

М. Ф. Каравай, В. С. Подлазов

Скрытые резервы системных сетей суперкомпьютеров Blue Waters (IBM) и Gemini (CRAY)

АННОТАЦИЯ. Рассматриваются простые возможности повышения характеристик интерконнекта суперкомпьютеров *Blue Waters* (IBM) и *Gemini* (CRAY) за счет использования системных сетей с прямыми каналами.

Ключевые слова и фразы: параллельные многопроцессорные вычислительные системы, системные сети, самомаршрутизируемые неблокируемые сети, прямые каналы, распределенные полные коммутаторы, некоммутируемые мультикольца.

Введение

В основе структуры суперкомпьютеров *Blue Waters* и *Gemini* [1], [2] лежит пара тесно связанных узлов — процессорного узла и высокоинтеллектуального связного узла с большим числом портов для организации системной сети. *Blue Waters* и *Gemini* имеют процессорные узлы с многоядерными процессорами. Узлы связи имеют 47 и 20 высокоскоростных дуплексных портов соответственно.

В системной сети *Blue Waters* каждый узел связи использует межузловые каналы трех видов: 7 каналов K_1 высокой пропускной способности V_1 , 24 канала K_2 низкой пропускной способности $V_2 = V_1/5$ и 16 каналов K_3 средней пропускной способности $V_3 = 2V_2$. Каналы K_1 выполнены медным кабелем, а каналы K_2 и K_3 — оптическим кабелем.

32 узла связи образуют суперузел (Рис. 1), в котором узлы связаны по схеме полного графа каналами K_1 и K_2 . Среди них выделяются 4 группы по 8 узлов, связанных каналами K_1 . Остальные узлы связаны каналами K_2 .

Каждый суперузел имеет 512 каналов K_3 . В максимальной конфигурации *Blue Waters* каждый такой канал используется для связи с другим суперузлом по схеме полного графа. В этом случае *Blue Waters* содержит 513 суперузлов и в них $513 \cdot 32 \cdot 4 > 64K$ процессоров, связанных каналами разной пропускной способности. Между любыми двумя узлами имеется единственный путь длиной не более 3-х скачков (смен передач пакетов по прямым каналам с промежуточной их буферизацией). Имеется единственный резервный путь длиной не более 5-и скачков.

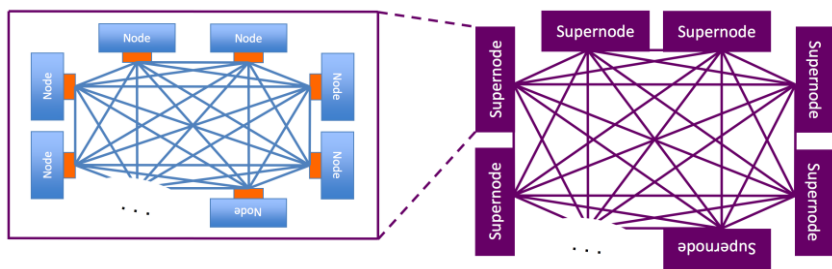


Рис. 1. Структура системной сети суперкомпьютера Blue Waters

Скрытым резервом данной системной сети является, во-первых, неоптимальное использование каналов K_1 максимальной пропускной способности для создания суперузла. Дополнительное использование при каждом узле связи коммутатора 7×7 каналов K_1 открывает возможность построения суперузлов с большим числом узлов, связанных только каналами K_1 , и освобождения каналов K_2 для связи с дополнительными суперузлами или для повышения пропускной способности системной сети. Так построенные суперузлы имеют топологию распределенного полного коммутатора в виде квазиполного графа или орграфа [3]–[8]. Эта оптимизация позволяет существенно увеличить как число узлов, связанных каналами максимальной пропускной способности, так и общее число узлов в системной сети. При этом узел связи остается неизменным.

Еще одним резервом является неоптимальное использование каналов K_3 для объединения суперузлов в систему. Дополнитель-

ное использование при каждом узле связи коммутаторов каналов K_3 , позволит увеличить число суперузлов и/или пропускную способность сети за счет замены полного графа на один или несколько минимальных квазиполных графов в виде распределенных полных коммутаторов.

В системной сети *Gemini* каждый узел связи входит в 10 дуплексных колец, образующих 3D-тор. Измерения x и y состоят из 4-х одинаковых дуплексных колец, измерение z — из 2-х одинаковых дуплексных колец. Общее число N узлов связи составляет величину $N=N_x N_y N_z$, где N_u — число узлов в кольце u -го измерения.

Скрытым резервом данной системной сети является неоптимальность использования множества колец. В каждом измерении все кольца имеют одинаковую топологию (последовательность соединения узлов). Использование колец с разной топологией открывает возможность существенного (в разы) повышения пропускной способности множества колец каждого измерения [8]–[10]. При этом узел связи остается неизменным и меняется только алгоритм выбора кольца для передачи пакета данных.

1. Метод расширения полных коммутаторов

Пусть у нас имеется исходная системная сеть с m дуплексными портами в виде полного коммутатора $m \times m$. Необходимо расширить ее до сети $PC(N)$ ($N > m$), сохраняя маршрутные свойства исходного коммутатора. Такими свойствами являются неблокируемость и самомаршрутизируемость на произвольной перестановке пакетов данных от N абонентов [3], [5].

Для этого строится расширенная сеть $PC(N)$ в виде квазиполного графа (Рис. 2).

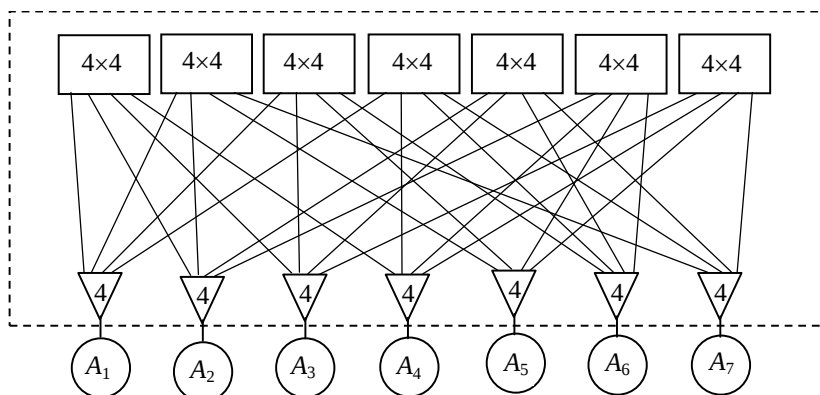


РИС. 2. Расширенная сеть PC(7,4,2)

Последний является двудольным графом, каждую долю которого составляют N вершин степени m , и путь любой длины 2 между любыми двумя вершинами одной доли проходит через одну и только одну вершину другой доли, и имеется точно σ разных путей между любыми двумя вершинами одной доли. В квазиполном графе, если он существует, число вершин одной доли задается выражением $N = m(m-1)/\sigma + 1$. Зададим вершины одной доли коммутаторами $m \times m$, ребра — дуплексными каналами, а вершины другой доли — абонентами, число портов в которых увеличено до m за счет использования разветвителя Pm дуплексного канала $1 \times m$. Тогда квазиполный граф задает схему распределенного полного коммутатора, который обладает свойствами неблокируемости и самомаршрутизируемости.

Схема межсоединений в квазиполных графах описывается таблицами, задающими симметричные блок-схемы $B(N, m, \sigma)$, изучаемые в комбинаторике. В ТАБЛИЦА 1 дается пример таблицы межсоединений для графа на РИС. 2. В ячейках строк таблицы находятся номера абонентов, подсоединенных к коммутатору, указанному в той же строке.

ТАБЛИЦА 1. Схема подсоединений абонентов к коммутаторам в $PC(7,4,2)$

4×4	$PC(7,4,2)$			
1	1	2	3	4
2	1	2	5	7
3	1	3	5	6
4	1	4	6	7
5	2	3	6	7
6	2	4	5	6
7	3	4	5	7

Часть $PC(N, m, \sigma)$ выше интерфейсов с абонентами (внутри пунктирного прямоугольника на Рис. 1 представляет собой распределенный полный коммутатор $N \times N - PK(N, m, \sigma)$. При $\sigma=1$ он имеет в $\sim N^{1/2}$ раз меньше каналов, чем сеть со структурой полного графа $N \times N$.

2. Использование распределенных коммутаторов в суперузле системной сети Blue Waters

Предположим, что каждый связной узел в *Blue Waters* дополнен одним полным коммутатором 7×7 для каналов K_1 . Пусть выбор выходных портов этого коммутатора осуществляется по их номерам в заголовке каждого пакета. Тогда можно создать суперузел с использованием только каналов K_1 связного узла и этих дополнительных коммутаторов. Такой суперузел имеет структуру распределенного полного коммутатора, в котором порты каналов K_1 выполняют роль портов разветвителя P_7 .

Сначала попробуем построить суперузел в виде распределенного коммутатора $PK(43, 7, 1)$ со структурой квазиполного графа. «В чистом виде» эта попытка неосуществима, т.к. блок-схема $B(43, 7, 1)$ не существует. Однако если допустить, что некоторые узлы связаны параллельно более чем через один коммутатор 7×7 , то можно построить $PK(39, 7, 1 \setminus 2)$ [12]. Схема дуплексных межсоединений для него обладает тем свойством, что каждый абонент i связан допол-

нительным путем через один коммутатор с абонентами, чьи номера задаются как $(i \times 1) \bmod 39$ и $(i \times 2) \bmod 39$.

Таким образом РК(39,7,1\2) позволяет создать суперузел из 39 узлов связи, объединенных неблокируемыми самомаршрутизируемыми каналами K_1 . В нем любые два узла связаны прямыми (без промежуточной буферизации пакетов) каналами, которые остаются бесконфликтными при любой перестановке пакетов данных между узлами. Они прокладываются путем червячной маршрутизации.

Теперь появляется возможность освободить каналы K_2 каждого узла в любом суперузле для связи с другими суперузлами не трогая топологию связей между суперузлами (полный граф). Их можно использовать двояко — оставить по одному каналу K_2 или K_3 между любой парой суперузлов или удвоить число таких каналов. В первом случае число суперузлов увеличится с 513 до $1561 = 39 \times (16 + 24) + 1$, т.е. более чем втрое. Во втором случае число суперузлов увеличится только до 781, но появляется возможность либо увеличить вдвое пропускную способность системной сети и/или снизить «длину» резервного пути между любыми узлами с 5 до 3 скачков.

Специально отметим, что повышение характеристик системной сети здесь достигается за счет добавления коммутатора 7×7 к каждому узлу связи. Коммутатор много проще узла связи (фактически он уже имеется в его составе) и имеет много меньшее энергопотребление. Тем самым отмеченное повышение характеристик системной сети достигается при незначительном увеличении ее энергопотребления.

3. Применение распределенных коммутаторов для связи между суперузлами Blue Waters

Аналогично можно оптимизировать структуру связей между суперузлами посредством замены полного графа на минимальный квазиполный граф. Для этого достаточно дополнить каждый суперузел с M узлами коммутатором $M \times M$ каналов K_3 и построить

распределенный коммутатор для суперузлов, используя от каждого узла только один (!) канал K_3 . При этом число суперузлов N достигает величины $N=M(M-1)+1$. Правда, для квазиполного графа возникает проблема его построения — для больших M она решена только для случая, когда $M-1$ является простым числом [4]. В частности, для суперузла с $M=39$ узлами такой квазиполный граф еще не построен. Однако, если взять $M=38$ (просто не подключая 39-й узел к коммутатору), то имеется возможность построить такой квазиполный граф. Авторы такой квазиполный граф построили.

Как следствие, мы можем построить системную сеть в которой число суперузлов N достигает величины $N=38\cdot 37+1=1407$ (против 513 в *Blue Waters*).

Можно использовать все 16 каналов K_3 для связи с другими суперузлами посредством использования 16 коммутаторов 38×38 . Все они должны иметь разные схемы межсоединений узлов и коммутаторов — например за счет перенумерации узлов, подключаемых к очередному коммутатору.

В результате, во-первых, каждый суперузел связан 16 независимыми путями (через разные узлы в нем) с любым другим суперузлом. Это в 15 раз повышает пропускную способность системной сети между суперузлами (используем один канал как резервный 3-х-скачковый путь). Во-вторых, каждый узел связан 16 разными 2-скачковыми и 3-скачковыми путями (через разные коммутаторы в суперузле) с любым узлом в любом другом суперузле. Это в 15 раз повышает пропускную способность системной сети между узлами в разных суперузлах. В-третьих, многократно повышается вероятность появления 2-скачковых путей (15/38 против 1/32 в *Blue Waters*). Это сокращает среднее время доставки пакетов по сети.

Таким образом мы получаем системную сеть с $1407\cdot 38=53466$ узлами (против 16416 в *Blue Waters* — примерно в 3,5 раз больше) с многократно повышенной пропускной способностью и отказоустойчивостью.

Накладные затраты здесь — это увеличение энергопотребления каждого суперузла за счет добавления 16 коммутаторов 38×38 . В

каждом суперузле имеется $39 \cdot 16 = 624$ дуплексных портов каналов K_3 . Добавлено $16 \cdot 38 = 608$ дуплексных портов каналов K_3 в коммутаторах. Порты в узлах и в коммутаторах работают на одной скорости и с одинаковой интенсивностью. Поэтому энергопотребление суперузла увеличивается не более чем вдвое при условии, что коммутаторы выполнены по той же технологии, что и узлы связи.

Для снижения энергопотребления коммутатор 38×38 должен быть оптоэлектронным и самомаршрутизируемым — оптическим по внутренним каналам и электронным по управлению их коммутацией. Такой коммутатор потребляет много меньше энергии — только на переключение и поддержание каналов. Существуют ли такие коммутаторы достаточно большого размера (на $30 \div 50$ портов) — это открытый вопрос, но он уже в «повестке дня» современной технологии [13].

К недостаткам рассмотренного выше «лобового» построения системной сети относятся использование больших коммутаторов, избыточное число независимых каналов и невозможность использования портов K_2 для повышения характеристик системной сети. Последние составляют топологический резерв, который может быть эффективно использован при другой структуре системной сети, при которой исправляются и отмеченные недостатки.

Такую структуру задает, например, модифицированный 3-мерный обобщенный гиперкуб, в котором полные графы каждого измерения заменены на квазиполные графы. Обобщенный гиперкуб — это многомерная решетка, каждая строка или столбец которой имеет связи со структурой полного графа. Для построения такой системной сети коммутаторы каналов K_1 , K_3 и K_2 размещаются при узлах связи. Они используются для построения в разных измерениях суперузлов в виде распределенных полных коммутаторов с разным быстрым действием (Рис. 3).

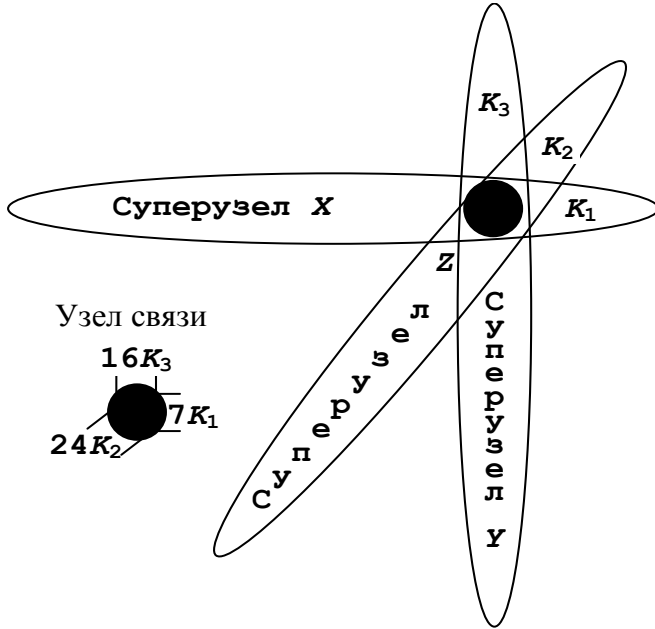


Рис. 3. Системная сеть как 3-мерный обобщенный гиперкуб с квазиполными графами в каждом измерении

Системная сеть при этом остается 3-х-скачковой и энергопотребление узлов связи с добавленными коммутаторами увеличивается не более чем в 2 раза.

Предложенная структура имеет большое число вариантов построения в зависимости от выбираемого вида квазиполного графа. Два из них приведены в Таблица 2. Почти все они позволяют многократно увеличивать число узлов и пропускную способность между ними. Однако для многих из этих вариантов еще не построены необходимые квазиполные графы (выделены подчеркиванием). Этой работой мы заняты в настоящее время.

ТАБЛИЦА 2. Параметры системной сети в виде модифицированного 3-мерного обобщенного гиперкуба

Системная сеть	Суперузел X: каналы K_1	Суперузел Y: каналы K_3	Суперузел Z: каналы K_2	Общее число узлов связи
Число коммутаторов при узле	1 – 7×7	2 – 8×8	3 – 8×8	$N=N_X N_Y N_Z = 126711$
Число узлов	$N_X = 39$	$N_Y = 57$	$N_Z = 57$	
Число путей	1	2	3	
Число коммутаторов при узле	1 – 7×7	1 – 16×16	1 – 24×24	$N=N_X N_Y N_Z = 93940$
Число путей	2	4	8	
Число узлов	$N_X = 21$	$N_Y = 61$	$N_Z = 70$	

4. Метод расширения произвольных системных сетей

Как уже отмечалось во введении, скрытым резервом системной сети *Gemini* является неоптимальность использования множества дуплексных колец. В каждом измерении все кольца имеют одинаковую топологию (последовательность соединения узлов). Использование набора колец разной топологии открывает возможность существенного повышения пропускной способности множества колец каждого измерения.

Этот набор колец строится посредством расширения дуплексного кольца методом инвариантного расширения произвольных системных сетей. Получаемый набор колец является мультикольцом с разреженными кольцами.

Рассмотрим кратко сам метод расширения произвольной сети. Пусть у нас имеется исходная произвольная системная сеть (СС) с K абонентами $ИС(K)$. Необходимо расширить ее до сети $РС(R)$ с R абонентами ($R > K$), сохраняя ее маршрутные свойства [3], [14].

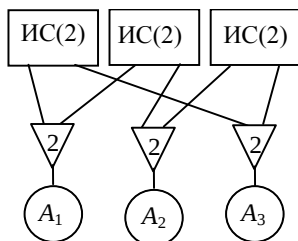


Рис. 4. ПРС(3,2,1)

Для этого сначала строится простейшая расширенная сеть ПРС($N, m, 1$), которая расширяет ИС(m) с фиксированным $m < K$. Она строится в виде минимального квазиполного графа (Рис. 1). Последний является двудольным графом, каждую долю которого составляют N вершин степени m , и путь между любыми двумя вершинами одной доли проходит через одну и только одну вершину другой доли, и в котором $N = m(m-1) + 1$. Вершины одной доли задают копии ИС(m), ребра — дуплексные каналы, а вершины другой доли — абонентов, число портов в которых увеличено до m .

Схемы межсоединений в квазиполных графах описываются таблицами, задающими симметричные блок-схемы $B(N, m, 1)$ [3], [14]. В Таблица 3 дается пример схемы межсоединений для графа на Рис. 4. в ячейках таблицы находятся номера абонентов.

ТАБЛИЦА 3. Схема подсоединений абонентов к ИС(2) в ПРС (3)

ИС(2)	ПРС(3,2,1)	
1	1	3
2	1	2
3	2	3

В общем случае $K > m$ сеть РС(R) набирается из ПРС($N, m, 1$) как из строительных блоков. Делается это следующим образом. Каждая копия ИС(K) разбивается на $\lfloor K/m \rfloor$ равных частей (каждая с m портами), которые нумеруются диапазоне $0 \leq j < \lceil K/m \rceil$. Объединим все j -ые части в каждой копии ИС(K) в j -ую ПРС($N, m, 1$). 0-ая ПРС должна иметь стандартную структуру. К j -ой ПРС подсоединим абонентов с номерами от jN до $N(j+1)-1$ так,

чтобы абоненты на одинаковых позициях имели номера на N больше, чем номера в $(j-1)$ -ой ПРС.

Таким образом, мы получим $PC(R)$, состоящую из $\lceil K/m \rceil$ ПРС($N, m, 1$), объединяющую R абонентов, где R находится в диапазоне: $N \lfloor K/m \rfloor \leq R \leq N \lceil K/m \rceil$. Для примера в Таблица 4 показана таблица межсоединений в $PC(18)$, построенная из ИС(12) на базе ПРС(3,2,1). Узлы с подчеркнутыми номерами можно исключить для сохранения числа узлов неизменным.

ТАБЛИЦА 4. Таблица подсоединений абонентов к ИС(12) в $PC(18)$

ИС(12)	PC(18)											
	1-я ПРС		2-я ПРС		3-я ПРС		4-я ПРС		5-я ПРС		6-я ПРС	
1	1	3	4	6	7	9	10	12	13	15	16	<u>18</u>
2	1	2	4	5	7	8	10	11	13	14	16	<u>17</u>
3	2	3	5	6	8	9	11	12	14	15	<u>17</u>	<u>18</u>

5. Расширение кольцевых системных сетей — разреженные некоммутируемые мультикольца

Пусть ИС(K) состоит из m дуплексных кратных колец с K узлами. Кратные кольца имеют цепочечную структуру, при которой пакет удаляется из кольца его приемником, освобождая кольцо для параллельного использования разными источниками. Кратное симплексное кольцо имеет пропускную способность $W=2v$, где v (бит/сек) — скорость передачи по кольцу [10]. Для дуплексного кратного кольца, состоящего из пары встречных симплексных колец, пропускная способность зависит от распределения длин маршрутов и расписания передач. При равномерном распределении длин маршрутов и при передаче пакетов в кольцо с кратчайшим маршрутом дуплексное кольцо имеет асимптотическую (при $K \gg$

1) пропускную способность $W=8v$ [10], [11], где v — скорость передачи по симплексному кольцу.

Для расширения ИС(K) будем использовать способ, рассмотренный в разделе 1. Для этого создадим у каждого абонента m дуплексных портов, например, с помощью разветвителя дуплексных каналов $1 \times m/m \times 1$. Пусть m является делителем K . В кольце это легко сделать введя несколько пустых узлов, не имеющих абонентов. РС(R) содержит $N=m(m-1)+1$ копий ИС(K), объединяет $R=KN/m$ абонентов и имеет пропускную способность $W=NW_2$. Последовательность подсоединения абонентов к узлам колец задается процедурой разбиения N копий ИС(K) на ПРС($N, m, 1$).

Таким образом построено некоторое мультикольцо, в котором любые два абонента связаны только через одно дуплексное кратное кольцо. В этом мультикольце дуплексные кольца являются «малыми» (K узлов при общем числе узлов $R \gg K$), имеют разную топологию (последовательности соединения узлов и переменный шаг между узлами). Это мультикольцо мы называем разреженным, т.к. в каждое дуплексное кольцо входит только часть узлов. Его удастся исследовать сразу аналитически — пропускная способность составляет $W=8Nv$, а суммарная длина кабеля — $E=2N$ у.е. (1 у.е. — длина кабеля в полном кольце со всеми узлами).

6. Использование разреженных мультиколец в Gemini

Использование разреженных мультиколец в каждом измерении открывает возможность существенного повышения пропускной способности каждого измерения. При этом в узле связи меняется только алгоритм выбора кольца для передачи пакета данных.

Для измерения z $m=2$. Возьмем 3 дуплексных кольца на 12 абонентов. Тогда мультикольцо объединит 15 абонентов (Таблица 4), а его пропускная способность составит $3 \times 7,5 = 22,5v$. Это в 1,5 раза больше, чем в измерении z Gemini.

Для измерений x и y $m=4$. Возьмем 13 дуплексных колец на 12 абонентов. Тогда мультикольцо объединит 39 абонентов (Таблица 5), а его пропускная способность составит $13 \times 7,5 = 97,5v$. Это в 3 с

лишним раза больше, чем в измерениях x и y *Gemini*. В ТАБЛИЦА 5 для наглядности ИС(12) заменены на дуплексные кольца.

ТАБЛИЦА 5. Подсоединения узлов к кольцам в мультикольце РС(39)

ДК(12)	Мультикольцо РС(39)											
	1-я ПРС(13,4,1)				2-я ПРС(13,4,1)				3-я ПРС(13,4,1)			
1	1	2	3	4	14	15	16	17	27	28	29	30
2	1	5	11	13	14	18	24	26	27	31	37	39
3	1	6	8	12	14	19	21	25	27	32	34	38
4	1	7	9	10	14	20	22	23	27	33	35	36
5	2	5	7	12	15	18	20	25	28	31	33	38
6	2	6	9	13	15	19	22	26	28	32	35	39
7	2	8	10	11	15	21	23	24	28	34	36	37
8	3	5	6	10	16	18	19	23	29	31	32	36
9	3	7	8	13	16	20	21	26	29	33	34	39
10	3	9	11	12	16	22	24	25	29	35	37	38
11	4	5	8	9	17	18	21	22	30	31	34	35
12	4	6	7	11	17	19	20	24	30	32	33	37
13	4	10	12	13	17	23	25	26	30	36	38	39

Отметим, что повышение пропускной способности каждого измерения достигнуто без изменения числа портов узлов связи. Фактически это означает, что за счет увеличения числа колец и выбора места подключения узлов к ним можно существенно повысить пропускную способность любого многоканального nD -тора. Иначе говоря, только за счет удлинения связного кабеля, выбора места и последовательности подсоединения узлов к нему можно в разы повысить пропускную способность такого тора, и, самое главное, сделать это без повышения энергопотребления.

Заключение

Распределенный полный коммутатор и некоммутируемое мультикольцо разрабатывались авторами как самостоятельные системные сети для МВС с несколькими десятками или сотнями абонентов, имеющие минимальные времена доставки пакетов данных по прямым каналам. Данная работа показывает, что их можно эффективно использовать и в более масштабных системных сетях.

Такая возможность возникает в результате появления в суперкомпьютерах многофункциональных узлов связи с большим числом сетевых каналов. Как следствие возникает задача наиболее эффективного использования этих каналов для решения задачи повышения характеристик способности системной сети. Данная работа показывает, что эту задачу можно решать за счет выбора схемы соединений узлов связи или даже последовательности их соединений.

Существенной особенностью рассмотренного подхода является то, что повышение характеристик системных сетей достигается с незначительным повышением их энергопотребления (*Blue Waters*) или даже без оного (*Gemini*).

Список литературы

- [1] Arimili B., Arimili R., Chung V., et al. *The PERCS High-Performance Interconnect* // 18th IEEE Symposium on High Performance Interconnects. 2009, p. 75–82.
- [2] Alverson R., Roweth D. and Kaplan L., Cray Inc. *The Gemini System Interconnect* // 18th IEEE Symposium on High Performance Interconnects. 2009, p. 83–87.
- [3] Подлазов В. С., Соколов В. В. *Метод однородного расширения системных сетей многопроцессорных вычислительных систем* // Проблемы управления. 2007. № 2, с. 22–27.
- [4] Каравай М. Ф., Пархоменко П. П., Подлазов В. С. *Комбинаторные методы построения двудольных однородных минимальных квазиполных графов (симметричных блок-схем)* // Автоматика и телемеханика. 2009. №. 2, с. 153–170.
- [5] Каравай М. Ф., Подлазов В. С. *Метод инвариантного расширения системных сетей многопроцессорных вычислительных*

- систем. *Идеальная системная сеть* // Автоматика и телемеханика. 2010. № 10, с. 166–176.
- [6] Каравай М. Ф., Подлазов В. С., Соколов В. В. *Метод расширения полных коммутаторов в фиксированном схемном базисе* // Труды 5-й международной конференции «Параллельные вычисления и задачи управления» (РАСО'2010). М. 2010, с. 295–305. URL: <http://paco.ipu.ru/pdf/A205.pdf>
- [7] Каравай М. Ф., Пархоменко П. П., Подлазов В. С. *Универсальная сетевая структура для отказоустойчивых многопроцессорных систем реального времени* // Труды конференции «Технические и программные средства систем управления, контроля и измерения» (УКИ'10). М. 2010, с. 583–597. URL: <http://cmm.ipu.ru/proc/index.html>.
- [8] Каравай М. Ф., Подлазов В. С. *Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных вычислительных систем* // Управление большими системами. Выпуск 34. М.: ИПУ РАН. 2011, с. 92–116. URL: <http://ubs.mtas.ru/upload/library/UBS3405.pdf>.
- [9] Алленов А. В., Подлазов В. С., Стецюра Г. Г. *Пропускная способность набора кольцевых каналов. I. Класс наборов колец. Наборы с простыми узлами* // Автоматика и телемеханика. 1996. № 3, с. 135–144.
- [10] Подлазов В. С., Подлазова А. В. *Обеспечение наращиваемости отказоустойчивых многопроцессорных систем с общей памятью с использованием многокольцевых некоммутируемых сетей связи с неоднородными узлами* // Труды Института проблем управления РАН. 2002. т. XVIII, с. 164–181.
- [11] Подлазов В. С. *Наращиваемые многокольцевые некоммутируемые сети связи для многопроцессорных вычислительных систем* // Проблемы управления. 2006. № 2, с. 50–57.
- [12] Каравай М. Ф., Подлазов В. С. *Расширенные блок-схемы для идеальных системных сетей* // Проблемы управления. 2012. № 4, с. 45–51.
- [13] Riter M. B., Vlasov Y., Kash J. A., and Benner A. *Optical technologies for data communication in large parallel systems* // Topical Workshop on Electronics for Particle Physics (TWEPP-10). 2010. Aachen. Germany. URL: <http://iopscience.iop.org/1748-0221/6/01/C01012>.

- [14] Подлазов В. С., Каравай М. Ф. *Системные сети с прямыми каналами для многопроцессорных вычислительных систем – идеальные системные сети* // Palmarium Academic Publishing. 2012. 168 с.

Рекомендовал к публикации

Программный комитет

Национального суперкомпьютерного форума НСКФ-2012

Об авторах:



Михаил Федорович Каравай

Учреждение Российской академии наук Институт проблем управления (ИПУ) РАН.

Зав. лаб., д.т.н., доцент.

104 печатных работ и 3 монографии.

Техническая диагностика и отказоустойчивость многопроцессорных вычислительных систем реального времени.

e-mail:

(mkaravay@ipu.ru)



Виктор Сергеевич Подлазов

Учреждение Российской академии наук Институт проблем управления (ИПУ) РАН.

Г.н.с., д.т.н., доцент.

101 печатная работа и 2 монографии.

Локальные и системные сети МВС, сетевое обеспечение отказоустойчивости МВС.

e-mail:

(podlazov@ipu.ru)

Образец ссылки на публикацию:

М. Ф. Каравай, В. С. Подлазов. Скрытые резервы системных сетей суперкомпьютеров Blue Waters (IBM) и Gemini (CRAY) // Программные системы: теория и приложения: электрон. научн. журн. 2013. Т. 4, № 3(17), с. 53–70.

URL:

http://psta.psiras.ru/read/psta2013_3_53-70.pdf

Mikhail Karavay, Viktor Podlazov. System area networks latent reserves of supercomputers Blue Waters (IBM) and Gemini (CRAY)

ABSTRACT. A method of enhancing of supercomputer interconnect characteristics based on the features of direct channels system area networks is offered. The interconnect for improvement of supercomputers Blue Waters (IBM) and Gemini (CRAY) is considered.

Key Words and Phrases: massive parallel multiprocessor computer, system area networks, self-routing nonblocking networks, direct channels, distributed full switches, nonswitched multirings.