зв'язку з постачальником послуг Internet в 56 Кбіт/с становить 23 %, а для лінії T1 (1.544 Мбіт/с) – 28 %. При тій самій кількості активних клієнтів використання лінії T1 і кешування відносно використання лінії зі швидкістю 56 Кбіт/с та відсутності кешування зменшує час відповіді на WEB-запит на 81 %. Очевидно, що із збільшенням кількості Cache Processor ці показники будуть кращими.

На рис. 5 і 6 наведені залежності коефіцієнтів завантаження відповідно Internet і Cache Processor від кількості активних клієнтів.

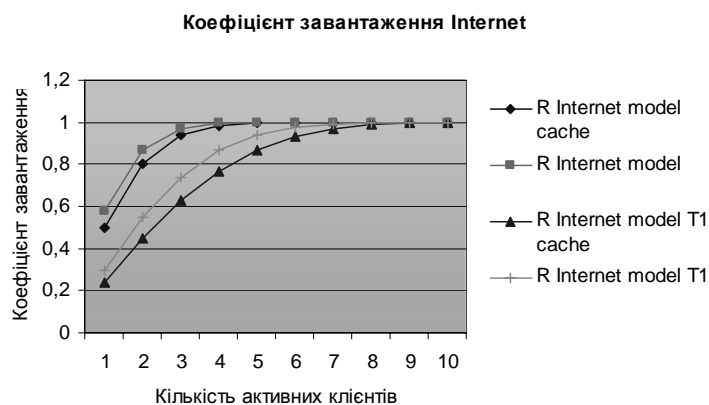


Рис. 5. Залежність коефіцієнта завантаження Internet від кількості активних клієнтів

Як видно з наведених залежностей, вузьким місцем є Internet, вплив якого на час відповіді на WEB-запит є значним. Тому використання WEB-кешування є тим шляхом, який дає змогу зменшити цей час.

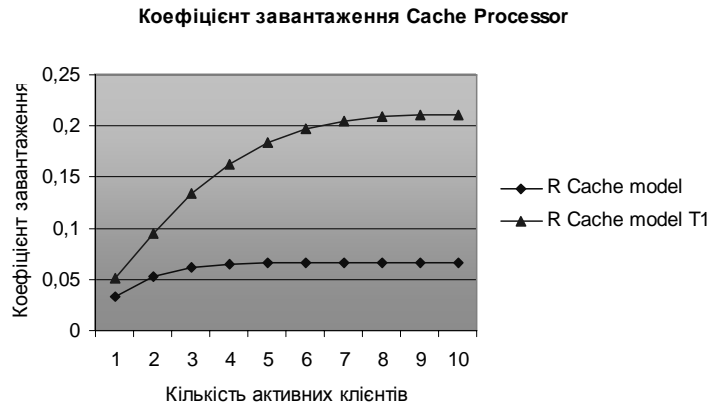


Рис. 6. Залежність коефіцієнта завантаження Cache Processor від кількості активних клієнтів

При швидкості лінії зв'язку з постачальником послуг Internet в 56 Кбіт/с коефіцієнт завантаження Cache Processor досягає свого максимального значення вже при кількості активних клієнтів, яка дорівнює чотирьом клієнтам, тоді як при швидкості в 1.544 Мбіт/с (лінія T1) цей показник досягає максимального значення при 10 активних клієнтах. Збільшення значення коефіцієнта завантаження Cache Processor при використанні лінії T1 порівняно з лінією із швидкістю 56 Кбіт/с пояснюється перерозподілом WEB-запитів між Internet і Cache Processor.

Висновки

Запропонована модель дає змогу оцінити ступінь впливу на продуктивність WEB-сервісу застосування WEB-кешування на основі протоколу WCCP. Це дає можливість оцінити якість обслуговування при зростанні кількості запитів в одиницю часу та визначити вузькі місця, які знижують продуктивність мережі. Модель і відповідні інструментальні засоби моделювання можуть використовуватися системними адміністраторами для прогнозування продуктивності WEB-сервісу при реконфігурації мережі або збільшенні трафіку запитів.

1. Cisco Systems, и др. Руководство по технологиям объединенных сетей, 3-е издание. : Пер. с англ. – М.: Издательский дом “Вильямс”, 2002. 2. Производительность Web-служб. Анализ, оценка и планирование: Пер. с англ./ Дэниел А. Менаске, Виргилио А. Ф. Альмейда. – СПб: ООО «ДиаСофтЮП», 2003. – 480 с. 3. M. R. Stadelman, “UNIXWeb Server Performance Analysis”, Proc. 1996 Comput. Measurement Group Conf., San Diego, California, Dec. 8-13, 1996, pp.1026-1033.