

УДК 681.32

А.З. МЕЛКОВ, Б.Г. ІСМАЙЛОВ

ДОСЛІДЖЕННЯ ПАРАМЕТРІВ ДИНАМІЧНИХ ПРІОРИТЕТІВ У СИСТЕМАХ ОБСЛУГОВУВАННЯ ІЗ ДВОМА ТИПАМИ ЗАЯВОК

Сумгайтський державний університет,
м. Сумгайт, Азербайджан, 35008, 43-й квартал,
Тел.: 4333714, Email: Balemi@rambler.ru

Анотація. Запропоновано й розроблено обчислювальні процедури з метою розрахунку оптимальних значень параметрів динамічних пріоритетів у системах обслуговування із двома класами заявок, де функція пріоритетності має мультиплікативний вигляд. Критерієм оптимальності є сумарна довжина черги різнотипних заявок, при цьому є обмеження на часі очікування заявок кожного типу.

Ключові слова: комутатор високошвидкісної мультимедійної мережі, пріоритет, заявка.

ВСТУП

У комутаторах високошвидкісних мультимедійних мереж із загальним буферним простором (Shared Buffering) з метою задоволення заданих рівнів якості обслуговування (Quality of Service) використовуються різні пріоритети. При цьому пріоритети, що визначають процедури прийняття в буфер різнотипних заявок, називаються просторовими пріоритетами (Space Priorities), а пріоритети, що задають правила вибору типу заявки з буфера, одержали назву тимчасових пріоритетів (Time priorities)[1].

У класичних схемах пріоритетного обслуговування, як правило, передбачається, що заявки певного типу мають одночасно високі (у порівнянні із заявками іншого типу) пріоритети обох видів. Разом з тим, у доступній літературі відомі (хоча в дуже незначній кількості) роботи, у яких вивчаються схеми з різними рівнями просторових і тимчасових пріоритетів (див. наприклад [2–4]).

Просторові пріоритети дозволяють, головним чином, управляти інтенсивностями (імовірностями) втрати різнотипних заявок, у той час як тимчасові пріоритети істотно впливають на час їхньої затримки (очікування) у буфері. У літературі запропоновані різні схеми визначення просторових пріоритетів з метою знаходження оптимальних (у відомому змісті) розмірів буферної пам'яті у вузлах різних мереж обслуговування (див. наприклад, [5-7])

Виходячи з реальних умов роботи високошвидкісних мультимедійних мереж обслуговування доцільно як тимчасові пріоритети використовувати динамічні пріоритети, що змінюються із часом. У системах з динамічними пріоритетами рішення про визначення типу заявки, обраної на обслуговування, залежить від значення деякої функції $J_i(t)$, визначальний миттєвий пріоритетний індекс заявки i -го

типу в момент t , $i = \overline{1, n}$, де n – загальна кількість типів заявки. Цю функцію аналітично можна визначити різними способами. Так, функція мультиплікативного характеру $J_i(t) = b_i \tau_{q_i}(t)$ визначена в роботі [8], де b_i – деякий коефіцієнт, що визначає швидкість зміни пріоритетності заявки i -го типу при її очікуванні в черзі; $\tau_{q_i}(t)$ – випадковий час очікування заявки i -го типу від моменту надходження до сучасного моменту t .

Аналіз динамічних пріоритетів, що залежать від часу, особливо актуальний при організації обслуговування заявок з обмеженим часом очікування [5], а також у системах з кінцевим часом старіння заявок [7].

Одна з обставин, що стримує широке застосування динамічних пріоритетів є відсутність методики визначення коефіцієнтів $b_i, i = \overline{1, n}$ у зазначеній вище формулі функції пріоритетності $J_i(t)$.

Разом з тим, саме за рахунок належного вибору цих коефіцієнтів b_i можна підвищити ефективність функціонування системи щодо обраного критерію якості.

Пропонується метод рішення завдання вибору оптимальних значень зазначених коефіцієнтів, що мінімізують сумарне значення довжини черги в системах із двома типами заявок.

ОПИС МАТЕМАТИЧНОЇ МОДЕЛІ

Розглядається вузол розподіленої комп'ютерної мережі, у якій для обслуговування надходять пуассонівські потоки заявок, при цьому інтенсивність i -го потоку дорівнює $\lambda_i, i = 1, n$. Передбачається, що час обслуговування заявки i -го типу має експоненціальний розподіл із середнім значенням $\mu_i^{-1}, i = 1, n$.

У мережі не обслужені заявки утворюють необмежену чергу в буферній пам'яті. Пріоритети обслуговування заявок змінюються залежно від тривалості їхнього очікування в черзі, тобто вибір типу заявки для обслуговування здійснюється з урахуванням поточного значення функції пріоритетності $J_i(t) = b_i \tau_{q_i}(t)$, що залежить для кожної заявки від часу її очікування в черзі. При цьому обслуговування заявок здійснюється в порядку відносного пріоритету, тобто в момент звільнення каналу на обслуговування надходить заявка, що володіє в цей момент часу максимальним значенням функції пріоритетності $J_i(t)$ із числа заявок, що перебувають у буферній пам'яті.

Тут досліджується два випадки - відносні пріоритети при лінійно зростаючій і лінійно спадаючій функції пріоритетності заявок у розподіленій мережі.

1. Динамічні пріоритети при лінійно зростаючій функції пріоритетності заявок.

Такі пріоритети особливо актуальні в системах з обмеженим часом очікування заявок у черзі. У цьому випадку у виразі функції пріоритетності $J_i(t)$ коефіцієнти b_i , що визначають швидкості зміни пріоритетності різнотипних заявок при їхньому перебуванні в буферній пам'яті мережі, є позитивними величинами, $b_i > 0, i = 1, n$. При використанні пріоритетів даного типу, якщо в деякий момент T_i надійшла заявка i -го типу, а в момент T_k заявка k -го типу, $T_k > T_i$, і при цьому виконується нерівність $b_k > b_i$, то відповідно до розглянутої дисципліни динамічного призначення пріоритетів до моменту $T_{ik} = (b_k T_k - b_i T_i) / (b_k - b_i)$ більш високим пріоритетом володіє заявка i -го типу, а при $t > T_{ik}$ заявка k -го типу, незважаючи на те, що вона перебуває в буферній пам'яті мережі менше часу, чим заявка i -го типу. Це означає, що якщо комп'ютер мережі звільниться від обслуговування деякої заявки до моменту T_{ik} , то на обслуговування буде обрана заявка i -го типу; у протилежному випадку, на обслуговування першої надійде заявка k -го типу. Таким чином, при даній дисципліні обслуговування навіть заявки з малим коефіцієнтом b_i при досить великих значеннях часу очікування в черзі можуть одержати перевагу на обслуговування перед всіма іншими типами заявок.

У стаціонарному режимі (тобто коли $\rho = \lambda / \mu < 1$) середній час очікування в черзі заявок i -го типу визначається так [8]:

$$\tau_{q_i} = \frac{\tau_{q_0} / (1 - \rho) - \sum_{j=i+1}^n \rho_j \tau_{q_j} (1 - b_j / b_i)}{1 - \sum_{j=1}^n \rho_j (1 - b_i / b_j)} \quad (1)$$

де $\tau_{q_0} = \frac{1}{2} \sum_{i=1}^n \lambda_i T_i^2 (1 + \nu_i^2)$ - середній час очікування завершення початого обслуговування

однієї заявки, \mathcal{G}_i - коефіцієнт варіації часу обслуговування заявок i -го пріоритету, $\rho_i = \lambda_i / \mu_i$ - завантаження мережі обслуговуванням заявок i -го пріоритету, $\rho = \lambda / \mu$ - загальне завантаження мережі всіма потоками заявок.

З формули (1) видно, що залежність τ_{q_i} від b_i визначається тільки відношеннями цих параметрів, а не їхніми власними значеннями.

2. Динамічні пріоритети при лінійно спадаючій функції пріоритетності заявок.

У цьому випадку функція пріоритетності заявок лінійно спадає при їх очікуванні в черзі, тобто коефіцієнти, що визначають швидкість спадання пріоритетності заявок, є негативними величинами, $b_i < 0, i = \overline{1, n}$.

Відзначимо, що така ситуація характерна, зокрема, для мереж зі швидко старіючою інформацією, цінність якої для функціонування мережі знижується пропорційно часу затримки перед обслуговуванням.

У цьому випадку також вибір заявки із числа, що очікує у черзі здійснюється таким чином, щоб на обслуговування надійшла заявка, для якої в момент звільнення комп'ютера функція пріоритетності мала максимальне значення.

У цьому випадку час очікування в черзі заявок i -го типу визначається так [8]:

$$\tau_{q_i} = \frac{\tau_{q_0} + \sum_{j=1}^{i-1} \rho_j \tau_{q_j} \left(1 - \frac{b_j}{b_i}\right)}{1 - \sum_{j=1}^i \rho_j - \sum_{j=i+1}^n \rho_j \frac{b_i}{b_j}} \quad (2)$$

Так само як і в попередньому випадку, значення τ_{q_i} залежать лише від відношення параметрів b_i .

АНАЛІЗ ПАРАМЕТРІВ ДИНАМІЧНИХ ПРІОРИТЕТІВ.

Розглядається завдання вибору оптимальних значень коефіцієнтів b_i для випадку двох типів заявок. При цьому як критерій оптимальності вибирається сумарна довжина черги різнотипних заявок, іншими словами потрібно знайти такі значення, $b_i^*, i = \overline{1, 2}$ щоб сумарна довжина черги була мінімальною при заданих обмеженнях на час очікування заявок кожного типу. Математично це завдання записується так:

$$L_q = \lambda_1 \tau_{q_1} + \lambda_2 \tau_{q_2} \rightarrow \min \quad (3)$$

при обмеженнях

$$\tau_{q_1} \leq \tau_{q_1}^{\max} \quad (4)$$

$$\tau_{q_2} \leq \tau_{q_2}^{\max} \quad (5)$$

де $\tau_{q_1}^{\max}$ і $\tau_{q_2}^{\max}$ є відомі величини.

Для першого випадку завдання (3)-(5) у явному виді записується так:

$$L_q = \frac{\tau_{q_0}}{1 - \rho} \cdot \frac{(\lambda_1 + \lambda_2)b_1 + (\lambda_1 - \rho\lambda_1)b_2}{(1 - \rho_1)b_1 + \rho_1 b_2} \rightarrow \min \quad (6)$$

при обмеженнях

$$\frac{\tau_{q_0}}{1 - \rho} \cdot \frac{(\lambda_1 - \rho\lambda_1)b_1 + \lambda_1 b_2}{(1 - \rho_1)b_1 + \rho b_2} \leq \tau_{q_1}^{\max} \quad (7)$$

$$\frac{\tau_{q_0}}{1-\rho} \cdot \frac{\lambda_2 b_1}{(1-\rho_1)b_1 + \rho_1 b_2} \leq \tau_{q_2}^- \quad (8)$$

Для другого випадку завдання (3)-(5) має такий вигляд:

$$L_q = \frac{\lambda_1 \tau_{q_0} b_2 (1-\rho) + \lambda_2 \tau_{q_0} (b_2 - (\rho_2 + \rho_1) b_1)}{(b_2 - b_2 \rho_1 - \rho_2 b_1)(1-\rho)} \rightarrow \min \quad (9)$$

при обмеженнях

$$\frac{\tau_{q_0} b_2}{b_2(1-\rho_1) - \rho_2 b_1} \leq \tau_{q_1}^- \quad (10)$$

$$\frac{\tau_{q_0} (b_2 - \rho_2 b_1 - \rho_1 b_1)}{(1-\rho)(b_2 - b_2 \rho_1 - \rho_2 b_1)} \leq \tau_{q_2}^- \quad (11)$$

Завдання (6)-(8) і (9)-(11) є завданнями дрібно-лінійного програмування (ДЛП). З використанням методів ДЛП ці завдання можуть бути зведені до основного завдання лінійного програмування (ОЗЛП).

Ці завдання розв'язані із застосуванням методу штучного базису. На основі наступних вихідних даних $\lambda_1 = (1/2100, 1/2250, 1/2300, 1/2500, 1/2600)$, $\lambda_2 = (1/2100, 1/2550, 1/4750, 1/11500, 1/11600)$ для оптимізаційних завдань отримані чисельні результати.

Аналіз результатів численних експериментів для першого випадку дозволяє зробити наступні висновки:

- з ростом $\tau_{q_2}^*$ зменшується значення співвідношення b_2^*/b_1^* ;
- з ростом $\tau_{q_1}^*$ збільшується значення співвідношення b_2^*/b_1^* ;
- з ростом загального навантаження зменшується значення співвідношення b_2^*/b_1^* .

Аналіз відповідних результатів для другого випадку дозволяє зробити наступні висновки:

- з ростом τ_2^* зменшується значення співвідношення b_1^*/b_2^* ;
- з ростом $\tau_{q_1}^*$ збільшується значення b_1^*/b_2^* ;
- з ростом загального навантаження зменшується значення співвідношення b_1^*/b_2^* .

Слід зазначити, що швидкості зміни значень величини b_1^*/b_2^* істотно відрізняються друг від друга при зростаючій і спадаючій функції пріоритетності при зміні $\tau_{q_1}^*$ й $\tau_{q_2}^*$. Разом з тим, для обох функцій пріоритетності швидкість зміни співвідношення b_1^*/b_2^* є майже однаковою.

ВИСНОВКИ

Запропоновано обчислювальні процедури для знаходження оптимальних значень параметрів динамічних пріоритетів, що залежать від часу очікування заявок у черзі. Отримані результати можуть бути застосовані у високошвидкісних мережах з різнотипними повідомленнями. Розв'язок зазначених завдань у випадку двох типів заявок зводиться до розв'язку певних завдань дрібно-лінійного програмування. У теперішній час проводяться дослідження з узагальнення розроблених процедур для розв'язку розглянутих проблем для випадку довільного числа типів заявок.

СПИСОК ЛІТЕРАТУРИ

1. Pitts J.M. Introduction to ATM design and performance / J.M.Pitts, J.A.Schormans - N.Y.: John Wiley & Sons, 1997. – P.206 – ISBN 0471963402
2. Chao H.J. Queue management with multiple delay and loss priorities for ATM switches / H.J.Chao, I.H. Peckan // International Conference on Communication, 1-5, May, 1994: Conference Proceedings. - N.Y.: IEEE Communication Society, 1994. - Vol.2 - PP.1184-1189.- ISBN 0-7803-1825
3. Lee Y. Queueing system with multiple delay and loss priorities for ATM networks / Y.Lee, B.D.Choi // Information systems. – Elsevier, 2001.- vol.138. - PP.7-29. - ISSN 0306-4379

4. Kim C.S. Approximation method for performance analysis of queuing system with multimedia traffics / C.S.Kim, A.Z.Melikov, L.A.Ponomarenko // Applied and Comput. Math. - Baku: Academy of Sciences of Azerbaijan, 2007. - vol.6,no.2. - PP.218-226. - ISSN 1683-3511
5. Пономаренко Л.А. Оптимизация структуры сетей передачи данных / Л.А.Пономаренко, А.З.Меликов, Б.Г.Исмаилов // Проблемы информатизации и управления. – Киев: издательство КМУГА,1997 г. - Выпуск 1. - С.155-160. - ISBN 5-7763-4415-8
6. Меликов А.З. Оптимизация иерархических структур сетей передачи данных / А.З.Меликов, Б.Г.Исмаилов // Известия НАН Азербайджана. Серия : физико-технических и математических наук. - Баку : Издательство «ЭЛМ»,1997. - №6. - С.69-73. - ISSN 002-3108
7. Исмаилов Б.Г Проектирование распределенной сети обслуживания объектов добычи и подготовки газа / Б.Г Ис-майлов // М.: Научно-технический сборник, 2001. - №1. - С.29-35. - ISBN 01886
8. Липаев В.В. Математическое обеспечение управляющих ЦВМ / В.В.Липаев, К.К.Колин, Л.А.Серебровский - М.: Советское радио, 1972. - 528 с. - ISBN 3.3-14/111-72

Надійшла до редакції 12.01.2009р.

МЕЛІКОВ АГАСІ ЗАРБАЛІ – д.т.н., проф., завідувач лабораторій інститут кібернетики НАН Азербайджану, тел: 4470640, *Email: agassi@science.az.*

ІСМАЙЛОВ БАЛАМІ ГАСИМ – к.т.н., доцент кафедри “Інформатики”, Сумгаїтський Державний Університет, м.Сумгаїт , Азербайджан, *Email: Balemi@rambler.ru*