

УДК 681.3

Б.Р. Дунець

Національний університет “Львівська політехніка”,
кафедра “Електронно-обчислювальні машини”**ШВИДКИЙ ПАРАЛЕЛЬНИЙ МЕТОД ПЛАНУВАННЯ КОМУТАЦІЄЮ**

© Дунець Б.Р., 2003

Запропоновано швидкий паралельний метод планування комутацією. Запропонований метод дозволяє побудувати комутатори пакетів матричного типу з вхідною буферизацією, які при меншій часовій складності, ніж у відомих, забезпечують максимальну пропускну здатність та мінімальний середній час затримки комутації пакетів при опрацюванні пакетів на швидкості надходження вхідних даних.

In the paper fast parallel scheduling method are proposed. The proposed fast parallel scheduling method allows creating input buffered crossbar switch fabrics, which provide maximum throughput on switch fabric rate equal input data rate. The proposed method with lowest computation complexity as others provides maximum switch fabric throughput with minimum cell delay.

Вступ

У сучасних комунікаційних системах та мережах, побудованих за принципом передачі пакетів фіксованої довжини (ATM, BISDN, Ethernet), широко використовуються комутатори [1]. Сучасні комутатори є складними пристроями, на які покладено задачі встановлення, підтримки та знищення зв'язку, а також керування потоком даних. Основним елементом комутатора, який виконує фізичну комутацію пакетів даних, є комутуюче середовище (КМС) [2]. Оскільки пакети даних, що надходять у комутуюче середовище, не є сплановані, то виникає ситуація, при якій декілька пакетів вхідних даних одночасно мають бути скомутовані на один вихідний порт. Така ситуація називається вихідним змаганням (конфліктом). Подолання вихідного конфлікту та існуючого в деяких типах комутаторів внутрішнього блокування є основною задачею, що вирішується різними методами побудови комутаторів із застосуванням ефективних алгоритмів розв'язання конфліктів [2]. Від ефективності вибраного методу побудови комутаторів та алгоритму розв'язання конфліктів залежать характеристики КМС та комутатора загалом.

За останні роки було запропоновано багато структур комутуючих середовищ, які поділяють на два класи: КМС з часовим характером комутації та КМС з просторовим характером комутації [2]. Крім того, КМС з часовим типом комутації можна поділити на КМС зі спільною пам'яттю та спільним середовищем, а КМС з просторовим розподілом – на три категорії: КМС матричного типу, бан'яного типу та КМС з N^2 розділеними зв'язками. Залежно від способу організації КМС необхідною є наявність певного типу буферної пам'яті для забезпечення нормальної його роботи. Буферна пам'ять може бути розміщена на вході – вхідна буферизація, на виході – вихідна буферизація, в КМС – розподілена чи суміщена. Необхідно зазначити, що, з одного боку, тип КМС однозначно визначає множину типів буферної пам'яті, які є ефективними поєднано з цим комутуючим середовищем, а з іншого – тип буферної пам'яті задає можливі типи КМС.

У статті при розгляді КМС та алгоритмів планування комутацією їх функціонування розглядається в дискретному часі – часових проміжках однакової тривалості. Прийом та передача пакета триває один часовий проміжок. Якщо КМС опрацьовує пакети на швидкості надходження вхідних даних, то КМС в один часовий проміжок виконує комутацію пакетів, кількість яких не перевищує кількості вихідних портів.

Основними характеристиками комутаторів та комутуючих середовищ, які використовуються при їх порівнянні, є пропускна здатність та середній час затримки пакета. *Пропускна здатність* (Throughput) – це співвідношення кількості прийнятих до кількості переданих пакетів. Пропускна здатність у цьому випадку може розглядатись як імовірність того, що пакет буде передано комутатором. *Середній час затримки* (Mean delay) – це кількість часових проміжків між моментом отримання пакета та моментом передачі. Числові значення пропускної здатності та середній час затримки залежать від характеру надходження вхідних даних.

У роботі [9] доведено, що мінімальний середній час затримки досягається КМС з вихідною буферизацією. Відомі алгоритми (MUCFA, CCF, JPM) [2] планування комутацією для КМС з віртуальною вхідною буферизацією, які забезпечують середній час затримки передачі пакетів, дорівнює середньому часу затримки КМС з вихідною буферизацією. Проте КМС повинні працювати на швидкості, більшій за швидкість надходження даних, і тому в цій статті вони не розглядаються.

Комутуючі середовища матричного типу з вхідною буферизацією є перспективним напрямком розвитку КМС, оскільки забезпечують максимальну пропускну здатність при опрацюванні пакетів на швидкості надходження вхідних даних [11]. Побудова комутаторів та комутуючих середовищ з вхідною буферизацією нашою вихідною проблемою обмеження пропускної здатності при використанні вхідної буферизації типу FIFO. В роботі [9] показано, що при використанні вхідної буферизації типу FIFO (рис.1а) пропускна здатність КМС є обмеженою на рівні 0,586, оскільки виникає блокування черги пакетом в заголовку (HOL Blocking). При виникненні ситуації, наведеної на рис.1а, є можливим, з теоретичної точки зору, виконати комутацію другого пакета першого вхідного каналу на перший вихідний порт. Проте доступ до цього пакета заблоковано першим пакетом першого вхідного порта, який очікує на комутацію у другий вихідний порт.

Усунення проблеми блокування вхідної черги пакетом в заголовку та досягнення максимальної пропускної здатності КМС досягається застосуванням вхідної буферизації, відмінної від FIFO, та ефективних алгоритмів планування комутацією.

Існує два методи побудови КМС з вхідною буферизацією, які дозволяють усунути блокування вхідної черги пакетом в заголовку, – це віртуальна вихідна буферизація (VOQ) на основі FIFO та Буферизація, відмінна від FIFO.

У першому випадку (рис.1б) використовується організація буферної пам'яті, при якій вхідні дані, що надходять вхідним каналом, розміщуються у відповідному буфері FIFO, номер якого відповідає номеру вихідного каналу. Проблема планування комутацією полягає в оптимальному встановленні зв'язку за допомогою комутуючої мережі між вибраними вхідними буферами та вихідними каналами. Існує ціла гама алгоритмів планування комутацією для КМС з віртуальною вихідною буферизацією, основними з яких є iSLIP, SLIP [3], LQF, LPF, iLQF, iLPF [4], MSM [5], MWM [6], MUCFA [7], LAURA[8]. Алгоритмічна складність кращих алгоритмів є доволі високою $O(N^2)$, що є перешкодою на шляху створення високопродуктивних комутаторів.

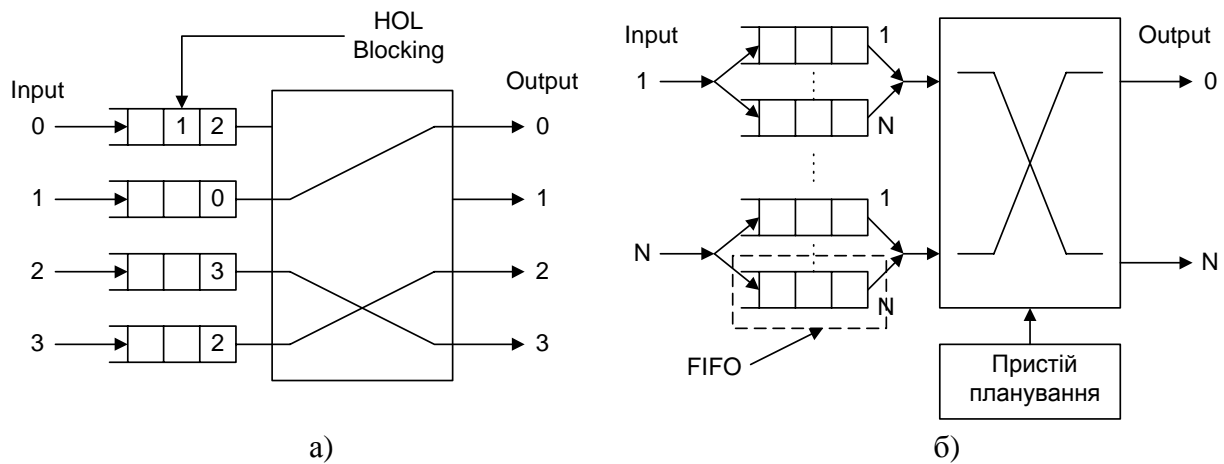


Рис. 1. КМС з вхідною буферизацією

У другому випадку вхідна буферизація організується так, що комутуюче середовище може одночасно опрацьовувати W пакетів кожного вхідного буфера. Величину W називають вікном обслуговування запитів. Відомо [9], що при зростанні величини вікна обслуговування W зростає відповідно і пропускна здатність КМС (табл. 1). Питання планування комутацією в даному випадку широко не досліджувалось, а в роботі [10] запропоновано лише використовувати нейронні мережі для планування комутації. Цей підхід не набув подальшого розвитку, оскільки використання нейронних мереж не забезпечує максимальної пропускної здатності КМС.

Таблиця 1

Залежність пропускної здатності від розміру вікна обслуговування та кількості портів комутатора

К-ть вхідних портів	Розмір вікна, W							
	1	2	3	4	5	6	7	8
2	0,75	0,84	0,89	0,92	0,93	0,94	0,95	0,96
4	0,66	0,76	0,81	0,85	0,87	0,89	0,91	0,92
8	0,62	0,72	0,78	0,82	0,85	0,87	0,88	0,89
16	0,60	0,71	0,77	0,81	0,84	0,86	0,87	0,88
32	0,59	0,70	0,76	0,80	0,83	0,85	0,87	0,88
64	0,59	0,70	0,76	0,80	0,83	0,85	0,86	0,88
128	0,59	0,70	0,76	0,80	0,83	0,85	0,86	0,88

Отже, створення високопродуктивних КМС з вхідною буферизацією наштовхується на низьку ефективність відомих методів побудови КМС, що базуються на складних алгоритмах планування комутацією. Складність алгоритмів обумовлена методом побудови КМС, при якому існує обмеження на вибір в часовому проміжку – не більше одного пакета з одного вхідного буфера. Такий підхід хоч і застарілий, але є доцільним при побудові комутатора на дискретних інтегральних елементах, де вхідні буфери будуються на блоках

пам'яті типу ОЗП [2]. Можливості сучасної технології НВІС та підходи до проектування систем на кристалі дозволяють створювати КМС з вхідною буферизацією, відмінною від вищенаведеної [12], та вимагають ефективних для організації алгоритмів планування комутацією.

1. Швидкий паралельний метод планування комутацією

З метою розв'язання проблеми блокування вхідної черги пакетом в заголовку запропоновано структуру КМС (рис. 2) та ефективний для цієї структури швидкий паралельний алгоритм планування комутацією – FPSA. При такій організації вхідна буферизація є відмінною від FIFO, а блок планування комутацією може одночасно опрацьовувати W пакетів кожного вхідного буфера. В роботі розглядається КМС розміром $M \times N$, де M – кількість входів, N – виходів, в якому за кожен часовий проміжок t надходить тільки один пакет в кожен з вхідних портів.

Означення 1. Множина процесів $A(t)=[A_i]$ – це сукупність дискретних в часі, стаціонарних та ергодичних незалежних процесів надходження вхідних пакетів $A_i(t)$ з i -го входу на j -й вихід з інтенсивністю λ_i, j та з відповідними ідентифікаторами $h_i(t)$, які вказують на вихідний порт, на який цей пакет має бути скомутований.

Означення 2. Множина вхідних процесів $A(t) = \{A_i(t); 1 \leq i \leq M\}$ є допустимою, якщо виконується умова:

$$\sum_i \lambda_i, j < 1, \sum_j \lambda_i, j < 1, \quad (1)$$

де $i=1,2,\dots,M$ – номер входу, $j=1,2,\dots,N$ – номер виходу, і недопустимою в усіх інших випадках.

$D_j(t)$ – є процес, що описує передачу пакетів j -го виходу з інтенсивністю $\mu_{i,j}$. Процес передачі пакетів може бути не визначений, якщо процес надходження вхідних пакетів нестационарний та неергодичний.

Означення 3. Матриця ідентифікаторів $H(t)=[h_i]$ – це матриця-стовпець, елементами якої є ідентифікатори $h_i(t)$, які в кожному часовому проміжку t разом з вхідними пакетами надходять в КМС та містять для кожного входу $i=1,2,\dots,M$, номер вихідного порта $j=1,2,\dots,N$, на який цей пакет має бути скомутований. Якщо $h(t)_i=0$, то пакет порожній.

Комуруюче середовище (рис.2) складається з вхідного буфера Q , що представляється масивом буферів пакетів $Q=[q_{i,k}]$, $i=1,2,\dots,N$, $k=1,2,\dots,W$, блоку планування комутацією (БПК) та комуруючої мережі. При надходженні в момент часу t множини вхідних пакетів $A(t)$, дані пакетів зберігаються в масиві $Q(t)$, а відповідні ідентифікатори $H(t)$ надходять у БПК. Для моменту часу t (часового проміжку) БПК на основі отриманих вхідних ідентифікаторів визначає вектор комутації $S(t)=[s_j]$, $j=1,2,\dots,N$, який містить слово комутації s_j для кожного j -го вихідного порта. Слово комутації s_j задає номер пакета масиву Q , який комутується за допомогою комуруючої мережі на вихід j .

Означення 4. Матриця ідентифікаторів $V(t)=[v_{i,k}]$ – це масив, елементами якого є ідентифікатори $v_{i,k}$, де $v_{i,k}=h_i(t-k-1)$, $i=1,2,\dots,M$, $k=1,2,\dots,W$.

Означення 5. Матриця вибору $G(t)=[g_{j,k}]$ – це масив, елементами якого є числові значення номеру стовпця матриці ідентифікаторів.

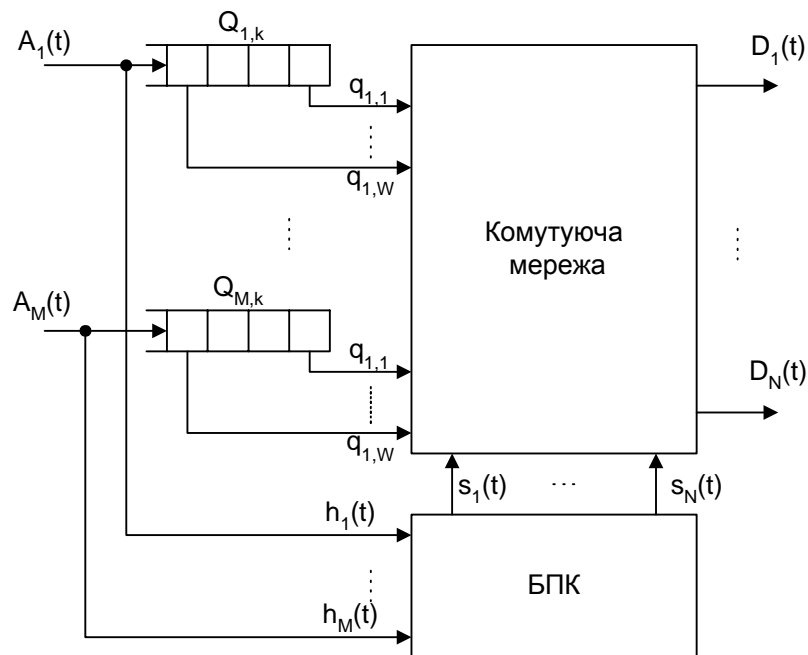


Рис. 2 Структура КМС з вхідною буферизацією, відмінною від FIFO

Для подальшого спрощення опису алгоритму планування комутацією введемо функцію вертикального вибору $V_SEL(V(t), j, k)$ та функцію горизонтального вибору $H_SEL(G(t), j)$, які виконують пошук та вибір елементів матриць $V(t)$ та $G(t)$ згідно з певними критеріями.

Функція вертикального вибору $V_SEL(V(t), j, k)$ виконує пошук елемента, значення якого дорівнює j серед елементів k -го стовпця матриці $V(t)$. Функція повертає номер рядка елемента чи 0 у випадку відсутності в рядку елемента, значення якого дорівнює j . У випадку наявності декількох елементів в k -му стовпці матриці $V(t)$ функція повертає найбільший номер рядка.

Функція горизонтального вибору $H_SEL(G(t), j)$ виконує пошук серед елементів j -го рядка матриці $G(t)$ значення ненульового елемента. Якщо такий елемент $g_{j,k}$ знайдено, то функція повертає пару чисел $\{i, k\}$, де $i = g_{j,k}$ та k – номер стовпця знайденого елемента. За наявності декількох ненульових елементів в j -му рядку матриці $G(t)$, функцією вибирається елемент з найбільшим номером стовпця. За відсутності в j -му рядку ненульових елементів функція повертає $\{0, 0\}$.

Швидкий паралельний алгоритм планування комутацією в кожному часовому проміжку t виконує планування комутацією за такою послідовністю дій:

Крок 1. Вертикальний пошук

Обчислення за допомогою функції вертикального вибору значень матриці вибору $g_{i,k} = V_SEL(V(t), j, k)$, для кожної пари j та k , де $j = 1, 2, \dots, N$ та $k = 1, 2, \dots, W$.

Крок 2. Горизонтальний пошук

Обчислення за допомогою функції горизонтального вибору значень вектора комутації $s_j = H_SEL(G(t), j)$, для кожного $j = 1, 2, \dots, N$.

Крок 3. Підтвердження та комутація

Для кожного вихідного порта $j=1,2,\dots,N$ виконується комутація за вектором комутації та встановлення в 0 значення відповідних комірок матриці комутації, що відповідають скомутованим пакетам на цьому кроці:

$$Dj(t)=v_{i,k}, v_{i,k}=0, \text{ де } \{i,k\}=s_i$$

Властивість. Запропонований алгоритм комутацією має властивість *паралельності*. Оскільки елементи матриці вибору $G(t)$ є між собою непов'язані, то можна обчислювати одночасно всі елементи матриці $G(t)$ за допомогою $N \cdot W$ паралельних викликів функції вертикального вибору. Аналогічно, за допомогою N паралельних викликів функції горизонтального вибору можна одночасно визначити всі елементи вектора комутації $S(t)$.

Приклад роботи алгоритму для $N=2, M=4, W=8$:

Нехай в момент часу t маємо таку матрицю ідентифікаторів:

$$V(t) = \begin{bmatrix} 1 & 2 & 2 & 0 & 0 & 1 & 0 & 1 \\ 1 & 2 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \end{bmatrix},$$

Крок 1. Під час горизонтального пошуку за допомогою 16-ти викликів функцій горизонтального вибору отримуємо таку матрицю вибору:

$$G(t) = \begin{bmatrix} 2 & 0 & 4 & 0 & 0 & 3 & 0 & 1 \\ 0 & 2 & 1 & 0 & 3 & 0 & 4 & 0 \end{bmatrix}.$$

Крок 2. Під час вертикального пошуку за допомогою 2-х викликів функцій вертикального вибору отримаємо такий вектор комутації:

$$S(t) = [\{1,8\} \quad \{4,7\}].$$

Крок 3. Виконання комутації на перший вихід пакет $q_{1,8}$, на другий – $q_{4,7}$. Встановлення в 0 відповідних комірок матриці ідентифікаторів: $v_{1,8}=0, v_{4,7}=0$.

2. Ефективність швидкого паралельного алгоритму планування комутацією

З метою порівняння запропонованого методу планування комутацією, основною складовою якого є швидкий паралельний алгоритм планування комутацією, скористаємося загальноприйнятими критеріями оцінки КМС, а саме: часовою та асимптотичною часовою складнощами алгоритмів планування комутацією, залежностями пропускної здатності від величини вікна обслуговування та середнього часу затримки від інтенсивності надходження вхідних пакетів.

Часова складність є основною характеристикою, що дозволяє оцінити час виконання алгоритму планування комутацією. Часова складність запропонованого алгоритму визначається як $L = N \cdot W \cdot L_{V_SEL} + N \cdot L_{H_SEL}$, але оскільки з властивості алгоритму випливає, що значення $g_{i,k}$ та s_j можуть обчислюватись паралельно, тоді $L = L_{V_SEL} + L_{H_SEL}$. Часові складності функцій горизонтального та вертикального вибору залежать від обраного алгоритму пошуку, при використанні паралельного алгоритму часові складності функцій

дорівнюють: $L_{V_SEL} = \log_2 M$, $L_{H_SEL} = \log_2 W$, тоді $L = \log_2 M + \log_2 W$. Асимптотична складність алгоритму при цьому дорівнює $O(\log_2 N)$. Порівняння часових складностей кращих [11] алгоритмів планування комутацією (табл. 2) та запропонованого швидкого паралельного алгоритму (FPSA) показує, що запропонований алгоритм має найменшу асимптотичну складність.

Пропускна здатність КМС з швидким паралельним методом планування комутацією залежить від величини вікна W . При збільшенні значення вікна пропускна здатність зростає та прямує до 1. На рис. 3 наведено графіки залежності пропускної здатності (Throughput) від розміру вікна для КМС з $N=M=4,8,16,32$. Залежності отримано поведінковим моделюванням КМС з швидким паралельним алгоритмом планування комутацією для допустимої множини вхідних процесів $A(t)$. Кожен з процесів надходження вхідних даних відповідає процесу Бернуллі з рівномірним розподілом вхідних запитів на вихідні порти $\lambda_i, j = \frac{1}{M}$.

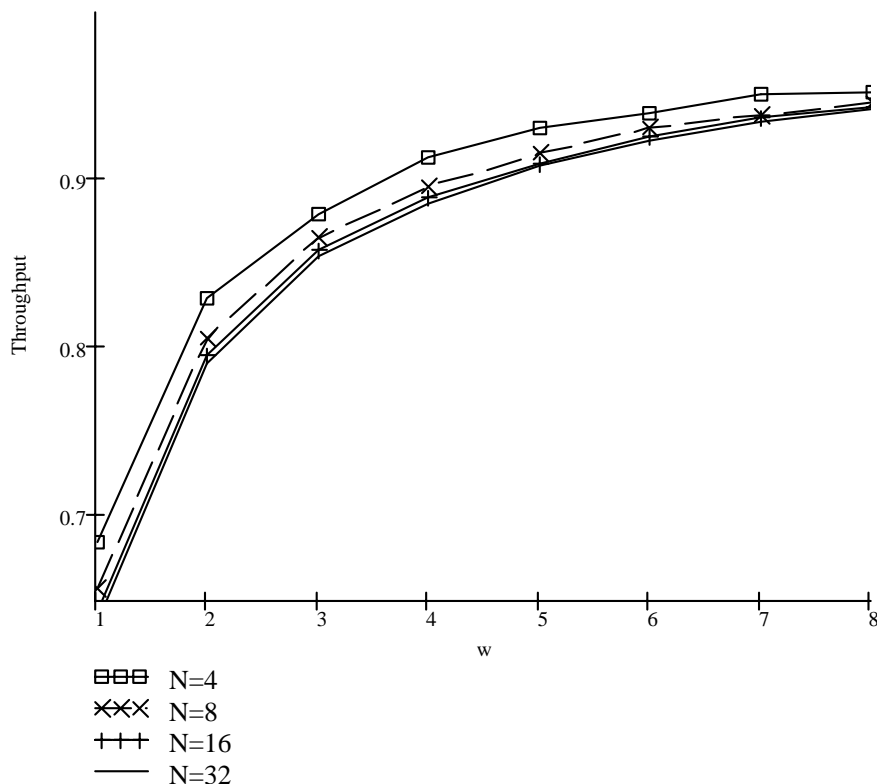


Рис. 3. Графік залежності пропускної здатності КМС від розміру вікна

На рис. 4 наведено графіки залежності пропускної здатності (Throughput) КМС, в яких використано відомі алгоритми планування комутацією, що використовують вікно обслуговування (McCulloch-Pitts, Hopfield Memory, Simple Heuristic) та КМС з використанням запропонованого (FPSA) від розміру вікна обслуговування W ($N=M=16$). Як і в попередньому випадку, результати симуляції отримано для характеру надходження вхідних пакетів, що описуються процесами Бернуллі з рівномірним розподілом вхідних запитів. Результати симуляції показують, що запропонований алгоритм порівняно з іншими відомими алгоритмами планування комутації досягає більшої пропускної здатності.

Складність алгоритмів планування комутацією

Алгоритм	К-ть додавань та віднімань	К-ть множень та ділень	К-ть порівнянь	Асимптотична складність
ISLIP	-	-	$3N^2 \log_2 N$	$O(N^2 \log_2 N)$
ILQF	-	-	$3N^2 \log_2 N$	$O(N^2 \log_2 N)$
IOCF	-	-	$3N^2 \log_2 N$	$O(N^2 \log_2 N)$
ILPF	$2N^2$	-	$N^2 + 2N \log_2 N$	$O(N^2)$
2DDR	-	-	N^2	$O(N^2)$
RC	-	-	N^2	$O(N^2)$
MUCS	$\approx N^3$	$\approx \frac{2}{3} N^3$	$\approx \frac{1}{3} N^3$	$O(N^2)$
RPA	-	-	$2N^2$	$O(N^2)$
IZIP	N^2	-	$2N^2$	$O(N^2)$
FPSA	-	-	$\log_2 M + \log_2 W$	$O(\log_2 N)$

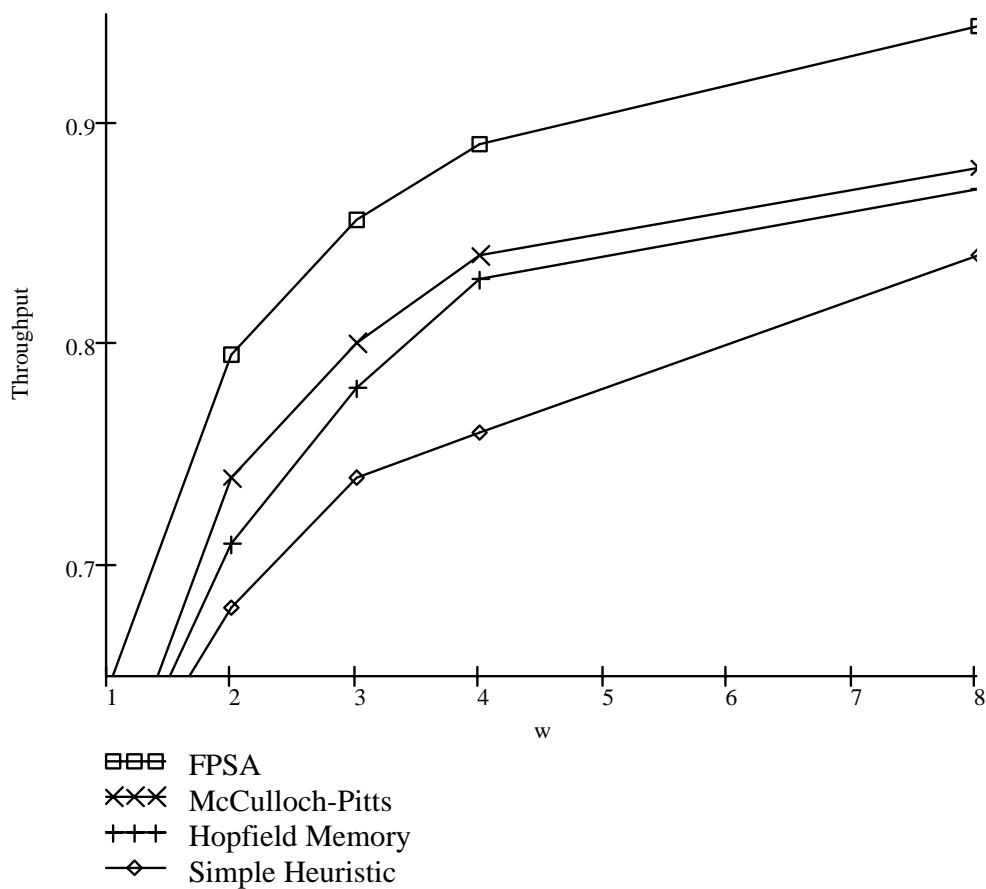


Рис. 4. Порівняння пропускної здатності відомих та запропонованого КМС

Одним з важливих критеріїв порівняння КМС є середній час затримки пакетів при комутації. Ця характеристика дозволяє оцінити середню затримку передачі даних в інформаційній мережі, де використовується КМС. На рис.5 наведено графіки залежності середнього часу затримки (Mean delay time) від інтенсивності надходження вхідних пакетів (*load*). Залежності отримано шляхом поведінкового моделювання КМС для допустимої множини вхідних процесів $A(t)$. Кожен з процесів надходження вхідних даних відповідає процесу Бернуллі з рівномірним розподілом вхідних запитів на вихідні порти з врахуванням інтенсивності: $\lambda_i, j = \frac{1}{M} load$.

Як вже зазначалось, мінімальний середній час затримки досягається в КМС з вихідною буферизацією. Як бачимо з результатів симуляції, середній час затримки КМС, в якому використано запропонований метод (крива FPSA), є ідентичний середньому часу затримки КМС з вихідною буферизацією (крива Output Queue). Для порівняння на рис. 5 наведено графіки залежностей для КМС з вхідною буферизацією типу FIFO (крива Input Queue), КМС з віртуальною вихідною буферизацією з використанням алгоритмів планування комутацією MWM (крива MWM) та КМС з вхідною буферизацією з використанням віконної технології та алгоритму планування комутацією на основі нейронних мереж (крива Neural).

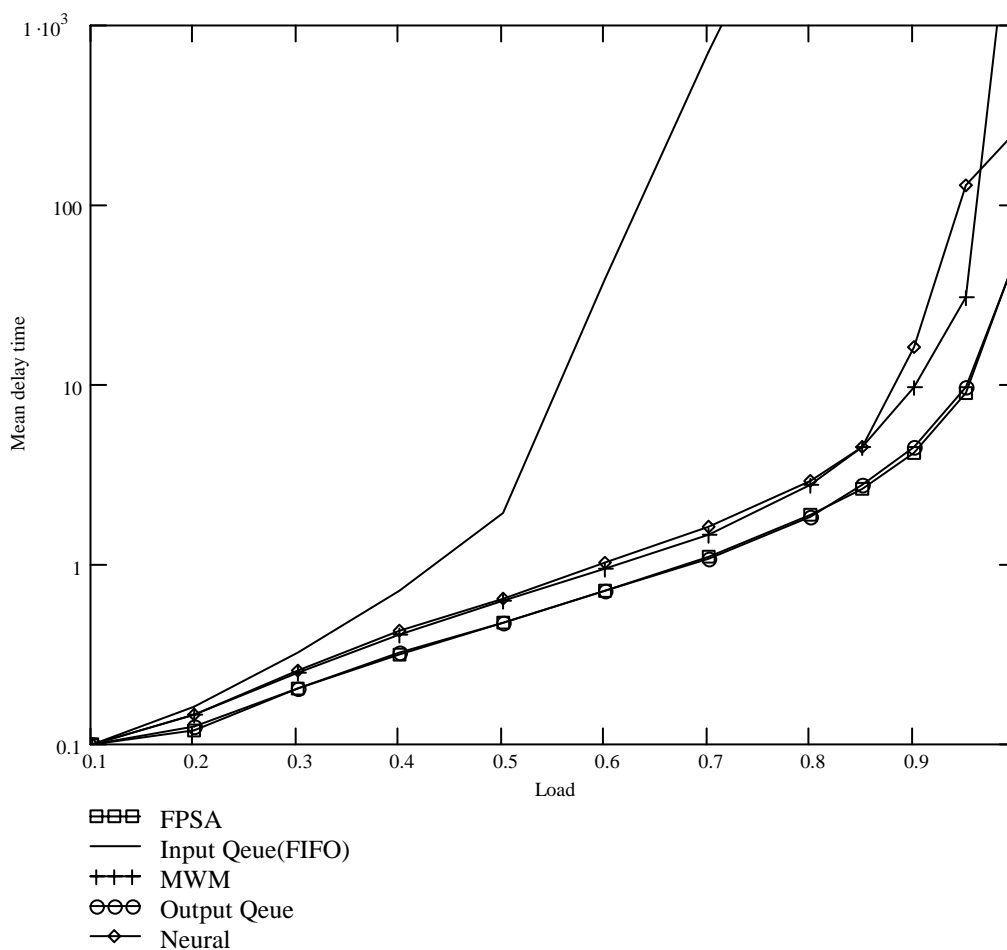


Рис. 5. Порівняння характеристик КМС, в яких використані різні алгоритми планування комутацією

Висновки

1. У роботі запропоновано новий швидкий паралельний метод планування комутацією для комутуючих середовищ з вхідною буферизацією, який порівняно з відомими використовує буферну пам'ять, відмінну від FIFO.

2. Запропонований алгоритм планування комутацією порівняно з відомими має найменшу асимптотичну часову складність.

3. Порівняно з відомими підходами до побудови КМС з плануванням комутацією на основі вікна обслуговування запропонований метод дозволяє побудувати КМС з більшою пропускнуою здатністю.

4. Результати симуляції показують, що середній час затримки передачі пакетів в КМС, побудованому на основі запропонованого методу, є на рівні оптимального.

1. Кульги М. *Технология корпоративных сетей: Энциклопедия*. – СПб.: Издательство “Питер”, 2000. – 704с. 2. H. Jonathan Chao, Cheuk H. Lam, Eiji Oki *Broadband Packet Switching Technologies: A Practical Guide to ATM Switches and IP Routers*. – New York.: John Wiley & Sons, 2001. – 459p. 3. Nick McKeown, Martin Izzard, Adisak Mekkittikul, Bill Ellersick and Mark Horowitz *The Tiny Tera: A Packet Switch Core // IEEE Micro Jan/Feb.* – 1997. – P. 26-33. 4. Adisak Mekkittikul, Nick McKeown *A Practical Scheduling Algorithm to Achieve 100% Throughput in Input-Queued Switches // IEEE Infocom 98, Vol 2, April 1998, San Francisco.* - P. 792-799 5. Sundar Iyer, Nick McKeown *Maximum Size Matching and Input Queued Switches // Proceedings of the 40th Annual Allerton Conference on Communication, Control and Computing.* 6. Isaac Keslassy, Nick McKeown “*Analysis of Scheduling Algorithms That Provide 100% Throughput in Input-Queued Switches*”, *Proceedings of the 39th Annual Allerton Conference on Communication, Control, and Computing. Monticello, Illinois, October 2001.* 7. Balaji Prabhakar, Nick McKeown *On the Speedup Required for Combined Input and Output Queued Switching // Computer Systems Technical Report CSL-TR-97-738. November 1997.* 8. Devavrat Shah, Paolo Giaccone, Balaji Prabhakar *An Efficient Randomized Algorithm for Input-Queued Switch Scheduling // IEEE Hot Interconnects 9, Stanford, CA, USA, August 2001.* 9. Karol, M.; Hluchyj, M. *Queueing in high-performance packet-switching // IEEE J. Selected Area Communications, Dec. 1988, vol.6, P.1587-1597.* 10. W.M.Moh, Y.-F. Chung *Design and evaluation od cell scheduling algorithms for ATM switches // Networks: The Next Millenium, Proc. of IEEE Singapore Int. Conf. on Networks, P. 355-369, World Scientific, 1997.* 11. Marco Ajmone Marsan, Andrea Bianco, Enrica Filippi, Paolo Giaccone, Emilio Leonardi, Fabio Neri *A Comparison of Input Queuing Cell Switch Architectures // IEEE BSS'99, 3rd International Workshop on Broadband Switching Systems, Kingston, Canada, June 1999* 12. Mei Yang, S.Q. Zheng *An Efficient Scheduling Algorithm for CIOQ Switches with Space-Division Multiplexing Expansion // IEEE Infocom 2003, Vol 2, April 2003, San Francisco.* – P. 740-749