

Проект комиссии Президента
по модернизации и технологическому развитию экономики России
«Создание системы подготовки высококвалифицированных кадров
в области суперкомпьютерных технологий и
специализированного программного обеспечения»

Конспект лекций дисциплины

«Архитектура и программное обеспечение высокопроизводительных вычислительных систем»

Разработчик: к.ф.-м.н. Богословский Н.Н.

Москва

Тема № 4 Сети передачи данных

В основе архитектуры любой многопроцессорной вычислительной системы (ВС) лежит способность к обмену данными между компонентами этой ВС. Коммуникационная система ВС представляет собой сеть, узлы которой связаны трактами передачи данных – каналами. В роли узлов могут выступать процессоры, модули памяти, устройства ввода/вывода, коммутаторы либо несколько перечисленных элементов, объединенных в группу. Организация внутренних коммуникаций вычислительной системы называется топологией.

Топологию сети межсоединений (СМС) определяет множество узлов N объединенных множеством каналов C . Связь между узлами обычно реализуется по двухточечной схеме (point-to-point). Любые два узла, связанные каналом связи, называют смежными узлами или соседями. Каждый канал $c = (x, y) \in C$ соединяет один узел-источник (source node) x с одним узлом-получателем (recipient node) y , где $x, y \in N$. Узел-источник, служащий началом канала c , будем обозначать s_c , а узел-получатель – второй конец канала – r_c . Часто пары узлов соединяют два канала, по одному в каждом направлении. Канал $c = (x, y)$ характеризуется шириной (w_c или w_{xy}) - числом сигнальных линий; частотой (f_c или f_{xy}) - скоростью передачи битов по каждой сигнальной линии; задержкой (t_c или t_{xy}) - временем пересылки бита из узла x в узел y . Для большинства каналов задержка находится в прямой зависимости от физической длины линии связи и скорости распространения сигнала.

В зависимости от того, остается ли конфигурация взаимосвязей неизменной, по крайней мере пока выполняется определенное задание, различают сети со статической и динамической топологиями. В статических сетях структура взаимосвязей фиксирована. В сетях с динамической топологией в процессе вычислений конфигурация взаимосвязей с помощью программных средств может быть оперативно изменена.

Узел в сети может быть терминальным, то есть источником или приемником данных, коммутатором, пересылающим информацию с входного порта на выходной, или совмещать обе роли. В сетях с непосредственными связями (direct networks) каждый узел одновременно является как терминальным узлом, так и коммутатором, и сообщения пересылаются между терминальными узлами напрямую. В сетях с косвенными связями (indirect networks) узел может быть либо терминальным, либо коммутатором, но не одновременно, поэтому

сообщения передаются опосредованно, с помощью выделенных коммутирующих узлов. (В дальнейшем для простоты изложения позволим называть оба варианта «прямыми» и «косвенными» сетями, также для краткости вместо терминального узла будем говорить «терминал»-, несмотря на некоторую языковую некорректность.) Существуют также такие топологии, которые нельзя однозначно причислить ни к прямым, ни к косвенным. Любую прямую СМС можно изобразить в виде косвенной, разделив каждый узел на два – терминальный узел и узел коммутации. Современные прямые сети реализуются именно таким образом – коммутатор отделяется от терминального узла и помещается в выделенный маршрутизатор. Основное преимущество прямых СМС в том, что коммутатор может использовать ресурсы терминальной части своего узла. Это становится существенным, если учесть, что, как правило, последний включает в себя вычислительную машину или процессор.

Тремя важнейшими атрибутами СМС являются:

- стратегия синхронизации;
- стратегия коммутации;
- стратегия управления.

Две возможных стратегии синхронизации операций в сети – это синхронная и асинхронная. В синхронных СМС все действия жестко согласованы во времени, что обеспечивается за счет единого генератора тактовых импульсов (ГТИ), сигналы которого одновременно транслируются во все узлы. В асинхронных сетях единого генератора нет, а функции синхронизации распределены по всей системе, причем в разных частях сети часто используются локальные ГТИ.

В зависимости от выбранной стратегии коммутации различают сети с коммутацией соединений и сети с коммутацией пакетов. Как в первом, так и во втором варианте информация пересылается в виде пакета. Пакет представляет собой группу битов, для обозначения которой применяют также термин сообщение.

В сетях с коммутацией соединений путем соответствующей установки коммутирующих элементов сети формируется тракт от узла-источника до узла-получателя, сохраняющийся, пока весь доставляемый пакет не достигнет пункта назначения. Пересылка сообщений между определенной парой узлов производится всегда по одному и тому же маршруту.

Сети с коммутацией пакетов предполагают, что сообщение самостоятельно находит свой путь к месту назначения. В отличие от сетей с коммутацией соединений, маршрут от исходного пункта к пункту назначения каждый раз может быть иным. Пакет последовательно проходит через узлы сети. Очередной узел запоминает принятый пакет в своем буфере временного хранения, анализирует его и делает выводы, что с ним делать дальше. В зависимости от загруженности сети принимается

решение о возможности немедленной пересылки пакета к следующему узлу и о дальнейшем маршруте следования пакета на пути к цели. Если все возможные тракты для перемещения пакета к очередному узлу заняты, в буфере узла формируется очередь пакетов, которая «рассасывается» по мере освобождения линий связи между узлами (если очередь также насыщается, согласно одной из стратегий маршрутизации может произойти так называемый «сброс хвоста» (tail drop), отказ от вновь поступающих пакетов).

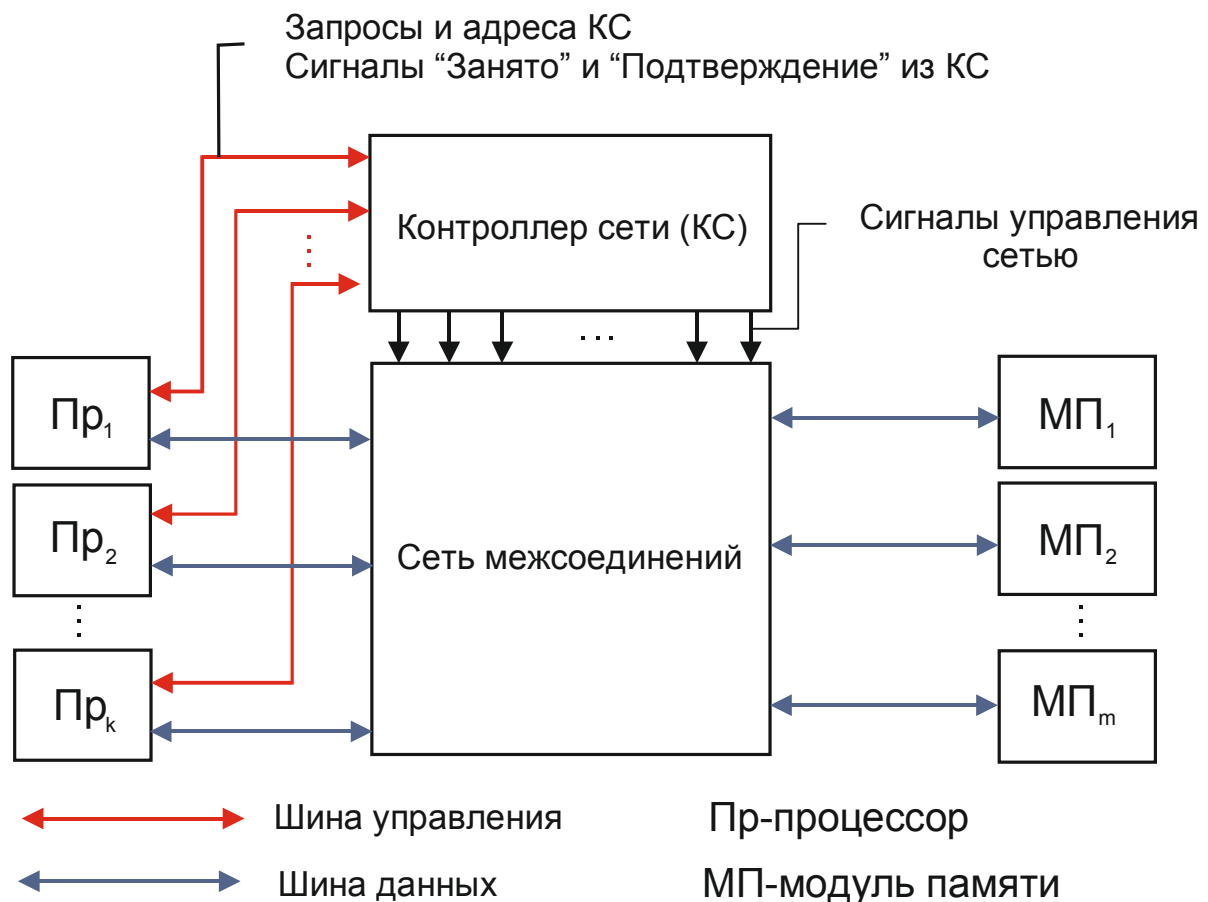


Рис 1. Структура сети с централизованным управлением

СМС можно также классифицировать по тому, как в них организовано управление. В некоторых сетях, особенно с переключением соединений, принято централизованное управление (рис. 1). Процессоры посылают запрос на обслуживание в единый контроллер сети, который производит арбитраж запросов с учетом заданных приоритетов и устанавливает нужный маршрут. К данному типу следует отнести сети с шинной топологией. Процессорные матрицы также строятся как сети с централизованным управлением, которое осуществляется сигналами от центрального процессора. Приведенная схема применима и к сетям с коммутацией пакетов. Здесь тег маршрутизации, хранящийся в заголовке

пакета, определяет адрес узла назначения. Большинство из серийно выпускаемых ВС имеют именно этот тип управления.

В схемах с децентрализованным управлением функции управления распределены по узлам сети.

Вариант с централизацией проще реализуется, но расширение сети в этом случае связано со значительными трудностями. Децентрализованные сети в плане подключения дополнительных узлов значительно гибче, однако взаимодействие узлов в таких сетях существенно сложнее.

В ряде сетей связь между узлами обеспечивается посредством множества коммутаторов, но существуют также сети с одним коммутатором. Наличие большого числа коммутаторов ведет к увеличению времени передачи сообщения, но позволяет использовать простые переключающие элементы. Подобные сети обычно строятся как многоступенчатые.

Метрики сетевых соединений

При описании СМС их обычно характеризуют с помощью следующих параметров:

- размер сети;
- число связей;
- диаметр;
- порядок узла;
- пропускная способность;
- задержки;
- связность;
- ширина бисекции;

Размер сети (network size) – численно равен количеству узлов, объединяемых сетью.

Число связей (number of links) – это суммарное количество каналов между всеми узлами сети. В плане стоимости лучшей следует признать ту сеть, которая требует меньшего числа связей.

Диаметр сети (network diameter) (называемый также коммуникационным расстоянием) – определяет минимальный путь, по которому проходит сообщение между двумя наиболее удаленными друг от друга узлами сети. Путь (path) в сети — это упорядоченное множество каналов $P = (c_1, c_2, \dots, c_n)$ по которым данные от узла-источника, последовательно переходя от одного промежуточного узла к другому, поступают на узел-получатель. Для обозначения отрезка пути между парой смежных узлов применяют термин переход (hop – в живой речи также «транзит» и «хоп»). Минимальный путь от узла x до узла y — это путь с минимальным числом переходов. С возрастанием диаметра сети

увеличивается общее время прохождения сообщения, поэтому разработчики ВС стремятся по возможности обходиться меньшим диаметром.

Порядок узла – каждый узел сети связан с прочими узлами множество каналов. Порядок узла x представляет собой число узлов сети, с которыми данный узел связан напрямую. Например, в сети, организованной в виде матрицы, где каждый узел имеет каналы только к ближайшим соседям (слева, справа, сверху и снизу), порядок узла равен четырем. Увеличение значения этой метрики ведет к усложнению коммутационных устройств сети и, как следствие, к дополнительным задержкам в передаче сообщений. С другой стороны, повышение порядка узлов позволяет реализовать топологии, имеющие меньший диаметр сети, и тем самым сократить время прохождения сообщения. Так, если сеть, состоящая из 4096 узлов, построена по матричной схеме, ее диаметр составляет 126, а для гиперкуба – только 12. Разработчики ВС обычно отдают предпочтение таким топологиям, где порядок всех узлов одинаков, что позволяет строить сети по модульному принципу.

Пропускная способность сети (network bandwidth) – характеризуется количеством информации, которое может быть передано по сети в единицу времени. Обычно измеряется в мегабайтах в секунду или гигабайтах в секунду без учета издержек на передачу избыточной информации, например битов паритета.

Задержка сети (network latency) – это время, требуемое на прохождение сообщения через сеть. В сетях, где время передачи сообщений зависит от маршрута, говорят о средней, минимальной и максимальной задержках сети.

Связность сети (network connectivity) – можно определить как минимальное число узлов или линий связи, которые должны выйти из строя, чтобы сеть распалась на две непересекающихся сети. Следовательно, связность сети характеризует устойчивость сети к повреждениям, то есть ее способность обеспечивать функционирование ВС при отказе компонентов сети.

Ширина бисекции (bisection width) – ширина бисекции позволяет оценить число сообщений, которые могут быть переданы по сети одновременно, при условии что это не вызовет конфликтов из-за попытки использования одних и тех же узлов или линий связи.

Функции маршрутизации данных

Кардинальным вопросом при выборе топологии СМС является способ маршрутизации данных, то есть правило выбора очередного узла, которому пересылается сообщение. Основой маршрутизации служат адреса узлов. Каждому узлу в сети присваивается уникальный адрес. Исходя из этих адресов, а точнее, их двоичных представлений, производится соеди-

нение узлов в статических топологиях или их коммутация в топологиях динамических. В сущности, принятая система соответствия между двоичными кодами адресов смежных узлов – функция маршрутизации данных – и определяет сетевую топологию. Последнюю можно описать как набор функций маршрутизации, задающий порядок выбора промежуточных узлов на пути от узла источника к узлу получателю. В некоторых топологиях используется единая для всей СМС функция маршрутизации, в других — многоступенчатых – при переходе от одной ступени к другой может применяться иная функция маршрутизации.

Статические топологии

К статическим топологиям СМС относят такие, где между двумя узлами возможен только один прямой фиксированный путь, то есть статические топологии не предполагают наличия в сети коммутирующих устройств. Если же такие устройства имеются, то используются они только перед выполнением конкретной задачи, а в процессе всего времени вычислений топология СМС остается неизменной.

Из возможных критериев классификации статических сетей чаще всего выбирают их размерность. С этих позиций различают:

- одномерные топологии (линейный массив);
- двумерные топологии (кольцо, звезда, дерево, решетка, систолический массив);
- трехмерные топологии (полносвязная топология, хордальное кольцо);
- гиперкубическую топологию.

Ниже рассматриваются основные виды статических топологий СМС без акцентирования внимания на какой-либо их классификации.

Линейная топология

В простейшей линейной топологии узлы сети образуют одномерный массив и соединены в цепочки (рис. 2). Линейная топология характеризуется следующими параметрами:

- размер сети N ;
- число связей $N - 1$;
- диаметр $N - 1$;
- порядок узла 2;
- связность 1;
- ширина бисекции 1;

Линейная топология не обладает свойством полной симметричности, поскольку узлы на концах цепочки имеют только одну коммуникационную линию, то есть их порядок равен 1, в то время как порядок остальных уз-

лов равен 2. Время пересылки сообщения зависит от расстояния между узлами, а отказ одного из них способен привести к невозможности пересылки сообщения. По этой причине в линейных СМС используют отказоустойчивые узлы, которые при отказе изолируют себя от сети, позволяя сообщению миновать неисправный узел.



Рис. 2. Линейная топология.

Кольцевые топологии

Стандартная кольцевая топология представляет собой линейную цепочку, концы которой соединены между собой (рис, 3 а). В зависимости от числа каналов между соседними узлами (один или два) различают однонаправленные и двунаправленные кольца. Кольцевая топология характеризуется следующими параметрами:

- размер сети N ;
- число связей N ;
- диаметр $\lceil \frac{N}{2} \rceil$
- порядок узла 2;
- связность 2;
- ширина бисекции 2;

Кольцевая топологии, по сравнению с линейной, традиционно была менее популярной, поскольку добавление или удаление узла требует демонтажа сети.

Один из способов разрешения проблемы большого диаметра кольцевой сети - добавление линий связи в виде хорд, соединяющих определенные узлы кольца. Подобная топология носит название хордальной. Если хорды соединяют узлы с шагом 1 или $N/2 - 1$, диаметр сети уменьшается вдвое. На рис. 3,б показана хордальная кольцевая сеть с шагом 3.

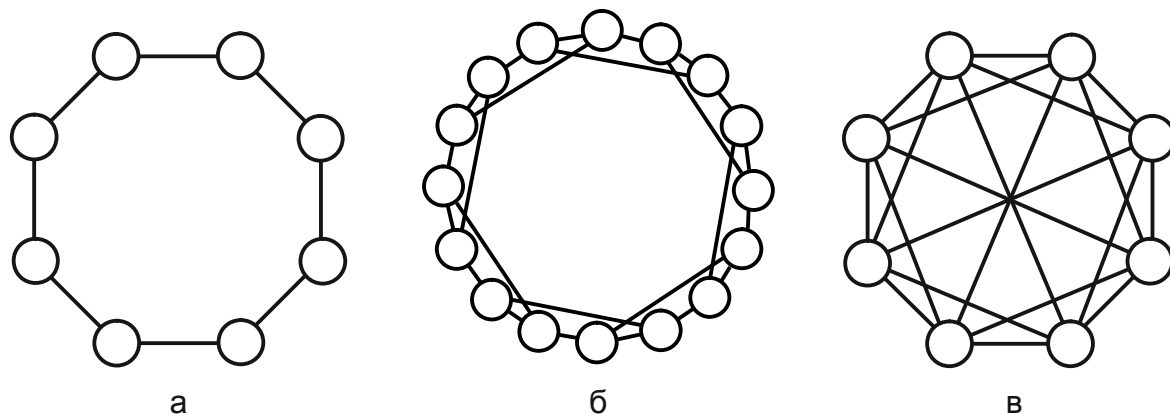


Рис. 3. Кольцевая топология: а – стандартная; б – хордальная; в – с циклическим сдвигом связей.

В качестве практических примеров для топологии кольца следует назвать сети Token Ring, разработанные фирмой IBM, а также вычислительные системы KSR1 и SCI.

Дальнейшее увеличение порядка узлов позволяет добиться еще большего сокращения тракта передачи сообщения. Примером такой топологии может служить, показанная на рис. 3, в, топология с циклическим сдвигом связей. Здесь стандартная кольцевая топология с N узлами дополнена соединениями между всеми узлами i и j , для которых $|i - j|$ совпадает с целой степенью числа 2. Алгоритмы маршрутизации для подобной сети чрезвычайно эффективны, однако порядок узлов по мере разрастания сети увеличивается.

Звездообразная топология

Звездообразная сеть объединяет множество узлов первого порядка посредством специализированного центрального узла - концентратора (рис. 4). Топология характеризуется такими параметрами:

- размер сети N ;
- число связей $N-1$;
- диаметр 2;
- порядок узла 1;
- связность 1;
- ширина бисекции 1;

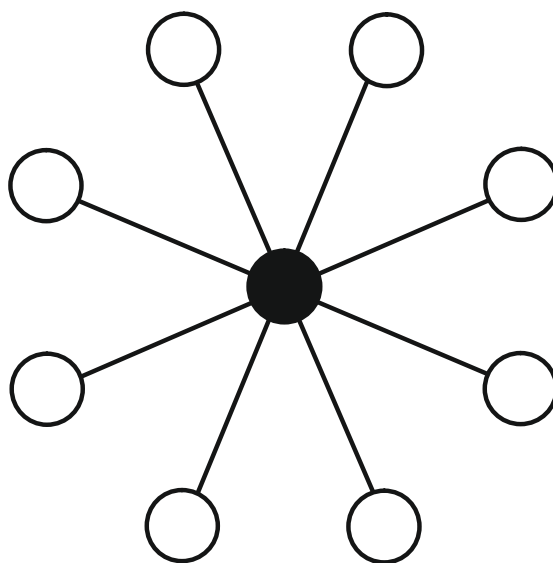


Рис. 4. Звездообразная топология.

Звездообразная организация узлов и соединений редко используется для объединения процессоров многопроцессорной ВС, но хорошо

работает, когда поток информации идет от нескольких вторичных узлов, соединенных с одним первичным узлом, например при подключении терминалов. Общая пропускная способность сети обычно ограничивается быстродействием концентратора, аналогично тому, как сдерживающим элементом в одношинной топологии выступает шина. По производительности эти топологии также идентичны. Основное преимущество звездообразной, схемы в том, что конструктивное исполнение узлов на концах сети может быть очень простым.

Древовидные топологии

Еще одним вариантом структуры СМС является древовидная топология (рис. 5, а). Сеть строится по схеме так называемого строго двоичного дерева, где каждый узел более высокого уровня связан с двумя узлами следующего по порядку более низкого уровня. Узел, находящийся на более высоком уровне, принято называть родительским, а два подключенных к нему нижерасположенных узла — дочерними. В свою очередь, каждый дочерний узел выступает в качестве родительского для двух узлов следующего более низкого уровня. Отметим, что каждый узел связан только с двумя дочерними и одним родительским. Такую сеть можно охарактеризовать следующими параметрами:

- размер сети N ;
- число связей $N-1$;
- диаметр $2(\log_2 N - 1)$;
- порядок узла 3;
- связность 1;
- ширина бисекции 1;

Так, вычислительная система из 262 144 узлов при решетчатой топологии (немного забегаем вперед) будет иметь диаметр 512, а в случае строго бинарного дерева — только 36. Топология двоичного дерева была использована в мультипроцессорной системе DADO из 1023 узлов, разработанной в Колумбийском университете.

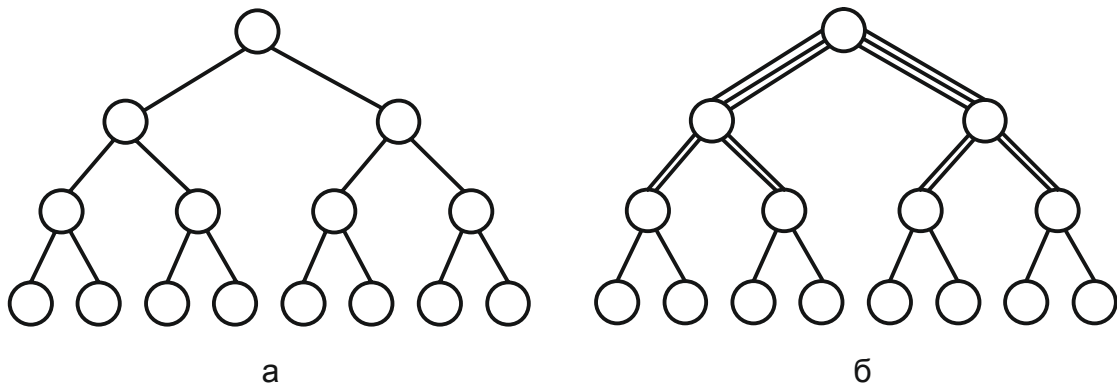


Рисунок 5. Древовидная топология: а - стандартное дерево; б - "толстое" дерево.

При больших объемах пересылок между несмежными узлами древовидная топология оказывается недостаточно эффективной, поскольку сообщения должны проходить через один или несколько промежуточных звеньев. Очевидно, что на более высоких уровнях сети вероятность затора из-за недостаточно высокой пропускной способности линий связи выше. Этот недостаток устраняют с помощью топологии, называемой "толстым" деревом (рис. 5, б).

Идея "толстого" дерева состоит в увеличении пропускной способности коммуникационных линий на прикорневых уровнях сети. С этой целью на верхних уровнях сети родительские и дочерние узлы связывают не одним, а несколькими каналами, причем чем выше уровень, тем больше число каналов. На рисунке это отображено в виде множественных линий между узлами верхних уровней. Топология «толстого» дерева реализована в вычислительной системе СМ-5.

Решетчатые топологии

Поскольку значительная часть научно-технических задач связана с обработкой массивов, вполне естественным представляется стремление учесть это и в топологии ВС, ориентированных на подобные задачи. Такие топологии относят к решетчатым (mesh), а их конфигурация определяется видом и размерностью массива.

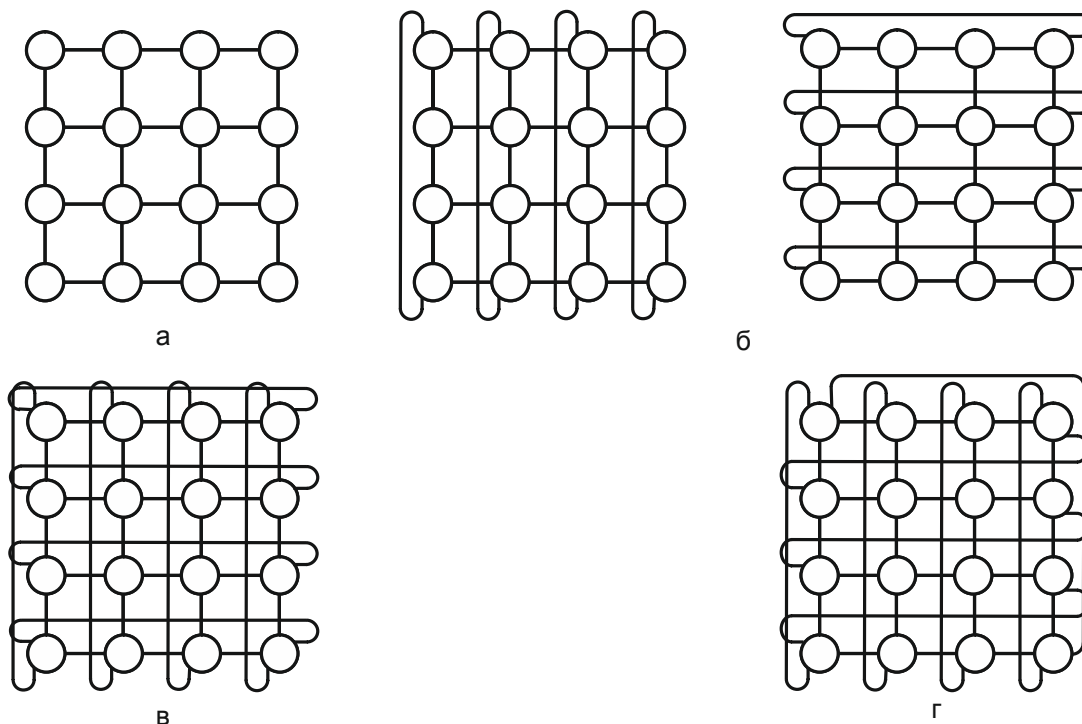


Рис. 5. Решетчатые топологии: а – плоская; б – цилиндрическая; в – тороидальная; г – витая-тороидальная.

Простейшими примерами для одномерных массивов могут служить цепочка и кольцо. Для двумерных массивов данных наиболее подходит топология плоской прямоугольной матрицы узлов, каждый из которых соединен с ближайшим соседом (рис. 13. а). Такая сеть размерности $m \times m$ ($m = \sqrt{N}$) имеет следующие характеристики:

- размер сети N ;
- число связей $2(N - m)$;
- диаметр $2(m - 1)$;
- порядок узла 4;
- связность 2;
- ширины бисекции \sqrt{N} ;

Если провести операцию свертывания (wraparound) плоской матрицы, соединив информационными трактами одноименные узлы левого и правого столбцов или одноименные узлы верхней и нижней строк плоской матрицы, то из плоской конструкции получаем топологию типа цилиндра (рис. 5, б). В топологии цилиндра каждый ряд (или столбец) матрицы представляет собой кольцо. Если одновременно произвести свертывание плоской матрицы в обоих направлениях, получим тороидальную топологию сети (рис. 5, в). Двумерный тор на базе решетки $m \times m$ обладает следующими параметрами:

- размер сети N ;
- число связей $2N$;
- диаметр $2 \left\lceil \frac{m}{2} \right\rceil$;
- порядок узла 4;
- связность 4;
- ширина бисекции $2m$;

Помимо свертывания к плоской решетке может быть применена операция скручивания (twisting). Суть этой операции состоит в том, что вместо колец все узлы объединяются в разомкнутую или замкнутую спираль, то есть узлы, расположенные с противоположных краев плоской решетки, соединяются с некоторым сдвигом. Если горизонтальные петли объединены в виде спирали, образуется так называемая сеть типа ILLIAC. На рис. 5, г показана подобная конфигурация СМС, соответствующая хордальной сети четвертого порядка и характеризуемая следующими метриками:

- размер сети N ;
- число связей $2N$;
- диаметр $m - 1$;

- порядок узла 4;
- связность 4;
- ширина бисекции $2m$;

Следует упомянуть и трехмерные сети. Один из вариантов, реализованный в архитектуре суперЭВМ Cray T3D, представляет собой трехмерный тор, образованный объединением процессоров в кольца по трем координатам: x , y и z .

Примерами ВС, где реализованы различные варианты решетчатых топологий, могут служить: ILLIAC IV, MPP, DAP, CM-2, Paragon и др.

Полносвязная топология

В полностью связной топологии (рис. 6), известной также под названием топологии «максимальной группировки» или полный граф, каждый узел напрямую соединен со всеми остальными узлами сети. Сеть, состоящая из N узлов, имеет следующие параметры:

- размер сети N ;
- число связей $\frac{N(N-1)}{2}$;
- диаметр 1;
- порядок узла $N-1$;
- связность $N-1$;
- ширина бисекции $\frac{N^2}{4}$;

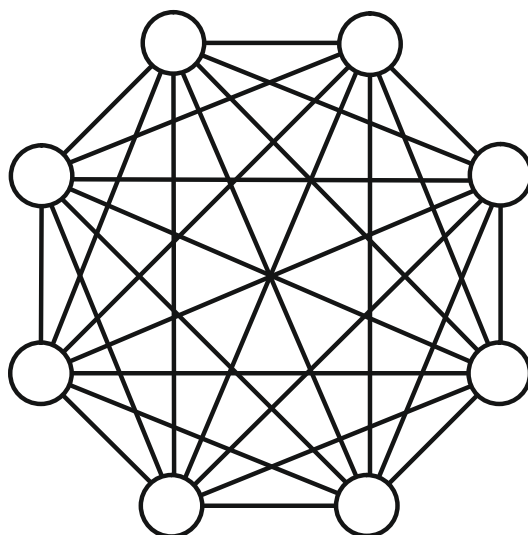


Рис. 6. Полносвязная топология.

Если размер сети велик, топология становится дорогостоящей и трудно реализуемой. Более того, топология максимальной группировки не

дает существенного улучшения производительности, поскольку каждая операция пересылки требует, чтобы узел проанализировал состояние всех своих $N - 1$ входов. Для ускорения этой операции необходимо, чтобы все входы анализировались параллельно, что, в свою очередь, усложняет конструкцию узлов.

Топология гиперкуба

При объединении параллельных процессоров весьма популярна топология гиперкуба, показанная на рис. 7. Линия, соединяющая два узла (рис. 7, а), определяет одномерный гиперкуб. Квадрат, образованный четырьмя узлами (рис. 7, б) - двумерный гиперкуб, а куб из 8 узлов (рис. 7, в) - трехмерный гиперкуб и т. д. Из этого ряда следует алгоритм получения m -мерного гиперкуба: начинаем с $(m - 1)$ - мерного гиперкуба, делаем его идентичную копию, а затем добавляем связи между каждым узлом исходного гиперкуба и одноименным узлом копии.

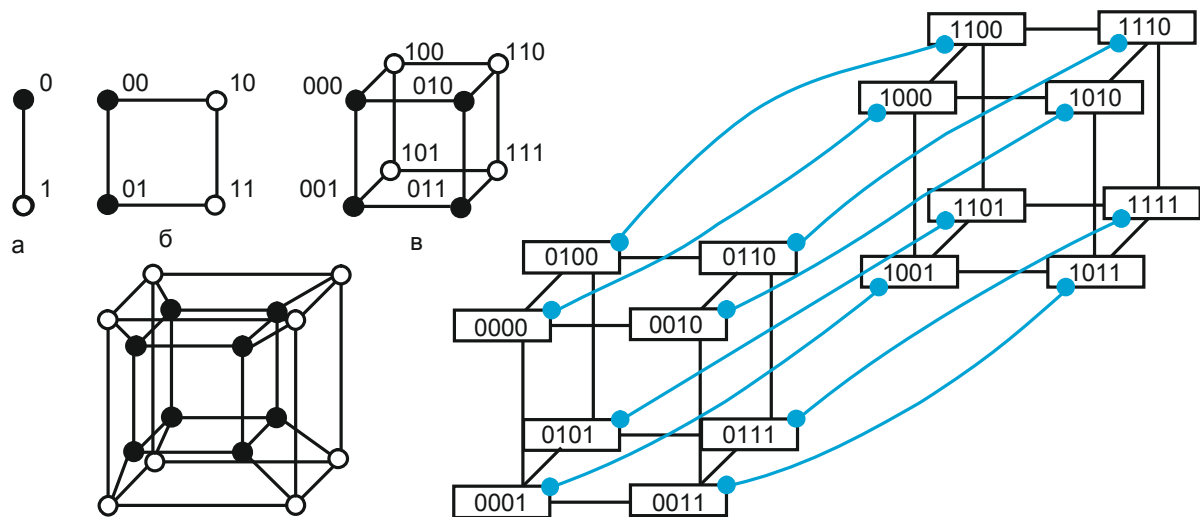


Рис. 7. Гиперкубическая топология.

Гиперкуб размерности $m(N - 2^m)$ имеет следующие характеристики:

- размер сети $m(N - 2^m)$;
- число связей $\frac{mN}{2}$;
- диаметр m ;
- порядок узла m ;
- связность $m / 2$;
- ширина бисекции m ;

Увеличение размерности гиперкуба на 1 ведет к удвоению числа его узлов, увеличению порядка узлов и диаметра сети на единицу. Обмен

сообщениями в гиперкубе базируется на двоичном представлении номеров узлов. Нумерация узлов производится так, что для любой пары смежных узлов двоичное представление номеров этих узлов отличается только в одной позиции. В силу сказанного, узлы 0010 и 0110 — соседи, а узлы 0110 и 0101 таковыми не являются.

Создание гиперкуба при большом числе процессоров требует увеличения порядка узлов, что сопряжено с большими техническими проблемами.

Динамические топологии

В динамической топологии сети соединение узлов обеспечивается электронными ключами, варьируя установки которых можно менять топологию сети. В отличие от ранее рассмотренных топологий, где роль узлов играют сами объекты информационного обмена, в узлах динамических сетей располагаются коммутирующие элементы, а устройства, обменивающиеся сообщениями (терминалы), подключаются к входам и выходам этой сети. В роли терминалов могут выступать процессоры или процессоры и модули памяти.

Если входы и выходы сети коммутирующих элементов разделены, сеть называют двусторонней (two-sided). При совмещенных входах и выходах сеть является односторонней (one-sided).

Обычно ключи в динамических СМС группируются в так называемые ступени коммутации. В зависимости от того, сколько ступеней коммутации содержит сеть, она может быть одноступенчатой или многоступенчатой. Наличие более чем одной ступени коммутации позволяет обеспечить множественность путей между любыми парами входов и выходов.

Блокирующие и неблокирующие многоуровневые сети

Минимальным требованием к сети с коммутацией является поддержка соединения любого входа с любым выходом. Для этого в сети с n входами и n выходами система ключей обязана предоставить n вариантов коммутации входов и выходов (перестановок - permutations). Проблема усложняется, когда сеть должна обеспечивать одновременную передачу данных между многими парами терминальных узлов (multicast), причем так, чтобы не возникали конфликты (блокировки) из-за передачи данных через одни и те же коммутирующие элементы в одно и то же время. Подобные топологии должны поддерживать n перестановок. С этих позиций все топологии СМС с коммутацией разделяются на три типа: неблокирующие, неблокирующие с реконfigurацией и блокирующие.

В **неблокирующих сетях** обеспечивается соединение между любыми парами входных и выходных терминалов без перенастройки коммутирующих элементов сети. В рамках этой группы различают сети

строго неблокирующие (strictly non-blocking) и неблокирующие в широком смысле (wide sense non-blocking). В строго неблокирующих сетях возникновение блокировок принципиально невозможно в силу примененной топологии. К таким относятся матричная сеть и сеть Клоша. Неблокирующими в широком смысле называют топологии, в которых конфликты при любых соединениях не возникают только при соблюдении определенного алгоритма маршрутизации.

В неблокирующих сетях с реконfigurацией также возможна реализация соединения между произвольными входными и выходными терминалами, но для этого необходимо изменить настройку коммутаторов сети и маршрут связи между соседними терминалами. Примерами таких сетей служат сети Бенеша, Бэтчера, «Мемфис» и др.

В блокирующих сетях, если какое-либо соединение уже установлено, это может стать причиной невозможности установления других соединений. К блокирующим относятся сети «Баньяна», «Омега», n-куб и др.