

М. Ф. Каравай, В. С. Подлазов
Mikhail Karavay, Viktor Podlazov

ПОВЫШЕНИЕ БЫСТРОДЕЙСТВИЯ МНОГОМЕРНЫХ ТОРОВ SPEED ACCELERATION OF MULTIDIMENSIONAL TOURSES

АННОТАЦИЯ. Рассматривается метод повышения быстродействия дуплексных колец многомерных торов за счет изменения их топологии, приводящее к значительному сокращению их диаметра и повышению пропускной способности.

ABSTRACT. A technique of multidimensional tourses speed increasing at the expense of change of component duplex rings topology is considered. The technique provides considerable reduction of ring diameters and increasing of ring throughputs.

Ключевые слова: высокопроизводительные многопроцессорные вычислительные системы, системные сети, многомерные торы, дуплексные кольца, мультикольца, сеть сервера GEMINI.

Keywords: massive parallel multiprocessor computer, system area networks, multidimensional t tours, duplex rings, multirings, network of server GEMINI.

Введение

Многомерные торы широко применяются для построения системных сетей современных высокопроизводительных параллельных МВС (суперкомпьютеров) [[1]–[3]]. В каждом измерении таких торов используется пара встречных симплексных колец – дуплексное кольцо.

Повышение быстродействия тора может быть достигнуто за счет сокращения времени передачи пакетов по кольцам отдельных измерений. Время передачи по отдельному кольцу складывается из времен доставки пакетов по кольцу и времени ожидания передачи в это кольцо. Время доставки по дуплексному кольцу может быть уменьшено за счет сокращения его диаметра. Время ожидания передачи в дуплексное кольцо (нахождения в очереди к нему) может быть уменьшено за счет повышения его пропускной способности.

Средняя время доставки пакета по кольцу T_0 прямо пропорциональна его диаметру D : – $T_0 = kD$. Для симплексного кольца $D=N$, а для дуплексного кольца $D = N/2$, где N – это число узлов в кольце. При промежуточной буферизации пакета в каждом узле $k = B/2$ для равномерного распределения длин маршрутов, где $B = b/v$ – длительность передачи пакета, а b – длина пакета.

Многомерные торы и применяются в первую очередь для сокращения времен доставки пакетов по системной сети и повышения ее пропускной способности за счет пространственного распараллеливания. Так в [[1]] используется 5-мерный тор, в [[2]] – 4-мерный и в [[3]] – 3-мерный, при числе узлов тора от 1,5 миллионов до десятков и сотен тысяч. Если обозначить число узлов в кольце измерения i ($1 \leq i \leq n$) как N_i , а его диаметр как D_i , то число узлов N в n -мерном торе выражается как $N = \prod_i N_i$, а его диаметр D – как $D = \sum_i D_i$.

В дуплексных кольцах пакеты удаляются из канала приемниками, а не источниками, как в кольцах с передачей жезла. Как следствие их пропускная способность w больше скорости передачи v , т.е. выражается как $w = cv$ [[4], [5]]. Коэффициент c называется емкостью кольца, а сами кольца являются кратными

Для кольца с передачей жезла $c \approx 1$. Для симплексного кратного кольца $c = 2$. Для дуплексного кратного кольца c зависит от распределения длин маршрутов по кольцу и размеров очереди пакетов для передачи в каждом узле. При передаче пакетов по кратчайшему маршруту для равномерного распределения длин маршрутов и большой очереди в несколько десятков пакетов достигается $c = 8$ при большом числе узлов [[4]]. При высокой загрузке кольца (больше $c/2$) время ожидания передачи в кольцо растет гиперболически и может оказаться больше времен доставки пакетов по кольцу. Поэтому повышение c может дать существенное сокращение общего времени передачи пакетов.

Задача повышения быстродействия системной сети является постоянно актуальной, особенно в связи с постоянным ростом числа и производительности процессоров в узлах суперкомпьютеров и растущим требованиям к характеристикам системной сети. Для многомерного тора ее можно дополнительно решать за счет сокращения диаметров колец разных измерений и повышения их пропускной способности. Последняя задача часто решается за счет увеличения числа колец в каждом измерении

[[3]]. В данной работе обе эти задачи решаются за счет изменения топологии пар колец каждого измерения, а именно за счет использования минимальных коммутируемых мультиколец, в которых число дуплексных портов M_i , используемых для включения узлов в кольцо i -го измерения, остается неизменным по сравнению с дуплексными кольцами. При наличии в i -ом измерении m_i дуплексных колец $M_i = 2m_i$.

1. Коммутируемые мультикольца

Мультикольцом мы называем обычное симплексное кольцо с одинаковым набором хордовых дуг разной длины, проведенных из каждого узла. Коммутируемым мультикольцом мы называем мультикольцо, в котором пакеты могут передаваться из любого узла по любой дуге. Минимальным мультикольцом мы называем мультикольцо, в котором набор хордовых дуг содержит только одну дугу.

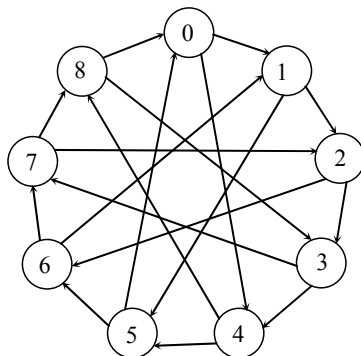


Рис. 1. Минимальное коммутируемое мультикольцо $\{9, (1, 4)\}$.

На Рис. 1 приведен пример минимального коммутируемого мультикольца с 9 узлами и с хордовой дугой длины 4. Для минимальное коммутируемое мультикольцо обозначим МКМ.

В остальной части работы мы будем использовать определение мультикольца как набора колец с разными шагами [[4], [5]]. Формально это можно выразить следующим образом.

Предположим, что узлы перенумерованы целыми числами из $[0, N - 1]$, где N – число узлов. Пусть мультикольцо состоит из $m \geq 1$ колец. В j -ом кольце номера узлов образуют последовательность ${}^jX_{i+1} = ({}^jX_i + {}^jS) \bmod N$, где $X_i \in [0, N - 1]$ ($i = 0, 1, \dots$) и ${}^jS > 0$ – шаг j -

ого кольца ($1 \leq j \leq m$). Мультикольцо задается набором шагов $S_m = ({}^1S = 1, {}^2S, \dots, {}^mS)$, где ${}^1S < {}^2S < \dots < {}^mS$ и обозначается как $\{N, S_m\}$.

Кольцо с положительным шагом ${}^iS \geq N/2$ будем также называть встречным кольцом с отрицательным шагом ${}^jS = -(N - S)$. Пример мультикольца на 9 узлов из 2-х колец с набором шагов $S_2 = (1, 4)$ приводится на Рис. 1. При обратном направлении хордовых дуг это было бы мультикольцо $S_2 = (1, 5) = (1, -4)$. В дальнейшем кольцо с шагом jS будем называть кольцом jS , а его дугу – дугой jS .

Будем различать два вида колец – обычные и расщепленные. Обычное кольцо имеет взаимно простые N и jS . В расщепленном кольце N и jS имеют наибольший общий делитель d , и последовательность X_i разделяется на d непересекающихся последовательностей jX_i с периодом N/d ($0 \leq j \leq d-1, {}^jX_0 = j$), которые в совокупности содержат все номера из $[0, N-1]$. Физически расщепленное кольцо состоит из d миниколец по N/d узлов в каждом.

2. Маршрутизация и диаметр

Минимальное коммутируемое мультикольцо (МКМ) с N узлами является мультикольцом $\{N, (1, p)\}$, где $2 \leq p \leq \lfloor N/2 \rfloor$.

Рассмотрим маршрутизацию некоторого пакета по МКМ, считая $p = \lfloor N/n \rfloor$. Длиной маршрута L является разница номеров (по модулю N) приемника и источника пакета. Выразим L как $L = Rp + r$, где $R = \lfloor L/p \rfloor$ и $0 \leq R \leq n-1$, а $0 \leq r \leq p-1$. Этапом маршрута назовем переход по дуге некоторого кольца.

Пусть сначала N и p взаимно просты. Маршрут длины L состоит из R последовательных этапов по кольцу p и r последовательных этапов по кольцу 1. Маршрут может начинаться с любого кольца. Однако маршруты всех пакетов должны начинаться с одного и того же кольца, иначе возможно возникновение маршрутных тупиков из-за циклической зависимости маршрутов по дугам.

Пусть теперь N и p имеют общий делитель. В этом случае мы рассматриваем только МКМ с $N = np$, в котором кольцо p состоит из n миниколец по p узлов каждое. Здесь маршрут содержит R

последовательных этапов по миникольцу, содержащему узел-источник, если маршрут начинается с кольца p , и по миникольцу, содержащему узел-приемник, если маршрут начинается с кольца 1.

При описанной маршрутизации диаметр МКМ выражается как $D = n+p-2$, что существенно меньше чем у дуплексного кольца, в котором $D = N/2$. Несмотря на меньший диаметр МКМ является несимметричной по задержкам сетью при обменах пакетами в произвольной паре узлов. Этот недостаток можно устранить при увеличении числа портов, т.е. при увеличении числа колец.

Если число дуплексных портов в каждом измерении больше 2 возможно дальнейшее уменьшение диаметра каждого измерения. Так в [[3]] все измерения содержит не менее двух дуплексных колец, т.е. число дуплексных портов $m_i \geq 4$. Если $m_i = 4$, то можно использовать дуплексное МКМ (ДМКМ), состоящее из двух встречных МКМ, и маршрутизацию в нем осуществлять по кратчайшему пути. ДМКМ является симметричной сетью.

Рассмотрим такую маршрутизацию при условии, что n и p являются четными и $N = np$. Здесь L выражается опять как $L = Rp+r$, однако числа этапов задаются другими ограничениями – $n/2 \leq R \leq n/2$, а $p/2 \leq r \leq p/2$ и задаются по следующему алгоритму.

Если $L < N/2$, то $R = L/p$, если $L > N/2$, то $R = (L-N)/p$, иначе $R = \pm N/2$ с равной вероятностью. Сначала случай $L \leq N/2$. Если $l = N-Rp < p/2$, то $r = l$, иначе $r = l-p$ и $R = R+1$. Теперь случай $L \geq N/2$. Если $l = N+Rp < p/2$, то $r = -l$, иначе $r = p-l$ и $R = R-1$.

Если $R > 0$, то для передачи пакетов используется кольцо p прямого МКМ, в противном случае – кольцо p встречного МКМ. Если $r > 0$, то для передачи пакетов используется кольцо 1 прямого МКМ, в противном случае – кольцо 1 встречного МКМ.

При такой маршрутизации диаметр ДМКМ выражается как $D = (n+p)/2$, что почти вдвое меньше, чем у МКМ.

3. Аналитическая модель МКМ

Оценим пропускную способность МКМ в тех же условиях, при которых для дуплексного кольца достигается $s = 8$. Это случай такой перегрузки сети, при которой каждый узел имеет пакет для передачи в кольцо всякий раз когда мимо него проходит по этому кольцу пустой сегмент. В этом случае расчет пропускной способности можно осуществлять по средним длинам маршрутов [[4], [5]]. Если предположить, что каждый узел генерирует один

пакет для передачи за период равный длительности пакета, то такой режим реализуется при достаточно большой очереди пакетов для передачи [[4]].

В модели предполагается передача пакетов постоянной длины в тактовых сегментах, циркулирующих независимо по обоим кольцам МКМ. Любой пакет, маршрут которого имеет $r = 0$, переносится в сегменте(ах) только кольца p , а пакет с $R = 0$ – сегменту(ах) только кольца 1. Пакет с $r \neq 0$ и $R \neq 0$ переносится в сегменте(ах) одного кольца, а затем коммутируется в сегмент другого кольца. Среднее число пакетов, перенесенных в сегменте за один его проход по i -му кольцу, называется его емкостью c_i . Примем для МКМ, что $i=0$ для кольца 1 и $i=1$ для кольца p .

Перенос пакет по кольцу может осуществляться прямо или с приостановкой в промежуточных узлах. В первом случае пакет переносится прямо в том сегменте, в который он был передан узлом-источником или узлом, осуществляющим коммутацию в другое кольцо. В втором случае некоторый промежуточный узел, имеющий пакет для передачи, заменяет пакет в сегменте на свой и заносит чужой пакет в свою очередь пакетов на передачу.

Среднее число пакетов, перенесенных за один оборот сегмента по i -му кольцу при условии наличия пакета для передачи в каждом узле, называется его емкостью c_i . Известно [[4], [5]], что при прямом переносе пакетов $c_i = N/\bar{L}_i$, где \bar{L}_i – средняя длина (в этапах) части маршрута по i -му кольцу. Число пакетов, доставленных узлам-приемникам за один оборот сегмента по i -му кольцу, называется его эффективной емкостью \hat{c}_i . Очевидно, что $\hat{c}_i \leq c_i$, так как большинство пакетов переносится в сегментах разных колец. Эффективная емкость МКМ составляет $\hat{C} = \hat{c}_0 + \hat{c}_1$. Оценим эти емкости для равномерного распределения длин маршрутов при прямом переносе пакетов.

Нетрудно видеть, что для кольца 1 имеем

$$\bar{L}_0 = \sum_{k=1}^{p-1} k/(p-1) = p/2 \text{ и } c_0 = 2N/p. \text{ Аналогично для кольца } p \text{ имеем}$$

$$\bar{L}_1 = \sum_{k=1}^{n-1} k/(n-1) = n/2 \text{ и } c_1 = 2N/n.$$

Расчет эффективной емкости зависит от выбора первого кольца для передачи пакетов с маршрутами по двум кольцам и от соотношения n и p .

Пусть первым является кольцо 1. На передачу по нему в среднем претендует $(p-1)$ 1-кольцевых пакетов (с $r \neq 0$ и $R = 0$) и $(N-n-p+1)$ 2-кольцевых пакетов (с $r \neq 0$ и $R \neq 0$). Их доля составляет $q_0 = (p-1)/(N-n)$ и $Q_0 = (N-n-p+1)/(N-n)$ соответственно. Первые доставляются к приемниками и их среднее число выражается как $c_0 q_0 = 2N(p-1)/[(N-n)p]$ и задает эффективную емкость кольца 1 $\hat{c}_0 = c_0 q_0$. Вторые коммутируются в кольцо p и их среднее число составляет величину $c_0 Q_0 = 2N/p - 2N(p-1)/[(N-n)p]$.

На передачу по кольцу p в среднем претендует $(n-1)$ 1-кольцевых пакетов (с $r = 0$ и $R \neq 0$) и скоммутированные 2-кольцевые пакеты (с $r \neq 0$ и $R \neq 0$). Их доля составляет $q_1 = (n-1)/(N-p)$ и $Q_1 = (N-n-p+1)/(N-p)$ соответственно. Первые доставляются к приемникам и их среднее число выражается как $c_1 q_1 = 2N(n-1)/[(N-p)n]$, но составляет только часть эффективной емкости кольца p . Вторые могут быть доставлены приемникам и их среднее число выражается как $c_1 Q_1 = 2N/n - 2N(n-1)/[(N-p)n]$.

Если общее число пакетов, претендующих на передачу по кольцу p превышает c_1 , то оно всех их пропустить не может, и в поэтому $\hat{c}_1 = c_1$, $\hat{c}_0 = c_0 q_0$ и $\hat{C} = \hat{c}_0 + \hat{c}_1 = 2N/n + 2N(p-1)/[(N-n)p]$. Эти формулы выполняются при $c_0 Q_0 + c_1 q_1 \geq c_1$ или $[1/p - (p-1)](N-p)/p \geq [1/n - (n-1)](N-n)/n$. Формально при $N = np$ они упрощаются до $\hat{c}_0 = 2$, $\hat{c}_1 = 2p$, $\hat{C} = 2p + 2$ и $n \geq p$.

В противном случае кольцо p работает с недогрузкой, т.е. нарушаются условия расчета по средним. В этом случае можно указать только нижнюю границу \hat{c}_1 . Она складывается из числа $c_1 q_1$ пакетов, доставленных прямо по кольцу p , и числа пакетов $c_0 Q_0$, доставленных после коммутации из кольца 1. Поэтому $\hat{c}_1 > c_1 q_1 + c_0 Q_0$, а $\hat{C} > \hat{c}_0 + c_1 q_1 + c_0 Q_0 = 2N/p + 2N(n-1)/[(N-p)n]$. При этом сохраняется ограничение $\hat{c}_1 \leq c_1$. Формально при $N = np$ эти неравенства упрощаются до $\hat{C} > 2n + 2$.

Таким образом при $N = np$ имеем:

$$(1) \quad \hat{C} = 2p + 2 \text{ при } n \geq p,$$

$$(2) \quad \hat{C} > 2n + 2 \text{ при } n < p.$$

Пусть теперь первым является кольцо p . Аналогичные рассуждения приводят при $N = np$ к формулам:

$$(3) \quad \hat{C} = 2n + 2 \text{ при } n \leq p,$$

$$(4) \quad \hat{C} > 2p + 2 \text{ при } n > p.$$

4. Имитационная модель

Для исследования свойств МКМ в различных режимах функционирования была создана его имитационная модель. Эта модель в исходном виде имитирует прямой (без приостановок) перенос пакетов постоянной длины в сегментах той же длины при взаимно простых N и p . Сначала сегмент обходит все узлы по одному кольцу (делает полный оборот по кольцу), затем другой сегмент обходит все узлы по другому кольцу. Это повторяется многократно ($K=10^4$) раз без очистки сегментов после каждого обхода. В процессе каждого обхода подсчитывается суммарное число доставленных приемникам пакетов по каждому кольцу и по обоим кольцам. После K обходов эти числа усредняются по K и задают эффективные емкости колец и МКМ в целом.

До начала обхода колец каждый узел генерирует 1 пакет с равномерным распределением длин маршрутов. Аналогично каждый узел генерирует 1 пакет и после обхода обоих колец. В каждом узле они поступают в конец очередей пакетов для передачи в соответствующие кольца. После обхода первого кольца скоммутированные в другое кольцо пакеты поступают в конец очереди к этому кольцу в том узле, где произошла коммутация. Для выполнения условий работоспособности аналитической модели размер очереди в имитационной модели берется равным числу тактов моделирования K .

Во втором варианте в кольце p используется p разных сегментов, из которых i -ый сегмент ($0 \leq i \leq p-1$) используется начиная с i -го узла до тех пор пока номер очередного узла не превысит i . После завершения полного оборота содержимое i -го сегмента переносится в $(i+1) \bmod p$ сегмент. Этот вариант является промежуточным для перехода к основному варианту с $N = np$.

В основном (третьем) варианте кольцо p расщепляется на p миниколец с n узлами каждое, в которых используется p

независимых сегментов. При этом суммарное число доставленных приемникам пакетов (за K обходов узлов) считается для кольца 1 и каждого миникольца и для всей их совокупности в целом. Усредненное суммарное значение и задает эффективную емкость кольца p .

Рассмотрение этих вариантов потребовалось для проверки правильности имитационной модели в основном варианте. Критерием ее правильности является совпадение значений эффективной емкости МКМ, полученных расчетом по аналитической модели и посредством эксперимента на имитационной модели во всех ее вариантах. Рассмотрим полученные результаты.

Пусть начальным кольцом для 2-кольцевых маршрутов, является кольцо 1 и $N = np + 1$. В Таблица 1 представлены результаты экспериментов с первым вариантом имитационной модели при четном n . Подчеркнутые значения задают нижнюю границу в области $n < p$. Отметим совпадение значений с точностью до нескольких процентов результатов аналитической и имитационной модели в области $n \geq p$ за исключением случая $p = 2$ где точность составила 15 процентов.

Таблица 1. Эффективная емкость МКМ в 1-ом варианте модели.
Начальная передача в кольцо 1

N	n	2	4	6	8	12	16	32	48
17	Имитация	6,4	9,8		5,1				
	Аналитика	<u>6,1</u>	10,5		6,1				
25	Имитация	6,3	12,5	9,9	8	5			
	Аналитика	<u>6,1</u>	<u>10,3</u>	10,3	8,2	6,1			
33	Имитация	6,2	12,3		9,9		5		
	Аналитика	<u>6,1</u>	<u>10,2</u>		10,2		6,1		
49	Имитация	6,1	12,2	17,3	13,7	9,9	8		
	Аналитика	<u>6</u>	<u>10,2</u>	<u>14,2</u>	14,2	10,2	8,1		
65	Имитация	6,1	12,2		17,6		10	5	
	Аналитика	<u>6</u>	<u>10,1</u>		18,2		10,1	6	
97	Имитация	6,1	12,1	18,2	24,2	17,7	13,9	8	5
	Аналитика	<u>6</u>	<u>10,1</u>	<u>14,1</u>	<u>18,2</u>	18,2	14,1	8,1	6
	Аналитика	<u>6</u>	<u>10,1</u>		<u>18,1</u>		18,1	10,1	

Результаты экспериментов со вторым вариантом имитационной модели оказались очень близкими. При этом необходимо отметить более пульсирующий характер трафика с большими значениями среднеквадратичного отклонения.

Таблица 2. Эффективная емкость МКМ в основном варианте.
Начальная передача в кольцо 1

<i>N</i>	<i>n</i>	2	4	6	8	12	16	32	48
16	Имитация	5,1	10		5,1				
	Аналитика	<u>6</u>	10		6				
24	Имитация	5	11,1	10	8	5			
	Аналитика	<u>6</u>	<u>10</u>	10	8	6			
32	Имитация	5	11,1		9,9		5		
	Аналитика	<u>6</u>	<u>10</u>		10		6		
48	Имитация	5	11,1	17,1	14	10	8		
	Аналитика	<u>6</u>	<u>10</u>	<u>14</u>	14	10	8		
64	Имитация	5	11		18		10	5	
	Аналитика	<u>6</u>	<u>10</u>		18		10	6	
96	Имитация	5	11	17,1	23	18	14	8	5
	Аналитика	<u>6</u>	<u>10</u>	<u>14</u>	<u>18</u>	18	14	8	6

Таблица 3. Эффективная емкость МКМ в основном варианте.
Начальная передача в кольцо *p*

<i>N</i>	<i>n</i>	2	4	6	8	12	16	32	48
16	Имитация	5	10		5,1				
	Аналитика	6	10		<u>6</u>				
24	Имитация	5,1	10	11,2	8,1	5			
	Аналитика	6	10	10	<u>8</u>	<u>6</u>			
32	Имитация	5,1	10		11,1		5		
	Аналитика	6	10		<u>10</u>		<u>6</u>		
48	Имитация	5,1	10	14	17,1	11,1	8,1		
	Аналитика	6	10	14	<u>14</u>	<u>10</u>	<u>8</u>		
64	Имитация	5	10		18		11	5	
	Аналитика	6	10		18		<u>10</u>	<u>6</u>	
96	Имитация	5	10	14	18	23,1	17	8	5
	Аналитика	6	10	14	18	<u>18</u>	<u>14</u>	<u>8</u>	<u>6</u>

Отметим хорошее совпадение результатов аналитической и имитационных моделей в областях точных оценок первой. При этом трафик в основном варианте оказался столь же равномерным как в первом варианте.

Результаты экспериментов с основным (3-м) вариантом имитационной модели представлены в Таблица 2 для случая, когда начальным для передачи является кольцо 1, и в Таблица 3 для случая, когда начальным является кольцо p .

Приведенные выше результаты позволяют сделать вывод о достаточной точности имитационной модели, что делает ее основным инструментом дальнейшего исследования свойств МКМ и ДМКМ.

7. Экспериментальные свойства МКМ и ДМКМ

Первое свойство МКМ состоит в том, что в максимальные значения \hat{C} в Таблица 2 для любых N находятся при больших или равных значениях n , чем в Таблица 3. Отметим, что такие кольца в МКМ требуются меньшего расхода кабеля.

Эксперименты с имитационной моделью показали, что МКМ при прямом переносе пакетов и при переносе с приостановками имеет разные эффективные емкости. При доставке пакетов с приостановкой эффективная емкость снижается на 20÷30%. Такое снижение является следствием увеличения реальной длины маршрутов. Действительно, пусть пакет со случайной длиной e ($1 \leq e \leq p-1$) приостанавливается после 1-го этапа. Тогда после приостановки пакета средняя длина \bar{e} его маршрута оказывается больше, чем до приостановки, а именно: $\bar{e} = 1 + (p-1)/2 > p/2 + 1/2$.

Таблица 4. Эффективная емкость ДМКМ в основном варианте.
Начальная передача в кольцо 1

N/n	2	4	6	8	12	16	32	48	64
16	15	16		13,3					
24	15,7	24	24	19,7	14,2				
32	16	31		29,3		14,5			
48	16,4	32,3	47	45,3	30,9	21,7			
64	16,5	33		60,9		31,6	15,3		
96	16,7	33,7	50	65,8	64,1	48,6	22,8	15,5	
128	16,8	34,1		67		65,6	32,8		15,7

Имитационная модель создавалась сразу для ДМКМ, а все предшествующие исследования проводились на ее односторонней половине. Такая нацеленность объясняется ожиданием [[4]], что эффективная емкость увеличится более чем в 2 раза при совместном использовании встречных МКМ. В Таблица 4 представлены результаты экспериментов в основном варианте имитационной модели для ДМКМ. Максимальные значения эффективной емкости в них выделены жирным шрифтом.

7. Практические ограничения

Для оценки практически достижимых эффективных емкостей МКМ и ДМКМ необходимо ограничить размеры очереди пакетов Q в каждом узле. Имитационное моделирование было проведено для $Q=10^9$, при $q=0, 1/2, 1, 2, 3, 4$. Полученные при таком моделировании эффективные емкости обозначим \hat{C}_q . В предыдущем разделе фактически рассматривались значения \hat{C}_4 , полученные в различных вариантах функционирования модели.

Рассмотрим величину отношения $\rho_q = \hat{C}_q / \hat{C}_4$ для условий Таблица 3 и Таблица 4 и тех N в них, при которых \hat{C}_q достигает максимальных значений.

Таблица 5. Значения отношений емкостей $\rho_q = \hat{C}_q / \hat{C}_4$ для МКМ

N	16	24	32	48	64	96
$q=2$	0,99	0,83	0,8	0,79	0,99	0,75
$q=1$	0,93	0,83	0,79	0,78	0,96	0,75
$q=0,5$	0,86	0,83	0,8	0,78	0,9	0,75
$q=0$	0,73	0,76	0,77	0,78	0,82	0,73

Таблица 6. Значения отношений емкостей $\rho_q = \hat{C}_q / \hat{C}_4$ для ДМКМ

N	16	24	32	48	64	96
$q=2$	1	1	0,99	1	0,99	0,96
$q=1$	1	1	0,94	0,93	0,87	0,92
$q=0,5$	0,96	0,92	0,82	0,77	0,72	0,82
$q=0$	0,74	0,68	0,62	0,57	0,54	0,64

В Таблица 5 приводятся значений ρ_q при разных N для МКМ, а в Таблица 6 – для ДМКМ. Видно, что при $Q = 3$ ($q = 1/2$) имеет

место приемлемое для практики уменьшение значений эффективных емкостей.

7. Сравнительные характеристики

Сравним полученные характеристики для МКМ и дуплексного кольца (ДК) и дуплексного МКМ (ДМКМ) и пары дуплексных колец (ПДК).

Сначала сравним их диаметры $D_{МКМ}$, $D_{ДМКМ}$ и $D_{ДК}$. На Рис. 2 приводятся значения отношений $\alpha = D_{ДК}/D_{МКМ}$ и $\beta = D_{ДК}/D_{ДМКМ}$ с учетом того $D_{ДК} = D_{ПДК}$. Видно, что в практически важном диапазоне $16 \leq N \leq 64$ имеет место уменьшение диаметра МКМ в 1,3÷ 2,3 раза и диаметра ДМКМ – в 2÷4 раза.

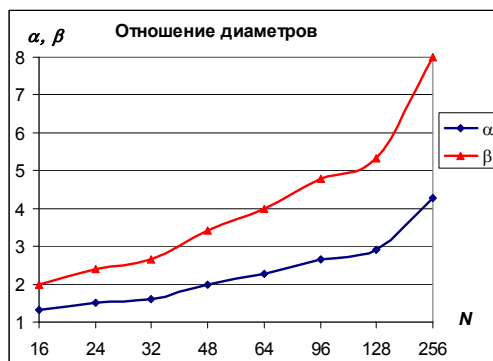


Рис. 2. Отношения диаметров $\alpha = D_{ДК}/D_{МКМ}$ и $\beta = D_{ДК}/D_{ДМКМ}$.

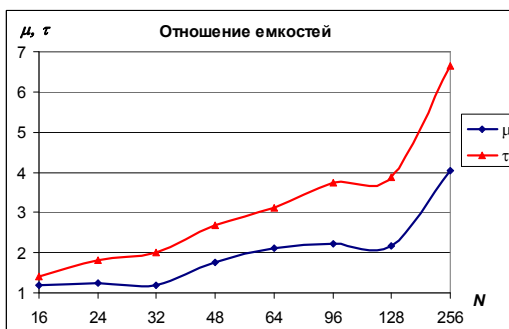


Рис. 3. Отношения эффективных емкостей $\mu = \hat{C}_{МКМ}/\hat{C}_{ДК}$ и $\tau = \hat{C}_{ДМКМ}/\hat{C}_{ПДК}$ при $Q = 3$.

Теперь сравним их пропускные способности, заданные максимальными эффективными емкостями. На Рис. 3 приводятся значения отношений $\mu = \hat{C}_{\text{МКМ}}/\hat{C}_{\text{ДК}}$ (Таблица 3), и $\tau = \hat{C}_{\text{ДМКМ}}/\hat{C}_{\text{ПДК}}$ (Таблица 4) для $Q = 3$ ($q = 1/2$). При этом значения $\hat{C}_{\text{ДК}}$ и $\hat{C}_{\text{ПДК}}$ получены в так же в имитационной модели для $q=1$. Видно, что в практически важном диапазоне $16 \leq N \leq 64$ имеет место увеличение пропускной способности МКМ в $1,2 \div 2,1$ раза и пропускной способности ДМКМ – в $1,4 \div 3,1$ раза.

Известно [[6]], что в модели $M/G/1$ теории массового обслуживания зависимость задержки передачи пакета $T(s)$ от загрузки кольца s для симплексного кратного кольца имеет вид:

$$(5) \quad T(s) = F/(c-s).$$

Здесь s – загрузка кольца $s = \lambda B$, где λ (сек⁻¹) – суммарная интенсивность генерации пакетов, $B = b/v$ (сек) – средняя длительность пакета, а b (бит) – средняя длина пакета. Функция $F(s, b, d)$ (сек) зависит от загрузки s и ее дисперсии, средней длины пакетов b и ее дисперсии d , но не зависит от емкости c . В формуле (5) емкость кольца задает положение вертикальной асимптоты гиперболической зависимости роста задержки от загрузки. При высокой загрузке ($s > 0,7c$) увеличение емкости в $1,5 \div 2$ раза приводит к уменьшению задержки передачи в несколько раз.

При равномерной загрузке узлов знаменатель формулы (5) можно выразить как $N(\gamma - \sigma)$, где $\gamma = c/N$ – удельная емкость, $\sigma = \lambda B$ – удельная загрузка и λ – интенсивность генерации пакетов узлом.

По аналогии можно предположить, что для МКМ и ДМКМ зависимость $T(s)$ имеет такой же вид с заменой емкости кольца c на эффективную емкость мультикольца \hat{C} . Поэтому можно сделать вывод, что оба отмеченных фактора преимущества МКМ над ДК и ДМКМ над ПДК по совокупности приводят к сокращению задержек доставки пакетов по системной сети в несколько раз.

8. ДМКМ с разреженными кольцами

В [[7]] был разработан метод повышения пропускной способности нескольких дуплексных колец (ДК), за счет образования их них мультикольца с разреженными кольцами. Он опирается на метод инвариантного расширения произвольных сетей с использованием топологии квазиполного графа [[8]], описываемого симметричными блок-схемами, рассматриваемыми в комбинаторике.

В данном разделе этот метод применяется для ДМКМ вместо ДК. В нем исходная сеть из 2-х одинаковых ДМКМ с N узлами по 8 портов в каждом преобразуется в сеть ДМКМРК из 3-х ДМКМ с разреженными кольцами, содержащую $N_{\text{ДМКМРК}} = 1,5N_{\text{ДМКМ}}$ узлов при неизменном числе портов в узлах. В Таблица 7 показан пример вставки узлов в кольца. Топология каждого ДМКМРК задается по порядковым номерам узлов.

Таблица 7. Таблица размещения узлов в кольцах при $N_{\text{ДМКМ}} = 10$,
 Π – порядковые номера узлов, в ячейках – номера узлов кольца

ДМКМ/ Π	1	2	3	4	5	6	7	8	9	10
1	1	3	4	6	7	9	10	12	13	15
2	1	2	4	5	7	8	10	11	13	14
3	2	3	5	6	8	9	11	12	14	15

Отметим, что представленный подход ориентирован на тор, измерения которого содержат 4 дуплексных кольца с 8 портами в узлах. По сравнению с этой сетью пропускная способность сети ДМКМРК в каждом измерении увеличится в $1,5\tau \hat{C}_{\text{ДМКМ}}(N_{\text{ДМКМ}})/\hat{C}_{\text{ПДК}}(N_{\text{ДМКМРК}})$ раз, где τ берется из графика Рис. 3, а \hat{C} – из Таблица 6. В результате емкость ДМКМРК больше емкости двух ПДК в $1,6 \div 3,0$ раз при $24 \leq N_{\text{ДМКМРК}} \leq 96$.

9. Резервы тора *Gemini*

3-мерный тор суперкомпьютера *Gemini* [[3]] имеет в 1-м измерении 2 одинаковых кольца с $N=16$ узлами. Во 2-м и 3-м измерениях этот тор содержит по 4 одинаковых дуплексных кольца с $N=32$ узлами. При этом 1-е измерение имеет диаметр $D=8$ и суммарную емкость $\hat{C} \leq 16$. Соответственно 2-е и 3-е измерения имеют $D=16$ и суммарную емкость каждого $\hat{C} \leq 32$. Отметим, что во всех измерениях удельная емкость $\gamma = \hat{C}/N$ не превосходит 1.

При использовании в 1-м измерении ДМКМ его диаметр уменьшится до $D=4$, а суммарная емкость \hat{C} вырастет в $\tau \approx 1,2$ раза, и удельная емкость станет $\gamma \geq 1$. При этом имеется возможность увеличивать число узлов N в несколько раз, обеспечивая пропорциональное повышение емкости \hat{C} и дальнейший рост удельной емкости γ .

Во 2-м и 3-м измерениях можно использовать ДМКМРК. Если его строить с $N_{\text{ДМКМ}} = 24$, то число узлов станет $N_{\text{ДМКМРК}} = 36$. При

этом диаметр останется таким же как в ДМКМ с $N_{\text{ДМКМ}} = 24$, т.е. $D = 5$, а суммарная емкость достигнет $\hat{C}_{\text{ДМКМРК}} = 3\hat{C}_{\text{ДМКМ}} \approx 66$ при $Q=3$ или $\hat{C}_{\text{ДМКМРК}} = 72$ при $Q=10$, а удельная емкость – $\gamma \approx 1,8$ или $\gamma = 2$. Отметим значительное улучшение рассматриваемых характеристик по сравнению с исходным вариантом тора *Gemini*.

Во 2-м и 3-м измерениях опять имеется возможность увеличивать $N_{\text{ДМКМРК}}$ в несколько раз, обеспечивая пропорциональное повышение емкости $\hat{C}_{\text{ДМКМРК}}$ с сохранением $\gamma \approx 2$.

Все это вместе взятое не только обеспечивает повышение быстродействия модифицированного тора, но и обеспечивает возможность достижения этого свойства при значительно большем числе узлов.

Заключение

В работе рассмотрен способ сокращения диаметра и повышения пропускной способности пары встречных цепочечных колец и двух-четырех пар таких колец. Результат получен за счет изменения топологии пары колец – за счет превращения их в пару коммутируемых колец с разными шагами в каждом кольце. Уменьшение диаметра и повышение пропускной способности ведет к значительному сокращению задержек доставки пакетов по сети, состоящей их таких пар колец.

Список литературы

- [1] Позднеев А. Blue Gene/Q: как обуздать массивный параллелизм // URL: <http://2013.nscf.ru/Presentations/IBM/AlexanderPozdneev-HowToHarnessMassiveParallelism.pdf>.
- [2] Михеев В.А. и др. Реализация высокоскоростной сети для суперкомпьютерных систем: проблемы, результаты, развитие // URL: http://2013.nscf.ru/TesisAll/Section%201/12_2761_SimonovAS_S1.pdf.
- [3] Iverson R., Roweth D. and Kaplan L. The Gemini System Interconnect // 18th IEEE Symposium on High Performance Interconnects. 2009. P. 3–87.
- [4] Алленов А.В., Подлазов В.С., Стецюра Г.Г. Пропускная способность набора кольцевых каналов I. Класс наборов

- колец. Наборы с простыми узлами // *АиТ*. 1996. № 3. С. 135–144.
- [5] Алленов А.В., Подлазов В.С. Пропускная способность набора кольцевых каналов II. Кольцевые коммутаторы // *АиТ*. 1996. № 4. С. 162–172.
- [6] Андреев Л.В. Однонаправленные кольцевые сети связи с коммутацией пакетов // *Проблемы передачи информации*. 1982. Т. 18. Вып. 4. С. 85–103.
- [7] Каравай М.Ф., Подлазов В.С. Топологические резервы суперкомпьютерного интерконнекта // *Управление большими системами*. 2013. вып. 41. М.: ИПУ РАН. с . 395–423.
URL: <http://ubs.mtas.ru/upload/library/UBS4114.pdf>.
- [8] Каравай М.Ф., Подлазов В.С. Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть. // *АиТ*. 2010. №. 10. С. 166–176.

СВЕДЕНИЯ ОБ АВТОРАХ

Каравай Михаил Федорович – зав. лабораторией технической диагностики и отказоустойчивости, учреждение Российской академии наук Институт проблем управления (ИПУ) РАН. Московский авиационный институт (МАИ) – 1963 г.
Д.т.н., доцент.

107 печатных работ и 3 монографии.

Диагностика, отказоустойчивость, многопроцессорность, системные сети.

mkaravay@ipu.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-90-00.

Mikhail Karavay, Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor, Profsoyuznaya st., 65, (495)334-90-00.

Подлазов Виктор Сергеевич – г.н.с., учреждение Российской академии наук Институт проблем управления (ИПУ) РАН. Московский физико-технический институт (МФТИ) – 1968 г.
Д.т.н., доцент.

111 печатных работ и 2 монографии.

Локальные и системные сети связи, сетевое обеспечение отказоустойчивости многопроцессорных вычислительных систем.

podlazov@ipu.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-78-31.

Viktor Podlazov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor, Profsoyuznaya st., 65, (495)334-78-31.