УДК 004.724.2 + 004.272.43 DOI: http://doi.org/10.25728/pu.2020.5.7

ОПТИМАЛЬНЫЕ ОТКАЗОУСТОЙЧИВЫЕ МНОГОМЕРНЫЕ ТОРЫ НА ОСНОВЕ МАЛОПОРТОВЫХ МАРШРУТИЗАТОРОВ И ХАБОВ

М.Ф. Каравай, В.С. Подлазов

Аннотация. Рассмотрен метод построения оптимальных системных сетей с топологией многомерных торов. Оптимизация выполнена по таким важным функциональным характеристикам сети как число ее абонентов (процессоров) и задержки передачи между ними, задаваемые диаметром сети. Оптимизация осуществлена в элементной базе малопортовых маршрутизаторов и разветвителей дуплексных каналов (хабов) путем применения сетей с топологией квазиполных графов. Оптимизация реализована благодаря инвариантному расширению многомерного тора и дуплексного канала с сохранением таких их маршрутных свойств, как способ маршрутизации и максимальные задержки передачи (диаметр сети). Показано, что оптимизация приводит к увеличению числа абонентов при неизменных задержках и к сокращению задержек при неизменном числе абонентов. Оптимизация сопровождается некоторым усложнением сети по схемным и кабельным затратам. При этом мера усложнения (в разах) оказывается меньше меры совместного улучшения обоих характеристик. Приведены сравнительные характеристики оптимальных торов и торов отечественной системной сети «Ангара». Доказано существенное увеличение числа абонентов и сокращение диаметров оптимальных торов по сравнению с сетью «Ангара».

Ключевые слова: системные сети суперкомпьютеров, сети с топологией многомерных торов, сети с топологией квазиполных графов, инвариантное расширение сетей, число абонентов сети, диаметр сети, оптимизация характеристик сети, сеть «Ангара».

ВВЕДЕНИЕ

В настоящее время используется небольшое число базовых структур системных сетей суперкомпьютеров — сложенная сеть Клоза, толстое дерево, многомерный тор, обобщенный гиперкуб и двухуровневая склейка полных графов. Среди них нет идеальных — выбор той или иной структуры сразу задает и ряд ее ограничений-недостатков. Так, сети с топологией D-мерных торов наименее сложные по схемным и кабельным затратам, но обладают наибольшими задержками передачи данных. Поэтому разработка методов устранения недостатков выбранной структуры в рамках ее базовых возможностей представляет собой актуальную задачу совершенствования современных системных сетей — задачу проектного управления характеристиками сети в целях улучшения (даже оптимизации) функционально важных ее характеристик.

В настоящей работе решается задача построения оптимальных по числу абонентов и задержкам передачи системных сетей с топологией многомерных торов. Она решается путем применения элементной базы, состоящей из малопортовых маршрутизаторов и дуплексных разветвителей (мультиплексоров/демультиплексоров) каналов $1\times m$, $m = 2, 3, 4$. Конкретно, применяются 8-портовый маршрутизатор сети «Ангара» [1, 2] и рыночные дуплексные разветвители 1×3 или 1×4 (хабы) интерфейса *PCI-express*. В различных вариантах их совместного применения имеется возможность увеличения масштабируемости сети (повышения числа процессоров), быстродействия сети (сокращения ее диаметра) и ее канальной отказоустойчивости.

В § 1 описывается метод инвариантного расширения системных сетей с сохранением их маршрутных свойств. Метод конкретизируется для исходных сетей, состоящих из неблокируемых коммутаторов, маршрутизаторов и дуплексных колец. В § 2 строятся распараллеленные дуплексные кольца (разреженные мультикольца) с малыми диамет-

рами. В § 3 строятся расширенные маршрутизаторы с увеличенным числом абонентов. В § 4 на базе разреженных мультиколец и расширенных маршрутизаторов строятся многомерные торы с увеличенным числом абонентов и с малыми диаметрами. Наконец, в § 5 сравниваются характеристики построенных многомерных торов и торов отечественной системной сети «Ангара». В Заключении перечисляются полученные результаты.

1. ИНВАРИАНТНОЕ РАСШИРЕНИЕ ПРОИЗВОЛЬНЫХ СЕТЕЙ

Пусть имеется исходная системная сеть (рис. 1), объединяющая K абонентов, ИсхС(K) с заданными маршрутными свойствами, которые задают число K абонентов сети и ее диаметр D. Для нее можно поставить задачу построения расширенной сети РасС(R) с числом абонентов $R > K$, которая состоит из копий сети ИсхС(K) с разными наборами абонентов и сохраняет неизменными (инвариантом) маршрутные свойства сети ИсхС(K) для абонентов. В данной работе маршрутные свойства — это параметры способов маршрутизации пакетов данных, которые описывают их задержки при передаче (диаметр) и число абонентов сети (процессоров).

Возможны два варианта поставленной задачи. В первом варианте исходной сетью служит маршрутизатор с K дуплексными портами и диаметром D или коммутатор $K\times K$ с диаметром $D = 1$. Во втором варианте исходной сетью служит дуплексное кольцо с K узлами и диаметром $D = \lfloor K/2 \rfloor$.

Будем решать поставленную задачу последовательно, начиная с малых значений $K = m$, $m = 2$, 3, 4. Решение основывается на использовании свойств такого математического объекта, как неполная уравновешенная симметричная блок-схема $B(N, m, \sigma)$ [3, 4], которая содержит N блоков и N элементов, размещенных по блокам так, что каждый блок содержит точно m различных элементов, а каждый элемент входит точно в m различных блоков и каждая пара элементов входит точно в σ блоков, а $N = m(m-1)/\sigma + 1$. При этом блок-схема $B(N, m, \sigma)$ задает *максимальное* N при заданных m и σ, и наоборот, *минимальное* m при заданных N и σ. В решаемой задаче блоки интерпретируются как копии сети ИсхС(m), элементы — как узлы степени m (с m дуплексными портами), а вхождение элемента в блок — как соединение их дуплексным каналом.

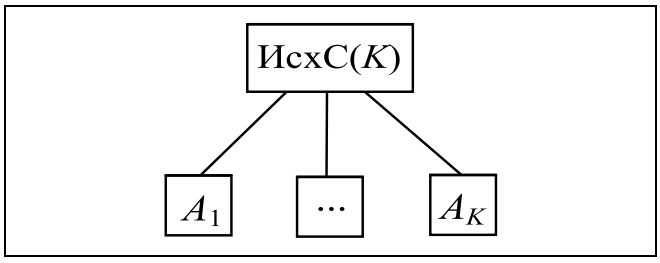

Рис. 1. Исходная произвольная сеть на K абонентов с одним портом у каждого абонента

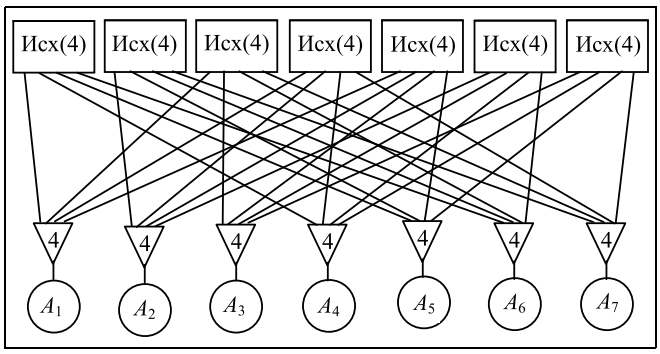

Рис. 2. Сеть ПС(7, 4, 2)

Таблица 1

Межсоединения в сети ПС(7, 4, 2)

Копия сети Исх(4)	Абоненты			
1	1	6	5	4
2	2	7	6	5
3	3	1	7	6
4	4	2	1	7
5	5	3	2	1
6	6	4	3	2
7	7	5	4	3

В такой интерпретации простейшая расширенная сеть, изоморфная блок-схеме $B(N, m, \sigma)$, задается двудольным графом, одна доля которого содержит N копий сети ИсхС(m), другая доля — N узлов (абонентов), соединенных m ребрами (дуплексными каналами) с разными копиями сети ИсхС(m). Между любыми абонентами имеется σ разных путей длины в два ребра, и любой путь проходит только через одну сеть ИсхС(m). Этот граф мы называем квазиполным графом, а изоморфную ему сеть — простейшей сетью ПС(N, m, σ). В ней абонент имеет степень m благодаря подключению к сети ИсхС(m) через разветвитель дуплексных каналов $1\times m$ — хаб(m).

Пример такой сети при $m = 4$ и $\sigma = 2$ дан на рис. 2. Схема соединений между копиями сети ИсхС(4) и абонентами в сети ПС(7, 4, 2) задается в табл. 1, в которой в каждой строке задается номер сети ИсхС(4) и номера подсоединенных к ней абонентов.

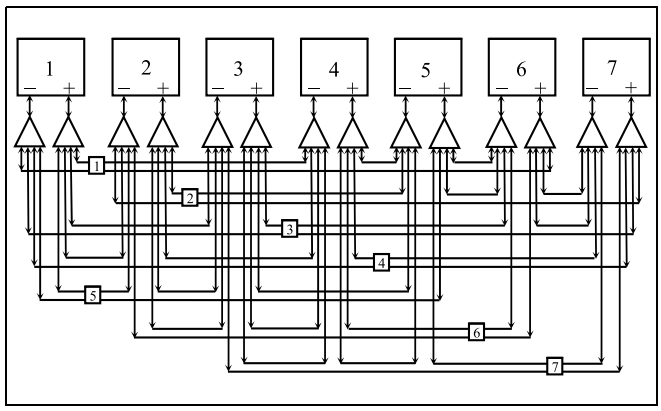

Рис. 3. Разреженное мультикольцо РазМк(7, 4, 2) с диаметром $D = 2$

Таблица 2

Межсоединения в разреженном мультикольце РазМк(7, 4, 2)

Кольца сети ИсхС(4)	Узлы			
1	1	6	5	4
2	2	7	6	5
3	3	1	7	6
4	4	2	1	7
5	5	3	2	1
6	6	4	3	2
7	7	5	4	3

Возможны следующие варианты организации исходной системной сети ИсхС(m).

- Если сеть ИсхС(m) является неблокируемым коммутатором $m\times m$, то сеть ПС(N, m, σ) является неблокируемым коммутатором $N\times N$, составленным из коммутаторов $m\times m$ и хабов(m). При этом сеть РасС(N, m, σ) имеет статическую самомаршрутизацию, при которой абоненты задают бесконфликтные маршруты независимо друг от друга. Эти маршруты прокладываются между абонентами по прямым каналам (без промежуточной буферизации пакетов). Поэтому сеть ПС(N, m, σ) имеет диаметр $D = 1$ и является $(\sigma - 1)$-отказоустойчивой по каналам и коммутаторам.
- Если сеть ИсхС(m) является дуплексным кольцом с m узлами, то сеть ПС(N, m, σ) является мультикольцом с N узлами, составленным из N колец с m узлами и хабов(m) [5]. Такую сеть ПС(N, m, σ) мы называем разреженным мультикольцом РазМк(N, m, σ), которое является $(\sigma - 1)$-отказоустойчивым по каналам.

Пример разреженного мультикольца при $m = 4$ и $\sigma = 2$ приведен на рис. 3. Схема соединений между кольцами сети ИсхС(4) и абонентами в разреженном мультикольце РазМк(7, 4, 2) задается табл. 2, в которой в каждой строке находятся номер кольца сети ИсхС(4) и номера подсоединенных к нему абонентов.

Мультикольцо РазМк(N, m, σ) также имеет статическую самомаршрутизацию, при которой любой путь между узлами проходит по одному кольцу сети ИсхС(m). Как следствие, мультикольцо РазМк(N, m, σ) имеет диаметр $D = \lfloor m/2 \rfloor$, а не $D = \lfloor N/2 \rfloor$, как имеет место в дуплексном кольце из N узлов.

- Наконец, если сеть ИсхС(m) является маршрутизатором с диаметром D, то сеть ПС(N, m, σ) также является маршрутизатором с диаметром $D + 1$, который складывается из скачка по ребру и скачков с входа на выход в сети ИсхС(m).
- В общем случае, когда $K > m$, инвариантное расширение сети ИсхС(K) осуществляется таким образом. Берется N копий сети ИсхС(K), каждая из которых разделяется на $t = \lfloor K/m \rfloor$

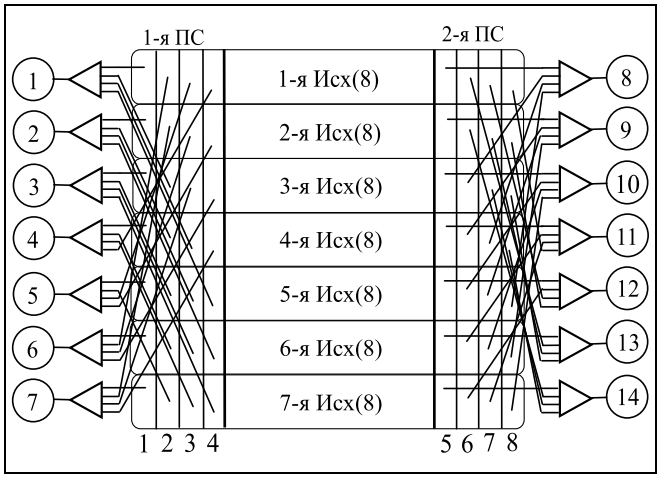

Рис. 4. Сеть РасС(14, 8, 4, 2) с диаметром $D = 2$

Таблица 3

Межсоединения в сети РасС(14, 8, 4, 2)

Копии сети Исх(8)	Порты							
	1	2	3	4	5	6	7	8
	1-я сеть ПС(7, 4, 2)				2-я сеть ПС(7, 4, 2)			
	Абоненты							
1	1	6	5	4	8	13	12	11
2	2	7	6	5	9	14	13	12
3	3	1	7	6	10	8	14	13
4	4	2	1	7	11	9	8	14
5	5	3	2	1	12	10	9	8
6	6	4	3	2	13	11	10	9
7	7	5	4	3	14	12	11	10

равных частей по m портов. К портам i-й части ($1 \le i \le t$) подсоединяются абоненты с номерами $(i-1)N + j$, $1 \le j \le N$, которые принадлежат к i-й сети ПС(N, m, σ). Если $K = tm + m - \sigma$, то $(t+1)$-я часть образует остаточную сеть ПС*($N - m$, $m - \sigma$, σ) с абонентами, номера которых задаются как $tN + j$, $1 \le j \le N - m$. Остаточная сеть ПС*($N - m$, $m - \sigma$, σ) изоморфна остаточной блок-схеме $B^*(N - m, m - \sigma, \sigma)$ [3], где $N - m = (m-1)(m-\sigma)/\sigma$. В результате расширенная сеть содержит либо $R = tN$, либо $R = tN + N - m$ абонентов и обозначается как РасС(R, K, m, σ).

- Если сеть ИсхС(K) имеет диаметр D, то сеть РасС(R, K, m, σ) имеет диаметр $D + 1$. Пример схемы соединений в сети РасС(R, K, m, σ) при $K = 8$, $m = 4$, $\sigma = 2$ и $D = 2$ дан в табл. 3, а самой сеть Рас(14, 8, 4, 2) — на рис. 4. По построению любые два абонента, номера которых не совпадают по mod$N$, соединены друг с другом сетью ИсхС(K) через одну копию сети ИсхС(K), но параллельно через σ разных копий сети ИсхС(K), и используют только маршрутные свойства сети ИсхС(K). С другой стороны, любые два абонента, номера которых совпадают по mod$N$, также соединены друг с другом последовательно через одну копию сети ИсхС(K), но параллельно через m разных копий сети ИсхС(K). Эти свойства обеспечивают сохранение в сети РасС(R) маршрутных свойств сети ИсхС(K). При этом образуются N подмножеств по t или $t + 1$ абонентов, с увеличенной в m раз пропускной способностью между ними.

2. РАЗРЕЖЕННЫЕ МУЛЬТИКОЛЬЦА С МАЛЫМ ДИАМЕТРОМ

Разреженное мультикольцо РазМк(N, m, σ) является кольцом с минимальным диаметром $D = \lfloor m/2 \rfloor$ скачков, которое заменяет кольцо с N узлами и диаметром $D = \lfloor N/2 \rfloor$. Однако в общем случае требуется заменить кольцо с P узлами, где $(p-1)N < P \le pN$, разреженным мультикольцом с меньшим диаметром. Такое мультикольцо обозначается как РазМк(P, N, m, σ) и строится таким образом.

Сначала описанным в § 1 методом строится таблица соединений для мультикольца РазМк(N, m, σ) с m кольцами. Для примера табл. 4 задает таблицу соединений в нем при $m = 2$ и $\sigma = 1$. Само мультикольцо с такой минимальной таблицей соединений представлено на рис. 5.

Затем минимальная таблица расширяется в таблицу соединений для мультикольца РазМк(P, N, m, σ), которая содержит N строк с номерами по pm узлов в каждой строке. Строки расширенной таблицы разделены на p зон по m столбцов в каждой. В зоне 1 находятся N узлов с номерами из минимальной таблицы и с тем же размещением по строкам. В зоне i ($1 < i \le p$) находится узел с номером $L + (i-1)N$ той же строки и в том же месте, что и узел с номером L в зоне 1 ($1 < L \le N$). Для примера табл. 5 задает расширенную таблицу с $p = 3$, $m = 2$ и $\sigma = 1$, полученную из табл. 4. Для примера мультикольцо РазМк(9, 3, 2, 1), построенное по табл. 5, представлено на рис. 6.

Для построения мультикольца РазМк(P, N, m, σ), узлы из каждой строки расширенной таблицы соединяются друг с другом в одноименном кольце в порядке возрастания их номеров. Для замыкания кольца узел с наибольшим номером подсоединяется к узлу с наименьшим номером. Если оказы-

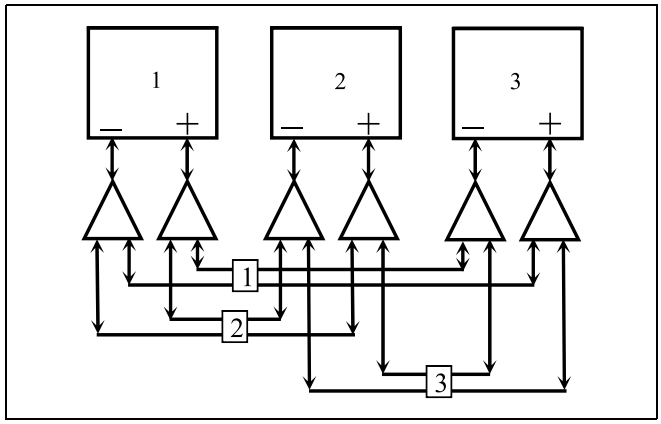

Рис. 5. Разреженное мультикольцо при $p = 1$, $m = 2$ и $\sigma = 1$ с диаметром $D = 1$

Таблица 4

Минимальная таблица подсоединения узлов к кольцам при $m = 2$ и $\sigma = 1$

Кольца ИсхС(2)	Узлы	
1	1	3
2	2	1
3	3	2

Таблица 5

Расширенная таблица подсоединения узлов к кольцам при $p = 3$, $m = 2$ и $\sigma = 1$

i	1		2		3	
Кольца	Узлы 1÷3		Узлы 4÷6		Узлы 7÷9	
1	1	3	4	6	7	9
2	2	1	5	4	8	7
3	3	2	6	5	9	8

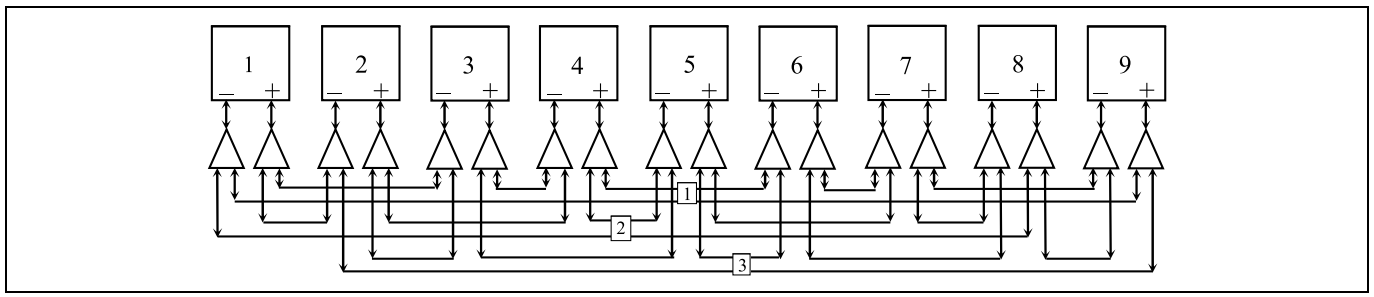

Рис. 6. Разреженное мультикольцо при $p = 3$, $m = 2$ и $\sigma = 1$ с $D = 3$

вается, что $(p - 1)N < P < pN$, то узлы с номерами больше P исключаются из расширенной таблицы и из мультикольца, а его кольца замыкаются через оставшиеся узлы с наибольшими номерами. Диаметр мультикольца РазМк(P, N, m, σ) задается как $D = \lfloor pm/2 \rfloor$, где $p = \lceil P/N \rceil$.

В дальнейшем нам потребуется мультикольцо РазМк(7, 3, 1) с диаметром $D = 1$, таблица соединений которого дана в табл. 6, а само мультикольцо — на рис. 7. На ее основе описанным методом можно построить мультикольцо РазМк(8, 7, 3, 1) с $p = 2$ и $D = 2$, РазМк(14, 7, 3, 1) с $p = 2$ и $D = 2$ или мультикольцо РазМк(16, 7, 3, 1) с $p = 3$ и $D = 3$.

3. ПРОСТЕЙШИЕ РАСШИРЕННЫЕ МАРШРУТИЗАТОРЫ

Рассмотрим 8-портовый маршрутизатор Е4 сети «Ангара» [1, 2] с диаметром $D = 1$. Возьмем его за исходную сеть Исх(8) и расширим до R-портового маршрутизатора РасС(R, 8, m, σ) при разных m и σ. Частично это уже сделано ранее для $m = 4$ и $\sigma = 2$ (см. табл. 3 и рис. 4).

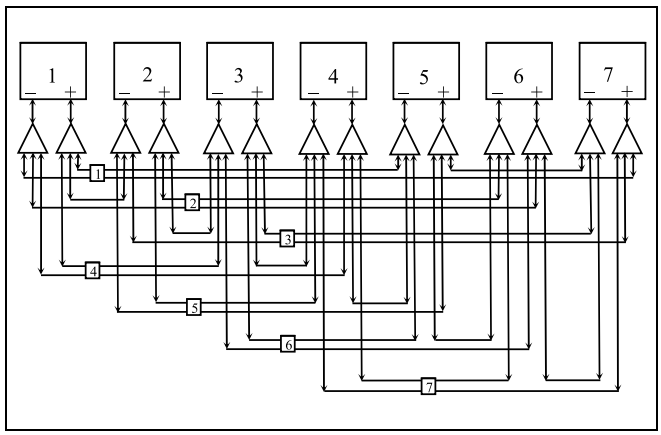

Рис. 7. Разреженное мультикольцо при $p = 1$, $m = 3$ и $\sigma = 1$ с $D = 1$

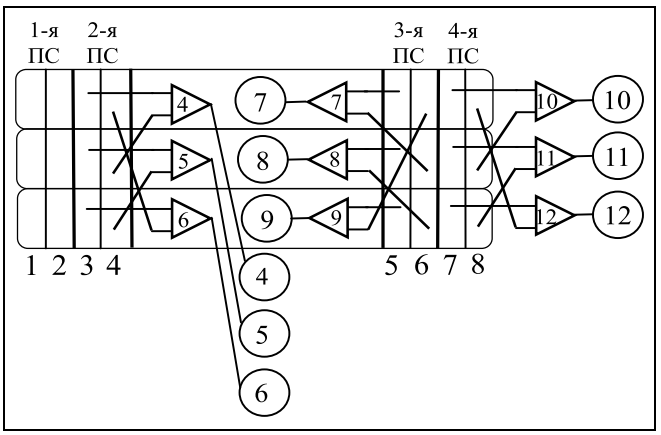

Рис. 8. Маршрутизатор РасС(12, 8, 2, 1) с диаметром $D = 2$

Таблица 6

Таблица подсоединения узлов к кольцам при $p = 1$, $m = 3$ и $\sigma = 1$

Кольца ИсхС(3)	Узлы		
1	1	7	5
2	2	1	6
3	3	2	7
4	4	3	1
5	5	4	2
6	6	5	3
7	7	6	4

Таблица 7

Межсоединения в маршрутизаторе РасС(12, 8,2, 1)

Копии сети Исх(8)	Порты сети Исх(8)							
	1	2	3	4	5	6	7	8
	1-я ПС(3, 2, 1)		2-я ПС(3, 2, 1)		3-я ПС(3, 2, 1)		4-я ПС(3, 2, 1)	
	Абоненты							
1	1	3	4	6	7	9	10	12
2	2	1	5	4	8	7	11	10
3	3	2	6	5	9	8	12	11

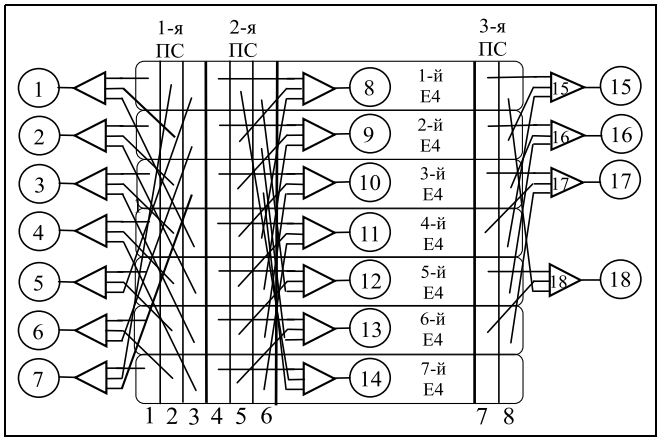

Рис. 9. Маршрутизатор РасС(18, 8, 3, 1) с диаметром $D = 2$

Таблица 8

Межсоединения в маршрутизаторе РасС(18, 8, 3, 1)

Копии сети ИсхС(8)	Порты сети ИсхС(8)							
	1	2	3	4	5	6	7	8
	1-я ПС			2-я ПС			3-я ПС	
1	**1**	7	5	**8**	14	12	**15**	18
2	2	**1**	6	9	**8**	13	16	**15**
3	3	2	7	10	9	14	17	16
4	4	3	**1**	11	10	**8**	17	**15**
5	5	4	2	12	11	9	18	16
6	6	5	3	13	12	10	18	17
7	7	6	4	14	13	11	—	—

В минимальном варианте $m = 2$ и $\sigma = 1$. Тогда схема соединений для маршрутизатора РасС(12, 8, 2, 1) дана в табл. 7, а сам маршрутизатор — на рис. 8.

Аналогично и при $m = 3$ и $\sigma = 1$. Оказывается, что в этом случае $R = 18$, как это видно из схемы соединений для РасС(18, 8, 3, 1), задаваемой табл. 8. Сам маршрутизатор с диаметром $D + 1$ представлен на рис. 9.

4. МНОГОМЕРНЫЕ ТОРЫ С МАЛЫМИ ДИАМЕТРАМИ

Расширенные маршрутизаторы (см. рис. 4, 7, 8) позволяют создавать r-мерные ($r = 1, 2, 3, 4$) торы. Для соединения маршрутизаторов в кольцах разных измерений достаточно отсоединить от каждого маршрутизатора по два абонента для каждого измерения. Освободившиеся дуплексные порты необходимо использовать для подсоединения к соседним маршрутизаторам и в направлении «+» и «−» в каждом измерении.

Простейший случай задает применение маршрутизаторов РасС(12, 8, 2, 1) и мультиколец РазМк(P, 3, 2, 1) (рис. 10). Характеристики создаваемых торов даны в табл. 9, в которой P задает число маршрутизаторов в кольце каждого измерения, а M — общее число абонентов в торе, где $M = P^r(12 - 2r)$. Диаметр D тора складывается из двух скачков между маршрутизатором и абонентами (источниками и приемником), скачков между маршрутизаторами по кольцам каждого проходимого измерения и двух скачков внутри маршрутизатора при смене измерения и $D = r\lfloor pm/2 \rfloor + 2(r - 1) + 1$, где $p = \lceil P/N \rceil$.

Дальнейшая оптимизация важных характеристик многомерных торов возможна, если их строить из маршрутизаторов РасС(18, 8, 3, 1) по рис. 9 и заменять в них кольца каждого измерения на муль-

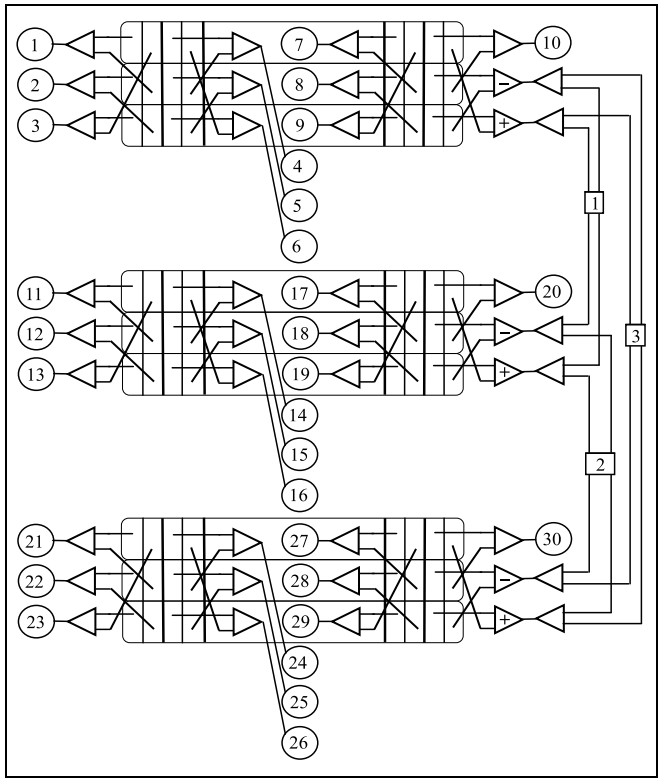

Рис. 10. Одномерный тор на базе маршрутизаторов РасС(12, 8, 2, 1) и мультиколец РазМк(P, 3, 2, 1) с диаметром $D = 3$

Таблица 9

Торы из маршрутизаторов РасС(12, 8, 2, 1) и мультиколец РазМк(P, 3, 2, 1)

P	3		6		9		12	
r	M	D	M	D	M	D	M	D
1	30	3	60	4	90	5	120	6
2	72	6	288	8	648	10	1152	12
3	162	9	1296	12	4374	15	10 368	18

Таблица 10

Торы из маршрутизаторов РасС(18, 8, 3, 1) и мультиколец РазМк(*P*, 7, 3, 1)

P	7		14	
r	*M*	*D*	*M*	*D*
1	112	3	224	4
2	686	6	2744	8
3	4116	9	32 928	12
4	24 010	11	384 160	16

Таблица 11

Торы из маршрутизаторов РасС(14, 8, 4, 2) и мультиколец РазМк(7*p*, 7, 4, 2)

P	7 ($p = 1$)		14 ($p = 2$)	
r	*M*	*D*	*M*	*D*
1	84	4	168	6
2	490	8	1960	12
3	2744	12	21 952	18
4	14 406	16	230 496	24

тикольца, построенные в § 2. В этом случае характеристики создаваемых торов даны в табл. 10. Значения для табл. 10 рассчитываются по формулам: $M = P^r(18 - 2r)$. В данном торе маршруты одинаковой длины в каждом измерении расщепляются по трем разным путям, что существенно снижает средние задержки передачи.

Несколько более слабые характеристики имеет 1-отказоустойчивые торы, которые строятся из маршрутизаторов РасС(14, 8, 4, 2) по рис. 4 и мультиколец РазМк(P, 7, 4, 2) в каждом измерении. Характеристики этих торов задаются в табл. 11. Значения для табл. 11 рассчитываются по формулам: $M = P^r(14 - 2r)$. Однако в данном торе маршруты одинаковой длины в каждом измерении расщепляются по четырем разным путям, что дополнительно снижает средние задержки передачи.

5. СРАВНЕНИЕ С СЕТЬЮ «АНГАРА»

Системные сети современных суперкомпьютеров строятся на базе многопортовых маршрутизаторов, например, единый однокристальный 48-портовый маршрутизатор *YARC* для 3-мерного тора *Gemini* и для 4-мерного обобщенного гиперкуба *Dragonfly* фирмы *CRAY* [6, 7]. При этом имеется тенденция перехода на топологию гиперкуба как сети с меньшим диаметром и большим быстродействием, но и большей сложности.

В РФ в настоящее время нет таких маршрутизаторов. Имеется функционально полный однокристальный маршрутизатор Е4 сети «Ангара» с 8 сетевыми дуплексными портами *PCI-express* и одним процессорным портом *PCI-express* [1, 2]. Первоначально предполагалось его использование в одноплатном варианте для построения системной сети в виде 4-мерного тора на $M = 4K$ процессоров и с диаметром в $D = 16$ скачков (8 маршрутизаторов в кольцах разных измерений). Потом было заявлено о возможности иметь на нем $M = 16K \div 32K$ с диаметрами $D = 24 \div 32$ скачков (до 16 маршрутизаторов в кольцах). Одноплатный вариант оказался не очень экономичным для построения малых сетей, так как расходовал один маршрутизатор для подсоединения к сети одного процессора.

Затем был создан однокорпусной 24-портовый маршрутизатор [1, 2, 8] путем сцепления четырех 8-портовых маршрутизаторов. Использование этого маршрутизатора резко упрощает построение сетей самого разного размера — от десятков процессоров до нескольких тысяч (в топологии одномерного или двумерного тора). В таком виде сеть «Ангара» считается базовой сетью для построения отечественных суперкомпьютеров.

Заметим, что 24-портовый маршрутизатор имеет внутренний диаметр в четыре скачка: один скачок от входного порта до соединительного порта в 8-портовом маршрутизаторе, два скачка между 8-портовыми маршрутизаторами и один скачок от соединительного порта до выходного порта. Одна-

Таблица 12

Сравнительные характеристики торов

	РасС(18, 8, 3, 1)								Ангара с 24-портовым маршрутизатором			
P	7		8		14		16		8		16	
r	*M*	*D*	*M*	*D*	*M*	*D*	*M*	*D*	*M*	*D*	*M*	*D*
1	112	3	128	4	224	4	256	5	128	8	256	16
2	686	6	896	8	2 744	8	3584	10	502	13	2048	21
3	4116	9	6144	12	32 928	12	49 152	15	—	—	—	—
4	24 010	11	40 960	16	384 160	16	653 600	20	—	—	—	—

ко в сети для связи между 24-портовыми маршрутизаторами используются четыре дуплексных канала между заданными 8-портовыми маршрутизаторами. Это делает проходную задержку равной одному скачку.

В табл. 12 (справа) представлены характеристики сети «Ангара». При этом максимальное число процессоров опять достигается при размещении 16 маршрутизаторов в кольцах. Однако использование больше 8 маршрутизаторов в кольце не увеличивает его пропускную способность, но увеличивает задержки передачи по нему. Правда, наличие четырех дуплексных колец в каждом измерении в значительной мере снижает эти задержки.

Заметим, что 3-мерный тор на базе 24-портового коммутатора уже не может быть создан из-за недостаточного числа портов, что делает невозможным дальнейшее увеличение числа процессоров в сети «Ангара» без увеличения числа узлов в кольцах и задержек передачи по ним.

Также в табл. 12 (слева) приведены характеристики торов, составленных из маршрутизаторов РасС(18, 8, 3, 1) и разреженных мультиколец РазМк(P, 7, 3, 1).

Видно, что построенный выше одномерный тор имеет в несколько раз меньший диаметр, чем одномерный тор сети «Ангара» при близком числе абонентов. Построенный двумерный тор имеет в два раза меньший диаметр, чем двумерный тор сети «Ангара» при несколько большем числе абонентов. В остальных случаях построенные торы имеют в несколько раз меньший диаметр и существенно большее число абонентов, чем может обеспечить сеть «Ангара» в любом варианте.

Построенные выше многомерные торы мы считаем оптимальными, так как они построены на базе оптимального распараллеливания сетевой структуры на основе квазиполного орграфа. Она позволяет строить расширенные маршрутизаторы с максимальным числом абонентов при заданных исходных маршрутизаторах. И наоборот, она позволяет иметь минимальный диаметр сети благодаря применению разреженных мультиколец.

Подобное распараллеливание структур сети сопровождается, конечно, увеличением ее аппаратных и кабельных затрат. Оценим их.

Примем, что сложность маршрутизатора пропорциональна квадрату числа портов. Тогда сложность s_K однокорпусного маршрутизатора «Ангара» составляет $s_K = 64c$, где c — коэффициент пропорциональности. Сложность s_X одного хаба 1×3 можно оценить как $s_X = 6c$ (мультиплексор + демультиплексор).

Любой одномерный тор является дуплексным кольцом, к узлам которого подсоединены абоненты (процессоры). Поэтому сложность одномерного тора «Ангара» с 8 корпусами в кольце задается как $S_{A,1} = 2\,048c$.

Любой двумерный тор является сетью, через узлы которой проходят дуплексные кольца разных измерений, что обеспечивает наличия в нем квадратичного числа узлов и абонентов. Поэтому сложность двумерного тора «Ангара» с 8 корпусами в кольце каждого измерения задается как $S_{A,2} = 16\,384c$. Они содержат $M_{A,1} = 128$ и $M_{A,2} = 502$ абонентов соответственно. Поэтому их удельная сложность составляет $s_{A,1} = S_{A,1}/M_{\mathbf{A},1} = 16c$ и $s_{A,2} = S_{A,2}/M_{A,2} = 32c$.

Маршрутизатор РасС(18, 8, 3, 1) содержит 7 маршрутизаторов Е4 сложности s_K и 18 хабов 1×3 сложности s_X. В результате сложность S_P расширенного маршрутизатора составляет $S_P = 7\times64c + 18\times6c = 556c$.

Одномерный тор содержит 7 таких маршрутизаторов и еще 14 хабов для образования разреженных мультиколец общей сложности $S_{P,1} = 3976c$. Он содержит $M_{P,1} = 112$ абонентов. Поэтому его удельная сложность составляет $s_{P,1} = 35{,}5c$.

Двумерный тор содержит 49 таких маршрутизаторов и еще 28 хабов в каждом одномерном измерении для образования разреженных мультиколец общей сложности $S_{P,2} = 28\,420c$. Он содержит $M_{P,2} = 686$ абонентов. Поэтому его удельная сложность составляет $s_{P,2} = 41{,}4c$.

Введем комплексную характеристику торов $\aleph$ как произведение диаметра на удельную сложность. Тогда $\aleph_{A,1} = 128c$ и $\aleph_{P,1} = 106{,}5c$, аналогично $\aleph_{A,2} = 416c$ и $\aleph_{P,2} = 250{,}2c$. Отсюда можно сделать вывод, что повышенная сложность торов из расширенных маршрутизаторов с избытком обеспечивает их малые диаметры. При этом одновременно обеспечивается и большее число абонентов. Однако маршрутизаторы с разреженными мультикольцами имеют в 7/4 раза больший расход кабеля.

ЗАКЛЮЧЕНИЕ

Рассмотрен метод построения оптимальных системных сетей с топологией многомерных торов.

Оптимизация в работе осуществляется по таким важным функциональным характеристикам сети, как число ее абонентов и задержки передачи, задаваемые диаметром сети. Оптимизация осуществляется в элементной базе малопортовых маршрутизаторов и хабов путем построения или применения сетей с топологией квазиполных графов.

Применяется метод инвариантного по маршрутным свойствам расширения сетей для увеличения в них числа абонентов и уменьшения их диаметра в заданной элементной базе.

Показана возможность применения рассмотренного метода для повышения масштабируемости и быстродействия отечественной системной сети «Ангара».

Совместная удельная сложность по числу абонентов и задержкам передачи в оптимизированной сети оказалась меньше, чем у сети «Ангара».

ЛИТЕРАТУРА

1. *Симонов А.С., Макагон Д.В., Жабин И.А.* и др. Первое поколение высокоскоростной коммуникационной сети «Ангара» // Наукоемкие технологии. — 2014. — Т. 15, № 1. — С. 21—28. [*Simonov, A.S., Makagon, D.V., Zhabin, I.A.,* et al. Pervoe pokolenie vysokoskorostnoi kommunikatsionnoi seti «Angara» // Naukoemkie tekhnologii. — 2014. — Vol. 15, no. 1. — P. 21—28. (In Russian)]
2. *Stegailov, V., Agarkov, A., Biryukov, S.,* et al. Early Performance Evaluation of the Hybrid Cluster with Torus Interconnect Aimed at Molecular Dynamics Simulations // International Conference on Parallel Processing and Applied Mathematics. — Springer. — Cham. — 2017. — P. 327—336.
3. *Холл М.* Комбинаторика. Главы 10—12. — Мир. М. — 1970. — 424 с. [*Hall, M.* Combinatorial Theory. — Waltham: Blaisdell Publishing Company, 1967. (In Russian)]
4. *Каравай М.Ф., Подлазов В.С.* Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть // Автоматика и телемеханика. — 2010. — № 12. — С. 166—176. [*Karavay, M.F., Podlazov, V.S.* An Invariant Extension Method for System Area Networks of Multicore Computational Systems. An Ideal System Network // Automation Remote Control. — 2010. — Vol. 71, no. 12. — P. 2644—2654.]
5. *Подлазов В.С.* Повышение характеристик многомерных торов // Управление большими системами. — 2014. — Вып. 51. — С. 60—81. [*Podlazov, V.S.* Boosting Performance of Multidimensional Tores // Automation and Remote Control. — 2017. — Vol. 78, no. 1. — P. 167—179.]
6. *Alverson, R., Roweth, D. and Kaplan, L.* The Gemini System Interconnect // 18th IEEE Symposium on High Performance Interconnects. — 2009. — P. 3—87.
7. *Alverson, R., Froese, E., Kaplan, L. and Roweth, D.* Cray XC® Series Network. — URL: https://www.cray.com/sites/default/files/resources/CrayXCNetwork.pdf.
8. *Каравай М.Ф., Подлазов В.С.* Расширение возможностей системной сети «Ангара» // Проблемы управления. — 2020. — № 2. — С. 47—56. [*Karavai, M.F., Podlazov, V.S.* Expanding the capabilities of the Angara system area network // Control Sciences. — 2020. — No. 2. — P. 47—56. (In Russian)]

Статья представлена к публикации членом редколлегии В.М. Вишневским.

Поступила в редакцию 16.04.2020, после доработки 10.06.2020. Принята к публикации 18.06.2020.

Каравай Михаил Федорович — д-р техн. наук,
✉ mkaravay@ipu.ru,

Подлазов Виктор Сергеевич — д-р техн. наук,
✉ podlazov@ipu.ru,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.

OPTIMUM MULTIDIMENSIONAL TORI BASED ON LOW-PORT ROUTERS AND HUBS

M.F. Karavay[1], V.S. Podlazov[2]

V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

[1]✉ mkaravay@ipu.ru, [2]✉ podlazov@ipu.ru

Abstract. A method for constructing optimal system networks with the topology of multidimensional tori is considered. The optimization was performed according to such important functional characteristics of the network as the number of its subscribers (processors) and transmission delays between them, set by the network diameter. Optimization was carried out in the element base of low-port routers and splitters of duplex channels (hubs) by using networks with the topology of quasi-complete graphs. Optimization is realized due to the invariant expansion of the multidimensional torus and the duplex channel with preservation of their route properties such as the routing method and maximum transmission delays (network diameter). It is shown that optimization leads to an increase in the number of subscribers with constant delays and to a reduction in delays with a constant number of subscribers. Optimization is accompanied by some complication of the network in terms of circuit and cable costs. In this case, the measure of complication (expressed in the number of times) is less than the measure of joint improvement of both characteristics. The comparative characteristics of the optimal tori and tori of the Angara domestic system network are given. A substantial increase in the number of subscribers and a decrease in the diameters of optimal tori in comparison with the Angara network have been proven.

Keywords: system-area networks of supercomputers, networks with the topology of multidimensional tori, networks with the topology of quasi-complete graphs, invariant expansion of networks, number of network subscribers and network diameter, optimization of network characteristics, Angara network.