УДК 681.321

М.Ф. КАРАВАЙ, В.С. ПОДЛАЗОВ

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

К ВОПРОСУ ПОСТРОЕНИЯ РАСШИРЕННОГО ПОЛНОГО КОММУТАТОРА – ИДЕАЛЬНОЙ СИСТЕМНОЙ СЕТИ

Предлагается новый способ построения большого полного коммутатора из малых полных коммутаторов, основанный на использовании неполных уравновешенных блок-схем, исследуемых в комбинаторике.

Ключевые слова: *системная сеть, полный коммутатор, идеальная сеть, симметричное расширение, отказоустойчивость, алгоритм маршрутизации.*

Предисловие

Сетевые технологии, характерные для коммутационных структур на плате и в мейнфрейм (mainframe) системах начинают глубоко проникать на уровень кристаллов [1,2,3], сохраняя в них те же основные структуры и протоколы взаимодействия. В современных многопроцессорных системах пакетная передача данных на уровне кристаллов оказывается более удобной, надёжной и производительной, чем обмен словами на уровне Процессор – Процессор или Процессор – Память. Естественно стремление найти идеальную структуру сети [4], подходящую как для информационного обмена внутри кристаллов, так и для системных сетей (СС) вне кристаллов. Необходимость этого поиска вызвана не только повышением производительности вновь разрабатываемых компьютерных систем, но и упрощением решения многих задач контролепригодного проектирования, в частности улучшения управляемости и наблюдаемости системы, тестирования и диагностирования отказов, введения резервирования по связям и абонентам. Известно, что реконфигурируемость системы и её отказоустойчивость оказываются напрямую связаны со структурой сети [5]. До сих пор ближайшими претендентами на звание "идеальной" сети были (мульти)кольца, гиперкубы, 2D и 3D решетки, перестраиваемые сети Клоза [6,7]. Но такие недостатки этих сетей как сложность нахождения бесконфликтной маршрутизации, наличие тупиков и связанная с этим необходимость большого числа буферов данных и виртуальных каналов для их разрешения, ставят под сомнение их способность считаться идеальными коммутационными структурами. Масштабирование этих сетей не сохраняет важных сетевых инвариантов (алгоритмов маршрутизации и задержек передачи) и симметричность доступа абонентов к ресурсам сети. В этом смысле "идеальной" является неблокируемая СС, в которой имеется возможность бесконфликтной реализации произвольной перестановки элементов данных (пакетов) между абонентами СС. Основными маршрутными свойствами идеальной СС являются неблокируемость и самомаршрутизируемость. Неблокируемость означает возможность бесконфликтно осуществлять произвольную перестановку пакетов данных между абонентами при параллельной передаче пакетов от всех абонентов, а самомаршрутизируемость – возможность прокладки маршрута при перестановке пакетов каждым абонентом самостоятельно независимо от других абонентов.

В частности, идеальной является СС, которая имеет структуру полного графа или, наоборот, состоит из одного полного коммутатора. Обе таких структуры непригодны для создания СС с большим количеством абонентов – первая вследствие большого числа портов абонентов и связей между ними, а вторая из-за невозможности создания коммутаторных чипов с необходимым числом портов. Перестраиваемая сеть Клоза на такое же число абонентов имеет приемлемые характеристики по сложности сети, но не обладает важнейшими логическими свойствами – неблокируемости и самомаршрутизируемости.

Рассматриваемый далее метод расширения системных сетей позволяет в значительной мере преодолеть перечисленные недостатки.

Введение

В настоящее время за сетями связи многопроцессорных вычислительных систем (МВС) утвердился термин системные сети (*System Area Network – SAN*) [8]. Идеальной системной сетью (СС) часто [4] считается та, которая обеспечивает прямые каналь-

ные соединения (без промежуточной буферизации данных) для любой пары абонентов сети при параллельной передаче пакетов данных от всех абонентов. В этом смысле идеальной является неблокируемая СС, в которой имеется возможность бесконфликтной реализации произвольной перестановки элементов данных (пакетов) между абонентами СС. В частности, в логическом смысле, идеальной является СС, которая имеет структуру полного графа или, наоборот, состоит из одного полного коммутатора.

Здесь рассматривается метод расширения любых СС с сохранением алгоритма маршрутизации. Он основывается на описании СС в терминах неполных уравновешенных блок-схем, давно уже исследуемых в комбинаторике [9,10]. Этот метод позволяет увеличивать число абонентов СС с одновременным резервированием всех каналов связи. При этом если в качестве исходной СС берется малый полный коммутатор, то в расширенном (большом) коммутаторе сохраняется возможность бесконфликтной независимой маршрутизации (самомаршрутизации) для каждой пары абонентов источник-приемник.

Задачу расширения СС решает и перестраиваемая сеть Клоза, В частности t-каскадная сеть (t=2k+1) расширяет полный коммутатор с m портами до сети с $R_t=m^{k+1}$ портами. Перестраиваемая сеть Клоза имеет отдельное бесконфликтное расписание для любой перестановки данных между входными и выходными портами, которое в общем случае неизвестно и требует выполнения отдельной длительной процедуры его построения. На практике [11,12] используется маршрутизация методом червоточины, которая не исключает возникновения конфликтов и, как следствие потерю пропускной способности и увеличение задержек передачи. Таким образом, перестраиваемые сети Клоза не являются неблокируемыми при маршрутизации от абонентов.

1. Инвариантное симметричное расширение полных коммутаторов

Термин “симметричное” здесь понимается как эквивалентность между собой абонентов СС.

В данном разделе рассматривается метод решения следующей задачи. Пусть имеется неблокируемая самомаршрутизируемая Системная Сеть на K абонентов (рис. 2), рассматриваемая как атомарная, т.е. неделимая. Эта СС может быть полным коммутатором на K портов. Необходимо расширить эту сеть до $R>K$ абонентов, сохранив как инвариант в расширенной сети саму ИС(K) и ее свойства неблокируемости и самомаршрутизируемости.

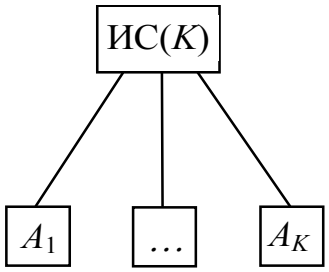

Рис. 1. Произвольная исходная системная сеть на K абонентов.

Будем называть СС на K абонентов *и*сходной *с*етью ИС(K), а создаваемую новую сеть на R абонентов – *р*асширенной *с*етью РС(R).

Сначала поставленная задача решается для малых K=m следующим образом. Предполагается, что для расширения СС можно использовать несколько исходных сетей ИС(m). Увеличивается число портов ввода/вывода (ВВ) **абонента** до m, что может быть реализовано подключением к его порту расширителя ВВ $1\times m$ – Рm. Возникает вопрос, можно ли в этих условиях построить расширенную сеть, сохранив все требования симметрии и атомарности, и так, чтобы расширение сети не вносило дополнительной новой избыточности, помимо существовавшей в исходной сети? Оказывается, что ответ здесь точно такой же, как и при синтезе уравновешенных симметричных блок-схем. При увеличении числа абонентов до R=N дополнительной избыточности не появляется только для соотношения числа копий ИС(K), нового числа абонентов K и портовости абонентов m, которые удовлетворяют условию существования уравновешенных симметричных блок-схем. Поэтому, если взять N=m(m–1)+1 копий ИС(K) и подсоединить к ним R=N абонентов так, что каждый абонент соединяется с любым другим абонентом *последовательно* только через одну копию ИС(K), и любая пара абонентов соединяется через единственную ИС(K), то упомянутые условия существования выполняются. Таким образом, безызбыточное расширение ИС(K) достигается только на верхней границе соответствующего диапазона существования блок-схем. Как в этих условиях построить расширенную СС, так чтобы сохранился алгоритм маршрутизации исходной СС? Оказывается, что для этого достаточно подсоединять абонентов к разным ИС(m) так, чтобы каждый абонент соединялся дуплексным каналом с любым другим абонентом *последовательно* только через одну ИС(m). При этом требуется использовать минимальное число N ИС(m) и подсоединить к расширенной СС максимальное число R абонентов.

Примеры расширенных СС, построенных описанным методом, приведены на рис. 2 и 3.

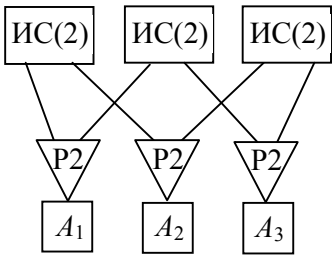

Рис. 2. РС(3) состоит из 3-х ИС(2).

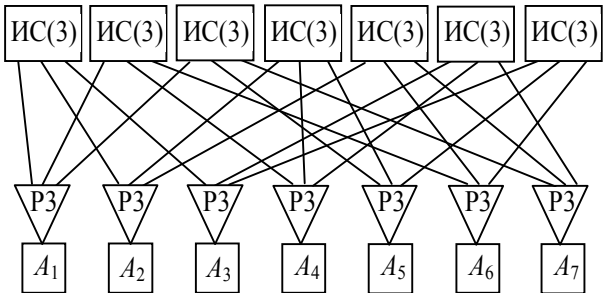

Рис. 3. РС(7) состоит из 7-и ИС(3).

Поскольку в расширенной СС любая пара абонентов последовательно только через одну исходную СС, то первая наследует неблокируемость и самомаршрутизируемость последней. Для прокладки маршрута в расширенной СС достаточно проложить путь от абонента-источника необходимой до ИС(m), выполнить маршрутизацию внутри ИС(m) и проложить путь от ИС(m) до абонента-приемника. При такой маршрутизации разветвитель Рm должен выполнять две функции. На выходе из абонента он должен направлять пакеты в необходимую ИС(m). При этом разветвитель может сам выбирать необходимую ИС(m) (маршрутная адресация), а может предавать эту функцию ИС(m (сетевая адресация). На выходе из ИС(m) разветвитель должен выполнять функцию множественного доступа в единственный выходной канал.

При маршрутной адресации кадр должен содержать номер выходного канала разветвителя. В этом случае функцию прокладки пути выполняет демультиплексор, управляемый номером выходного канала. При сетевой адресации разветвитель выполняет только функцию копирования кадра во все выходные каналы и в предельном случае сводится к мощному выходному элементу. В этом случае каждая ИС(m) должна отвергать кадры, адресованные другим ИС(m). В этом же режиме выполняются multicast and broadcast transfers.

Однако здесь возникает вопрос о существовании описанных схем соединений, и при положительном ответе, какими формулами задаются числа R абонентов и N исходных СС. Как оказывается, эти схемы являются частным случаем неполных уравновешенных блок-схем [9], а именно симметричными блок-схемами. Для них $N=m(m-1)+1$ и $R=N$. Общее определение блок-схемы дается ниже.

2. Квазиполные графы как неполные уравновешенные блок-схемы

Определение 1. Неполной уравновешенной блок-схемой $B(N, M, m, k, \sigma)$ называется совокупность из M блоков, в которые входят N разных элементов так, что в каждый блок входит точно k разных элементов, каждый элемент имеет вхождение точно в m блоков, а каждая пара элементов входит точно в σ блоков.

В этом определении блоки могут интерпретироваться как *исходные СС*, элементы - как *абоненты,* вхождение *разных* элементов в блок - как дуплексное подсоединение этих абонентов к соответствующей исходной СС*,* вхождение каждого элемента в *разные* блоков – как дуплексное подсоединение каждого абонента к *разным* исходным *СС,* а вхождение *пары элементов в разные блоки* интерпретируется как число различных маршрутов между любыми данными абонентами через разные исходные СС. При таком определении блок-схемы ее параметры связаны следующими соотношениями:

$$Nm=kM \text{ и } N=m(k-1)/\sigma+1. \qquad (1)$$

Обратим внимание, что блок-схемы определены с точностью до перестановки элементов в блоках и/или самих блоков. Поэтому блок-схемы, различающиеся порядком вхождения элементов в блоках и/или порядком блоков являются эквивалентными.

Математическая теория блок-схем построена для случая $k \le m$, $N \le M$. Минимальное количество блоков M и максимальное количество элементов N обеспечивается в случае симметричных блок-схем, когда $k=m$ и $M=N$. Такие блок-схемы $B'(N, m, \sigma)$ в основном и рассматриваются далее.

Теперь заменим в определении 1 блок на некоторую ИС(k), элемент – на абонента с m портами, вхождение элемента в блок – на дуплексное подсоединение абонента через один порт к некоторой ИС(k) и, наконец, блок-схему – на расширенную сеть РС(N). Тогда последняя состоит из M ИС(k), к каждой ИС(k) одсоединено k разных абонентов, каждый абонент подсоединен к m разными копиям ИС(k), а между каждой парой абонентов имеется σ параллельных соединений, которые проходят через разные копии ИС(k). При этом каждый путь *последовательно* проходит только через одну копию ИС(k). Поэтому РС(N) наследует неблокируемость и самомаршрутизируемость ИС(k). При этом добавляется (σ–1)-отказоустойчивость по связям. Такую расширенную сеть будем обозначать как РС(N, M, m, k, σ).

Таблица 1

Симметричная блок-схема при m=2 и σ=1

Блок-схема B(3, 2, 1)		
Номера блоков	Номера элементов в блоках	
1	1	2
2	1	3
3	2	3

Таблица 2

Симметричная блок-схема при m=3 и σ=1

Блок-схема B(7, 3, 1)			
Номера блоков	Номера элементов в блоках		
1	1	2	3
2	1	4	6
3	1	5	7
4	2	4	5
5	2	6	7
6	3	4	7
7	3	5	6

Нетрудно видеть, что ПС(3, 2, 1) и ПС(7, 3, 1) соответствуют РС(3) и РС(7), представленным на рис. 2 и 3.

В табл. 3 приводятся значения N для симметричных блок-схем при малых m для σ=1. Напомним, что *значение* σ характеризует число каналов связи между любой парой абонентов.

Заметим, что блок-схема B(43 7, 1) не существует по теории, а блок-схема B(111 11, 1) до сих пор не построена.

Таблица 3

Параметры N и m для B(N, m, 1)

m	2	3	4	5	6	7
N	3	7	13	21	31	43–
m	8	9	10	11	12	m
N	57	73	91	111?	133	N

Аналогично табл. 4 задает параметры возможных блок-схем для σ=2 и σ=3.

Общую таблицу возможных и построенных блок схем для m<22 и σ<10 можно найти в [13].

Таблица 4

Параметры N и m
возможных блок-схем B(N, m, 2) и B(N, m, 3)

B(N, m, 2)							
m	2	3	4	5	6	9	11
N	2	4	7	11	16	37	56 ?
B(N, m, 3).							
m	3	4	6	7	9	10	12
N	3	5	11	15	25	31	45 ?

3. Инвариантное несимметричное расширение полных коммутаторов

До сих пор, при симметричном расширении, число портов K в исходной СС (совпадающее с числом абонентов, подсоединённых к ИС) было ограничено соотношением K=k=m. Это следовало из предположения об эквивалентности абонентов. Далее рассматривается возможность несимметричного расширения с использованием большего числа абонентов.

Предположим, что в ИС(K) K=hm. В этом случае расширенная СС будет обозначаться как РС(R, N, m, K, σ), где R задает общее число абонентов, N – число копий исходной СС, m – число портов абонента, K – число портов исходной СС ИС(K) и σ – число параллельных соединений в расширенной сети любой пары абонентов через разные ИС(K). Надо понимать, что расширенная СС в этом случае уже не будет в математическом смысле симметричной уравновешенной *блок-схемой*.

Расширенная СС РС(R, N, m, K, σ) складывается из h копий ПС(N, m, σ). Каждая копия ИС(K) делится на h равных частей по m портов. Совокупность одноименных частей по всем копиям ИС(K) образует копию ПС(N, m, σ). К 1-ой копии ПС(N, m, σ) подсоединяются те же абоненты с номерами от 1-го до N-го и по той же схеме, что и к отдельной ПС(N, m, σ). В j-ой копии ПС(N, m, σ) номера подсоединенных абонентов увеличиваются по сравнению с 1-ой копией на jN без изменения мест их подсоединения. Поэтому такая расширенная СС объединяет R = hN абонентов.

Таблица 5

Расширенная сеть при K=6, h=3 и m=2

РС(9, 3, 2, 6, 1|2)						
Номера копий ИС(6)	Номера подсоединенных к копиям ИС(6) абонентов					
	1-я копия ПС(3, 2, 1)		2-я копия ПС(3, 2, 1)		3-я копия ПС(3, 2, 1)	
1	1	2	4	5	7	8
2	1	3	4	6	7	9
3	2	3	5	6	8	9

В табл. 5 представлена расширенная СС, построенная описанным методом из 3-х ИС(6) и 3-х ПС(3, 2, 1).

В такой расширенной СС любые два абонента соединяются *последовательно* только через одну исходную СС. Поэтому эта расширенная СС наследует неблокируемость и самомаршрутизируемость исходной СС. Однако в отличие от простейшей СС в ней некоторые пары абонента могут соединяться *параллельно* через m разных исходных СС.

Два абонента, номера которых i и j сравнимы по модулю N, т.е. связаны отношением $i \equiv j (\mathrm{mod}\ N)$, соединяются параллельно через m разных исходных сетей, а остальные только через σ исходных сетей. Поэтому такая расширенная сеть обозначается как РС($R, N, m, K, \sigma|m$). В табл. 6 представлена РС (21, 7, 4, 12, 2|4) построенная аналогично из 7-и копий ИС(12) и 3-х копий ПС(7, 4, 2).

Таблица 6

РС (21, 7, 4, 12, 2|4), собранная из 3 ПС(7, 4, 2). Состоит из 7 ИС(12).
Число параллельных путей между любой парой абонентов не меньше 2

Копии ИС(12)	РС (21, 7, 4, 12,2|4)											
	1-я копия ПС(7, 4, 2)				2-я копия ПС(7, 4, 2)				3-я копия ПС(7, 4, 2)			
1	1	2	3	4	8	9	10	11	15	16	17	18
2	1	2	5	7	8	9	12	14	15	16	19	21
3	1	3	5	6	8	10	12	13	15	17	19	20
4	1	4	6	7	8	11	13	14	15	18	20	21
5	2	3	6	7	9	10	13	14	16	17	20	21
6	2	4	5	6	9	11	12	13	16	18	19	20
7	3	4	5	7	10	11	12	14	17	18	19	21

Рассмотрим пример расширения полного коммутатора на 16 портов. Какие полные коммутаторы большего размера можно создать, используя его как исходную сеть ИС(16)?

Ответ дает табл. 7. Добавки в скобках получаются за счет использования портов копий ИС(16), которые не заняты полными копиями ПС($N, m,$ 1). Видно, что данным методом коммутатор для 86 абонентов построить невозможно, а коммутатор для 241 абонента неизвестно как строить, т.к. еще не построена соответствующая симметричная блок-схема.

Таблица 7

Полные коммутаторы как расширение коммутатора 16×16.

h	8	5	4	3	2	2	2	1
m	2	3	4	5	6	7	8	16
N	3	7	13	21	31	43	57	241
R	24	35(+1)	52	63(+1)	62(+3)	86–	114	241 ?

Заключение

Рассмотрен метод расширения произвольных системных сетей (СС) с сохранением алгоритма маршрутизации.

Он основывается на описании СС в терминах неполных уравновешенных блок-схем, исследуемых в математической комбинаторике [9].

Этот метод позволяет увеличивать число абонентов СС с одновременным резервированием всех каналов связи.

При этом если в качестве исходной СС берется малый полный коммутатор, то в расширенном (большом) коммутаторе сохраняется возможность бесконфликтной независимой маршрутизации (самомаршрутизации) для каждой пары абонентов источник-приемник. Подобные сети наиболее близки к требованиям так называемых “идеальных” системных сетей [4].

Литература

1. Ivanov A. Networks – what’s new and different when these are on-chip?/ A. Ivanov // Proc. IEEE EWDTW, Odessa, Sept. 17, 2005.

2. Kistler M. Cell Multiprocessor Communication Network: Built for Speed / M. Kistler, M. Perrone, F. Petrini // IEEE MICRO, May-June 2006. – P. 10-23.

3. Flich J. Logic-Based Distributed Routing for NoCs / J. Flich, J. Duato // IEEE COMPUTER ARCHITECTURE LETTERS. – Vol .7. No 1. – January-June 2008. – P. 13-16.

4. Kumar A. Toward ideal on-chip communication using express virtual channels / A. Kumar, L-S. Peh, P. Kundu, N.K. Jha // IEEE MICRO. – № 1. – Jan-Feb, 2008. – P. 80-90.

5. Каравай М.Ф. Минимальное отказоустойчивое вложение в произвольные гамильтоновы графы и реконфигурация при отказах / М.Ф. Каравай // Автоматика и телемеханика, № 12, 2004. – P. 2003-2018.

6. Clos C. A study of non-blocking switching networks. / C. Clos // Bell System Tech. J. – 1953. – Vol. 32. – P. 406-424.

7. Pipenger N. On rearrangeable and non-blocking switching networks / N. Pipenger J. // Comput. Syst. Sci. – 1978. – Vol 17. – P 307-311.

8. Rzymianowicz L. Designing efficient network interfaces for system area networks / L. Rzymianowicz [электронный ресурс]. – Режим доступа к статье: http://bibserv7.bib.uni-mannheim.de/madoc/volltexte /2002/54 /pdf/54_1.pdf.

9. Холл М. Комбинаторика / М. Холл // Изд. «Мир», Москва, 1970.

10. Каравай М.Ф. Комбинаторные методы построения двудольных однородных неизбыточных квазиполных графов (симметричных блок-схем) / М.Ф. Каравай, П.П. Пархоменко, В.С. Подлазов // Автоматика и телемеханика. – 2009. – № 2.– С. 153-170.

11. Guide to myrinet-2000 switches and switch networks [электронный ресурс]. – Режим доступа к руководству: http://www.myti.com/myrinet/m3switch/guide/.

12. Scott S. The black widow high-radix Clos network / S. Scott, D. Abts, J. Kim, W. Dally // Proc. 33rd Intern. Symp. Comp. Arch. (ISCA'2006), 2006 [электронный ресурс]. – Режим доступа к статье: http://cva.stanford.edu/people/jjk12/isca06.pdf.

13. Nikolaev A. The Fault-tolerant Extension of System Area Networks of Multiprocessor System / A. Nikolaev, V. Podlazov // Proceedings of the 17th World Congress. The International Federation of Automatic Control, Seoul, Korea. – July 2008. – P. 10622-10667.

Поступила в редакцию 11.02.2010

Рецензент: д-р техн. наук, профессор, профессор кафедры автоматизации и компьютерных технологий В.А. Краснобаев, Харьковский национальный технический университет сельского хозяйства им. Петра Василенко, Харьков, Украина.

ДО ПИТАННЯ ПОБУДОВИ РОЗШИРЕНОГО ПОВНОГО КОМУТАТОРА – ІДЕАЛЬНОЇ СИСТЕМНОЇ МЕРЕЖІ

М.Ф. Каравай, В.С. Подлазов

Пропонується новий спосіб побудови великого повного комутатора з малих повних комутаторів, заснований на використанні неповних урівноважених блок-схем, досліджуваних в комбінаториці.

Ключові слова: системна мережа, повний комутатор, ідеальна мережа, симетричне розширення, відмовостійкість, алгоритм маршрутизації.

TO THE QUESTION OF CONSTRUCTING THE FULL EXPANDED SWITCHBOARD – IDEAL SYSTEM NETWORK

M.F. Karavay, V.S. Podlazov

A new approach based on incomplete balanced block-designs to implement large complete interconnect networks from small switch bars is proposed. Structurally these networks are the equivalent to quasicomplete graphs. Networks of such structure have the best characteristics of controllability, observability, testability and fault tolerance in relation to the network's components (processors, devices, sensors, links and so on).

Key words: the system network, the full switchboard, the ideal network, symmetric expansion, fault tolerance, routing algorithm.

Каравай Михаил Федорович – д-р техн. наук, профессор, заведующий лабораторией, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, e-mail: mkaravay@ipu.rssi.ru.

Подлазов Виктор Сергеевич – д-р техн. наук, с.н.с, ведущий научный сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, e-mail: podlazov@gmail.com.