

РАБОТА ЛОКАЛЬНЫХ КОМПЛЕКСОВ GRID В РЕЖИМЕ ЖЕСТКОГО РЕАЛЬНОГО ВРЕМЕНИ

Г. Г. Стецюра

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

Рассмотрено применение в локальных GRID групповых операций и совмещения процессов передачи и обработки данных, позволяющее на предлагаемых GRID решать задачи жесткого реального времени.

ВВЕДЕНИЕ

Существует два типа сетей GRID — глобальные, с выходом в сети дальней связи, например, в Internet, и локальные, в которых компьютеры объединены локальной сетью. Сети GRID обычно рассматриваются как распределенный компьютер, однако характер выполнения вычислений в GRID весьма специфичен. Основная задача объединенных сетью компьютеров — обслуживать владельцев этих компьютеров, а распределенная по сети задача решается в фоновом режиме. Алгоритмы решения распределенных задач в глобальных GRID создаются с учетом непредсказуемой задержки вычислений и частого отключения удаленных компьютеров. Локальные сети GRID могут предоставлять сервис, существенно отличающийся от того, который дают глобальные GRID. Тем не менее, принципиальных отличий в реализованных сетях нет. Например, широко известная локальная GRID Condor [1] функционирует подобно глобальной GRID, но является более скоростной.

В настоящей статье рассмотрено использование локальных сетей GRID для решения задач жесткого реального времени (ЖРВ). Это могут быть задачи управления работой технических средств или вычислительные задачи, в которых требуется выдерживать временную согласованность выполнения частей задачи на разных компьютерах сети.

Будем называть GRID, способную решать такие задачи, комплексом GRID, чтобы подчеркнуть наличие в ней дополнительных структурных и технических решений, приближающих сетевую структуру к компьютеру.

Существуют два режима, в которых GRID может решать задачи ЖРВ. Во-первых, после окон-

чания рабочего дня пользователи покидают свои компьютеры, и для всей совокупности таких компьютеров распределенная задача становится основной. Во-вторых, даже при наличии пользователей, для распределенной задачи, требующей режима ЖРВ, можно, как это обычно делается в сетях для передачи срочных сообщений, выделить квант времени, который гарантированно предоставляется распределенной задаче. Во втором варианте права владельцев ущемляются, но появляется возможность гарантированно управлять ответственными процессами, например, физическим экспериментом. Таким образом, здесь речь идет о расширении функций локальных сетей GRID.

Легко прослеживается аналогия между указанными режимами работы GRID и работой автоматизированных систем управления, в частности, промышленных АСУ. Для таких систем в Институте проблем управления РАН ранее были разработаны методы, позволяющие существенно повысить реактивность систем [2, 3]. В последние годы они были дополнительно развиты и предлагаются в этой статье для применения в локальных комплексах GRID.

Центральная особенность разработанных методов состоит в использовании групповых операций. Под групповой операцией здесь понимается логическая или арифметическая операция, которую выполняет группа взаимодействующих устройств над данными, хранящимися в этих устройствах. При этом вычисления производятся в процессе передачи промежуточных результатов операции по каналу без их задержки на обработку [2]. Указанная независимость времени выполнения операции от количества участвующих в ней компьютеров — основная особенность групповых операций. Отсутствие задержки на вычисление достигается за счет совме-



щения обработки данных с их передачей по каналам связи.

Ниже кратко изложены особенности разработанных в ИПУ методов и применение их для реализации функций GRID в указанных выше режимах работы.

1. ПРИОРИТЕТНОЕ УПРАВЛЕНИЕ ДОСТУПОМ К РЕСУРСАМ

В работах [2, 4] изложены способы децентрализованного приоритетного управления (ДПУ) доступом компьютеров к средствам связи. Управление осуществляется на основе приоритетов компьютеров. Приоритеты могут изменяться динамически с учетом потребностей компьютеров и обеспечивают однозначный выбор компьютера, которому и предоставляется доступ. Способы ДПУ, прежде всего, обеспечивают доступ за гарантированное время, что необходимо для систем ЖРВ. Однако обеспечивается не только доступ к средствам связи. Средства связи — лишь общая среда общения компьютеров. Приоритеты ДПУ могут относиться к любым ресурсам GRID. Запросы компьютеров могут вначале отбираться, исходя из доступности ресурсов, которые требуются компьютерам. Только после этого, для обеспечения однозначности доступа, привлекаются другие параметры. Время доступа при ДПУ логарифмически зависит от числа компьютеров.

В способах ДПУ используются сигналы, которые позволяют обнаружить наличие в канале связи сигналов, соответствующих двоичной единице, при одновременном с ними нахождении сигналов двоичного нуля. Этого достаточно для выполнения операции *max*, которая последовательно сравнивает разряды передаваемых двоичных кодов приоритетов и выделяет наибольшее значение приоритета.

В ИПУ были разработаны также способы доступа, решающие рассмотренную задачу на основе введения различных временных задержек сигнала, соответствующих разным приоритетам — это методы децентрализованного пространственно-временного управления.

2. СОВМЕЩЕНИЕ ОБРАБОТКИ ДАННЫХ С ИХ ПЕРЕДАЧЕЙ

Для проведения обработки данных без задержки (“на лету”), в процессе передачи используется канал с направленной передачей данных, в котором от компьютера к компьютеру последовательно передаются сигналы. Сигнал по каналу поступает к одному из компьютеров, без задержки обрабатывается им, преобразованный сигнал поступает к

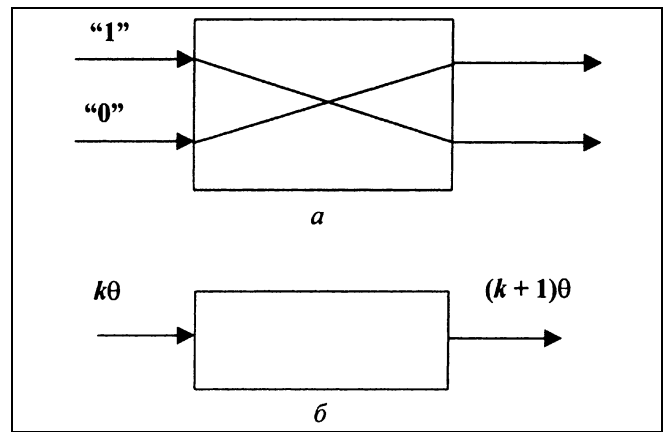
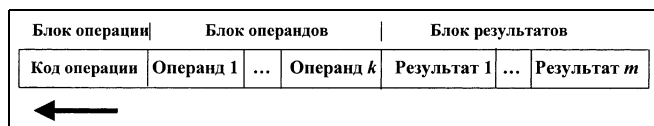


Рис. 1. Варианты инвертирования сигнала:
a — переключение линий; *b* — вращение плоскости поляризации

следующему компьютеру и т. д. Без задержки можно выполнить операцию только в том случае, если для выбора необходимого действия не требуется знать значение приходящего по каналу сигнала. Трудности при этом возникают с единственной операцией — инвертированием значения сигнала. Так как предполагается использование двоичных сигналов, то речь идет о превращении сигнала “1” в сигнал “0” и наоборот. На рис. 1 показаны способы выполнения инвертирования с нулевой временной задержкой [4]. На рис. 1, *a* сигнал в инвертор поступает по одной из двух линий и выходит из него также по двум линиям. Если сигнал поступает по верхней линии, то это соответствует приходу двоичной единицы. Сигнал на нижней линии соответствует двоичному нулю. В инверторе линии перекрещены, что и обеспечивает инвертирование значения сигнала. На рис. 1, *b* показан другой вариант инвертирования [5]. Сигнал предполагается поляризованным. Он проходит через среду, вращающую плоскость поляризации сигнала на угол $\theta = 360^\circ/n$, где n — четное число. При выполнении этого условия по поляризации сигнала можно определить его двоичное значение, во-вторых, инвертирование, как и в первом варианте реализации этой операции, происходит без задержки.

Отметим отличие приведенных элементов от других известных реализаций. В последних входной сигнал не покидает элемент. Его энергия используется для управления созданием выходного сигнала за счет внешнего источника энергии. В разработанных в ИПУ вариантах входной сигнал передается непосредственно на выход элемента.

С помощью указанных преобразований реализуются логические и арифметические операции. В публикациях [2, 4] такие операции названы “вы-

**Рис. 2. Структура групповой команды**

числением в общем канале” (ВОК). Несмотря на последовательную поразрядную передачу операнда ВОК выполняется очень быстро — за время перемещения операнда по каналу выполняется количество операций, равное количеству компьютеров комплекса. Например, для суммирования N чисел, распределенных по одному между N компьютерами, по каналу передается значение суммы, которое последовательно формирует каждый из N компьютеров. Это оказывается возможным только вследствие отсутствия задержки при выполнении ВОК. В работе [6] приведены области вычислений, где применение ВОК наиболее эффективно.

3. ГРУППОВЫЕ КОМАНДЫ, РАСПРЕДЕЛЕННЫЕ ПРОГРАММЫ

Групповые команды (ГК) — основа рассматриваемого в этом разделе распределенного программирования [5]. Их можно представить как сообщение, состоящее из трех блоков, показанных на рис. 2. Первый блок содержит код операции, требующий от компьютеров комплекса определенного действия. Второй блок содержит набор операндов. Компьютеры используют этот блок совместно с хранящимися в них данными для получения результата операции. Третий блок — блок результатов — содержит результаты выполнения команды. Он хранит результаты двух видов. Это может быть последовательность индивидуальных результатов, поочередно переданных в ГК участвовавшими в ее выполнении компьютерами. Но блок может также содержать результаты совместной работы компьютеров, выполненные с применением ВОК. Пример — создание суммы чисел непосредственно в групповой команде. Таким образом, ГК, перемещаясь между компьютерами, наполняется результатами заданной в ней операции, и за один обход всех компьютеров выполняется сложная совокупность действий.

В распределенной программе, естественно, необходимы операции ветвления процесса вычисления. Помимо традиционных способов ветвления здесь возможны ветвления, требующие одновременного участия большого количества компьютеров. При таком ветвлении источник программы посылает всем компьютерам групповую команду, указывающую, какими свойствами должны обладать компьютеры, чтобы иметь право в дальней-

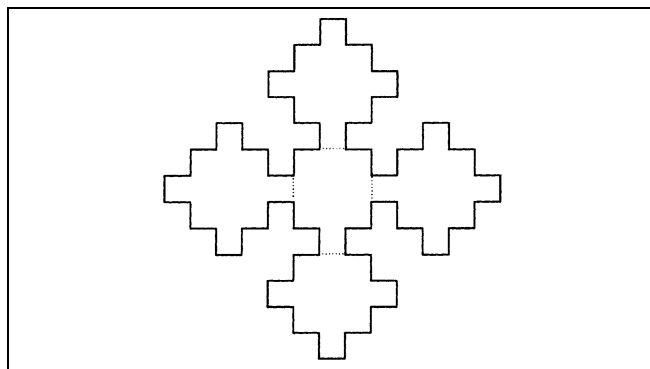
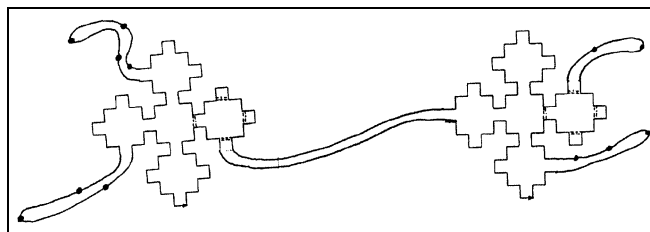
шем получить управление обработкой данных. Из этих компьютеров должен быть выделен один. В § 1 показано, как это делается.

Таким образом, при распределенном программировании последовательно выполняются групповые команды.

4. СТРУКТУРЫ СВЯЗЕЙ В ЛОКАЛЬНЫХ GRID С ГРУППОВЫМИ ОПЕРАЦИЯМИ

В рассматриваемых комплексах GRID применимы многие виды структур связи [2, 4]. Здесь останемся на новой, фрактальной структуре канала связи, разработанной специально для комплексов, использующих групповые операции [4]. Канал связи имеет вид фрактальной кривой Серпинского и строится итерациями с применением L -системы Лиденмауэра [7]. На рис. 3 показан канал, полученный после второй итерации. Структура связи имеет регулярный вид, ребра графа на рис. 3 представляют собой каналы связи, вершины — компьютеры с дополнительным оборудованием для выполнения ГК. Вся структура разделяется на отдельные кластеры с целью разделения потоков данных в комплексе и изоляции его дефектных частей [4]. На рис. 3 пунктиром показаны места, в которых целесообразно производить указанные разделения.

В реальных GRID каналы связи на рис. 3 могут иметь различную длину, как это показано на рис. 4.

**Рис. 3. Фрактальный канал****Рис. 4. Фрактальные связи с серверами**



Здесь части со связями равной длины будем считать расположенными в сервере комплекса. Удаленные связи подходят непосредственно к компьютерам на рабочих местах (показаны точками). Рассмотренные выше разбиения на кластеры производятся в пределах сервера. На рис. 4 показаны два сервера, каждый их них — это многопроцессорная структура с указанными фрактальными связями. В этой структуре компоненты серверов и клиентские компьютеры включаются в единый процесс обработки информации с применением групповых операций.

5. ГРУППОВЫЕ КОМАНДЫ И НАДЕЖНЫЕ ШИРОКОВЕЩАТЕЛЬНЫЕ СООБЩЕНИЯ

Надежность широковещательных сообщений совершенно необходима для режима ЖРВ. Должна быть гарантия того, что все участвующие в операции компьютеры приняли сообщение. Если это условие не соблюдается, то для корректировки действий надо как можно быстрее выяснить подробности сложившейся ситуации. Групповые команды здесь помогают следующим образом. Групповая команда проверяет работоспособность приемников широковещательного сообщения. В блоке результата этой команды приемники с помощью ВОК формируют сумму, к которой каждый работоспособный приемник прибавляет по единице. Получив эти данные, источник команды определяет состояние комплекса в целом. Если это не случайный сбой, то производится реорганизация системы. Посылается еще одна групповая команда, которая соберет в своем блоке результатов сведения о неработоспособных приемниках [4]. Далее, как показано выше в примере с ветвлением распределенной программы, источник запрашивает замену, и к решению задачи подключаются другие компьютеры. Аналогичным способом могут периодически осуществляться распределенная диагностика комплекса и восстановление его работоспособности.

Групповые команды позволяют быстро определять не только поступление данных в компьютеры, но и их идентичность. Самый быстрый способ состоит в следующем. Каждый компьютер вычисляет контрольную сумму принятого данного. После этого посылается групповая команда, в блоке результата которой компьютеры складывают контрольные суммы. Так как источник знает контрольную сумму посланного им сообщения и количество приемников, то он способен определить идентичность всех принятых сообщений. Так обнаруживаются и сообщения с правильной контрольной суммой, но не соответствующие посланному широковещательному сообщению.

Рассмотренные в настоящем разделе операции были применены в АСУТП [3].

6. СИНХРОНИЗАЦИЯ РАСПРЕДЕЛЕННЫХ РАБОТ В ЛОКАЛЬНОЙ GRID

Далее приведены два характерных для GRID примера синхронизации распределенных работ.

6.1. Распределенная общая память

Одна из важных задач режима ЖРВ заключается в синхронизации блоков распределенной общей памяти (РОП) — частей общей памяти комплекса, расположенных в разных компьютерах [8]. Каждый компьютер решает только часть общей задачи, но для ее решения требуется иметь полную информацию о состоянии объекта управления, создаваемую в разных местах комплекса. Для этого источник должен рассылать новые данные в удаленные блоки РОП. При этом требуется, во-первых, разрешить только поочередную работу источников данных, во-вторых, передающий данные источник должен быть уверен, что процедура обновления содержимого РОП прошла успешно. Быстрое и надежное копирование данных осуществимо с применением методов, изложенных в § 5. Накладные расходы на проверку завершения операции здесь минимальны.

6.2. Синхронизация выполнения идентичных заданий

В GRID поведение распределенных компьютеров непредсказуемо — задание на вычисление может прерываться другим, более приоритетным заданием, компьютер может быть выключен пользователем и т. п. Поэтому на практике одно и то же задание посылается на выполнение группе компьютеров, их результаты заказчик собирает и формирует из них окончательный результат.

Групповые операции существенно ускоряют управление такой процедурой. Выделение группы для выполнения задания выполняется способом, аналогичным поиску компьютеров при передаче управления в распределенной программе (см. § 3). Но здесь выделяется не один, а несколько компьютеров. Количество выделенных компьютеров определяется посылкой одной групповой операции. Так же быстро выполняется передача исходных данных и сбор результатов. Для наблюдения за продвижением процесса вычисления достаточно послать ГК управления с несколькими полями результата в блоке результата. Задается насколько уровней оценки продвижения вычисления и каждому уровню ставится в соответствие один результат. Каждый компьютер группы прибавляет единицу к тому результату, который соответствует уровню продвижения вычислений в данном ком-

пьютере. Фактически в блоке результата ГК строится гистограмма. Общение с компьютерами производится с указанием имени результата, к которому отнес себя компьютер.

7. О ПРОГРАММНОЙ И АППАРАТНОЙ РЕАЛИЗАЦИИ КОМПЛЕКСОВ GRID

Необходимые для построения комплексов GRID аппаратура и программы достаточно просты. Сложность состоит в отличии протокола взаимодействия компьютеров в комплексе GRID от обычно применяемого в локальных сетях протокола TCP/IP. Заметим, что практика применения в специализированных локальных сетях протоколов, отличных от TCP/IP, довольно распространена. Так, в промышленных сетях широко применяется протокол I²C, многие микроконтроллеры имеют встроенные средства, поддерживающие этот протокол. Расширение приложений сетей, например, создание сетей интеллектуальных датчиков, также порождает новые, существенно отличные от традиционных протоколы [9].

Однако для локальных сетей общего применения, а именно на такие должны ориентироваться комплексы GRID, практически обязательно использование TCP/IP. Поэтому надо уметь совмещать в одной системе два протокола — TCP/IP для локальной сети и протокол группового взаимодействия для объединения компьютеров комплекса. Для этого проще всего потоки данных, относящиеся к разным протоколам, посылать по разным каналам. С этой целью, например, в локальных сетях Ethernet достаточно задействовать имеющиеся в кабелях связи таких сетей неиспользуемые линии.

ЗАКЛЮЧЕНИЕ

Применение коллективного взаимодействия компьютеров при выполнении групповых операций — важный шаг в превращении локальной

GRID в действительно распределенный компьютер. Еще раз подчеркнем различие между организацией обработки данных в GRID без групповых операций и с ними. В первом случае стремятся распределить между компьютерами предельно независимые части задачи. Взаимодействие с ними осуществляется обычным сетевым способом. Во втором случае на совокупности компьютеров решается общая задача, и для выполнения вычислений используются средства связи.

ЛИТЕРАТУРА

1. *Thain D., Tannenbaum T., Livny M.* Condor and the Grid // Grid Computing: Making The Global Infrastructure a Reality. — John Wiley, 2003.
2. *Прангишвили И.В., Подлазов В.С., Стецюра Г.Г.* Локальные микропроцессорные вычислительные сети. — М.: Наука, 1984.
3. *Аленов А. В., Подлазов В. С., Слободчук, Смирнов А. Н., Стецюра Г. Г.* Локальные управляющие сети для тесно связанных распределенных систем жесткого реального времени // Вычислительная техника, системы управления. — 1991. — Вып. 6.
4. *Стецюра Г. Г.* Возможности применения фрактальных связей и групповых операций в многопроцессорных системах с перестраиваемой структурой для эволюционных вычислений // Автоматика и телемеханика. — 2003. — № 12.
5. *Стецюра Г. Г.* Возможности технической реализации клеточных автоматов с дальними связями и групповыми операциями // Автоматика и телемеханика. — 2004. — № 8.
6. *Подлазов В. С., Смирнов А. Н., Стецюра Г. Г.* Вычисления в общем канале // Многопроцессорные вычислительные системы с перестраиваемой структурой — М.: Научный совет по комплексной проблеме “Кибернетика”, 1981.
7. *Кроновер Р. М.* Фракталы и хаос в динамических системах. — М.: Постмаркет, 2000.
8. *Илларионов Ю. А., Стецюра Г. Г.* Взаимодействие процессов при разделении ресурсов в распределенных вычислительных системах // Измерения, контроль, автоматизация. — 1987. — № 3.
9. *Стецюра Г. Г.* Групповые команды для сетей мобильных датчиков // Датчики и системы. — 2004. — № 4.

☎ (095) 334-78-31

E-mail: stetsura@ipu.ru



ВНИМАНИЕ!

Журнал "Проблемы управления" на компакт-диске

Появилась возможность приобретения компакт-диска, полностью воспроизводящего все номера журнала "Проблемы управления" за 2003 г.

ЭТО УДОБНО И НЕДОРОГО

Стоимость диска равна примерно половине стоимости одного номера журнала. Его смогут приобрести не только библиотеки, но и кафедры вузов, отделы и лаборатории, ученые и специалисты.

Цена диска — договорная.

Заказать диск можно в редакции журнала "Проблемы управления": 117997, ГСП-7, Москва,

Профсоюзная ул., 65, офис 104

Тел./факс (095) 330-42-66, тел. 334-92-00

E-mail: datchik@ipu.ru