

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.722.25

УПРАВЛЕНИЕ ОТКАЗАМИ В ДЕЦЕНТРАЛИЗОВАННЫХ ГРИДАХ, ОБЛАДАЮЩИХ СВОЙСТВАМИ ТЕСНОГО МИРА

© 2011 г.

С.А. Анисимов

Нижегородский государственный технический университет им. Р.Е. Алексеева

Sergey.Anisimov@inbox.ru

Поступила в редакцию 02.11.2010

Рассмотрен децентрализованный грид, построенный с помощью фракталов с ограниченной степенью вершин. Предложена методика обнаружения неисправностей в сегментах сети. Разработаны различные алгоритмы управления отказами, выбор применения которых обусловлен типом грида, техническими возможностями и правилами поддержки, установленными в виртуальной организации.

Ключевые слова: сеть тесного мира, децентрализованный грид, управление отказами.

Введение

В настоящее время существуют сложные научные, производственные и инженерные задачи, для решения которых в разумные сроки недостаточно одного компьютера, кластера или даже специализированного суперкомпьютера. Для решения данных задач в настоящий момент используется технология «грид» – аппаратно-программная инфраструктура, которая обеспечивает надежный, устойчивый, повсеместный и недорогой доступ к высокопроизводительным компьютерным ресурсам, их скоординированное разделение и решение задач в динамически меняющихся виртуальных организациях со многими участниками [1, 2]. Основное отличие технологии «грид» от других парадигм распределенных вычислений заключается в возможности объединять, в том числе, и гетерогенные ресурсы [3].

Архитектура грида

Подробно существующая архитектура грида, стандарты и примеры систем показаны в работах [1–8]. Несколько упрощенно грид можно представить следующим образом: есть множество ресурсов, объединенных в сеть. Существует несколько основных типов ресурсов:

- *вычислительный ресурс* предоставляет задаче пользователя грида процессорные мощности. Основной характеристикой вычислительного ресурса является производительность;

- *ресурс памяти* представляет собой пространство для хранения данных. Основной ха-

рактеристикой ресурса памяти является его объем;

- *сетевой ресурс* является связующим звеном между другими элементами сети. Основной характеристикой сетевого ресурса является скорость передачи данных;

- *каталог (информационный ресурс)* есть особый тип ресурса памяти. Используется для хранения, предоставления метаданных и информации о других ресурсах грида. Каталог позволяет структурированно хранить массивы информации о состоянии грида, эффективно выполнять поиск ресурсов и осуществлять мониторинг ошибок в системе.

Нетрудно заметить, что каталог ресурсов является корневым сервером грида, в котором содержится список узлов сети и их состояние. Соответственно, при постановке в очередь задачи и запросе необходимого количества ресурсов одним из узлов грида происходит обращение к корневому серверу и поиск по каталогу ресурсов, удовлетворяющих искомому запросу. В настоящее время во всех гридах существует централизованный каталог ресурсов. Недостатки построения грида способом, описанным выше, следующие:

- в случае выхода из строя корневого сервера весь грид неработоспособен;

- наличие устаревшей информации о состоянии узлов либо парализация корневого сервера грида частыми запросами.

Если вероятность отказа системы можно снизить с помощью репликации информационных ресурсов, то второй недостаток традицион-

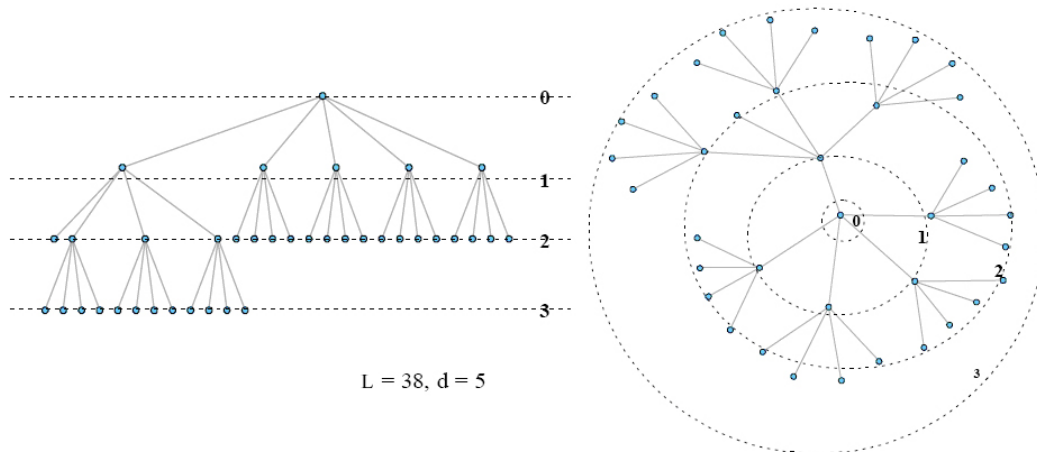


Рис. 1. Уровни LDF сети. Ступенчатое (слева) и круговое (справа) представление

ных гридов устранить нельзя. В работах [9–12] показано, что оптимальной топологией децентрализованного грида (графа) является фрактал с ограниченной степенью вершин Limited Degree Fractal (LDF), в котором наблюдается эффект тесного мира [12, 13].

Постановка задачи

В работе [14] были показаны методика построения децентрализованного грида и протокол поиска ресурсов в ней. Основная идея заключалась в организации каталога на каждом ресурсе грида, однако первый должен содержать информацию только о соседних узлах сети. В работе [15] было показано преимущество такой организации информационных ресурсов системы по сравнению с традиционной организацией грида, а именно:

- найдено аналитическое выражение для распределения нагрузки относительно количества узлов сети одного уровня в зависимости от количества запросов, степени вершины грида и радиуса сети для LDF топологии;
- проанализирована динамика изменения максимальной нагрузки децентрализованного грида с увеличением числа узлов и количества захватываемых ресурсов;
- предложены изменения алгоритма для компенсации нагрузки на центральные узлы при большой нагрузке LDF сети;
- проведено сравнение максимальных нагрузок LDF сети с нагрузками для классического грида.

Таким образом, грид при авариях на каналах связи либо на ресурсах, степени вершин которых не равны единицам, распадется на несколько работоспособных сегментов. В настоящее время не существует методики обнаружения неисправностей в каналах связи и узлах сети и алгоритмов управления отказами для LDF гридов.

Цель данной работы – предложить методику обнаружения различных неисправностей в сегментах сети, разработать различные алгоритмы управления отказами, выбор применения которых обусловлен типом грида, техническими возможностями и правилами поддержки, установленными в виртуальной организации.

Алгоритмы управления отказами

Необходимо в соответствии с [14, 15] понимать, что каталог ресурсов осуществляет мониторинг состояния только соседних узлов. Реакция системы на обнаружение той или иной ошибки может быть чрезвычайно разнообразной в зависимости от вида грида, технической возможности, сути ошибки: от уведомления администраторов домена до реконфигурации сети. Все неисправности концептуально можно разделить на два типа:

- некорректная работа узла – ресурс доступен, каталог ресурсов работает корректно. Для части заданий K , выполненных за последние N единиц времени, был возвращен код ошибки либо были возвращены некорректные значения тестовых заданий;
- узел недоступен – под этим необходимо понимать либо повреждение каналов связи, либо аварийное завершение работы ресурса без уведомления соседних узлов.

В первом случае для обнаружения некорректной работы узла используются отчуждаемые ресурсы элемента сети, которые в зависимости от типа последнего периодически осуществляют процедуру записи-чтения на диск с выводом результата либо решение тестового задания с выводом результата и скорости выполнения. Пороговые значения получаются эмпирическим исследованием для каждого конкретного грида. Отрицательным фактором данного способа является необходимость частично от-

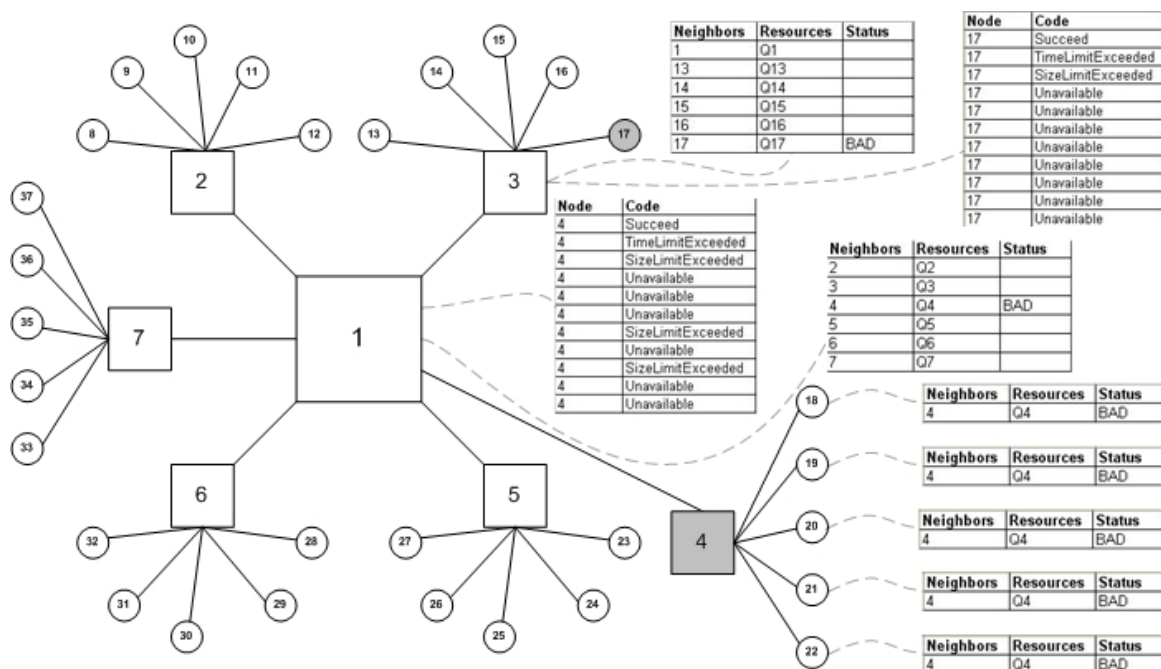


Рис. 2. Динамический алгоритм управления отказами

чуждать ресурсы, в противном случае при полном занятии последних задачи мониторинга выполнены не будут. В случае анализа возвращаемых протоколом Lightweight Directory Access Protocol (LDAP) кодов ошибок необходимо лишь поддерживать таблицу последних N кодов ошибок [7], получаемых при завершении работы задания. При превышении порога происходит уведомление администратора домена о некорректной работе узла. Однако для более эффективной работы был разработан динамический алгоритм управления отказами, реализация которого описана ниже (в таблицах рис. 2 показаны только необходимые поля).

Пусть в узле 17, степень вершины которого равна единице, произошел сбой. В таблице кодов ошибок информационного ресурса 3 через некоторое время произойдет превышение порогового значения. Необходимо уточнить, что заполнение таблиц происходит тогда и только тогда, когда ресурс, выполнявший задание, является соседним с узлом, где заполняются коды ошибок. После отправки сообщения администратору домена о неисправности узла 17 в основной таблице соседних ресурсов каталога 3 статусу неисправного компьютера будет присвоено значение «BAD». Следовательно, в ресурсном запросе узла 3 будет фигурировать требование того, что внутренний идентификатор элемента грида не должен быть равен 17. Большой интерес представляет ситуация, когда степень вершины неисправного ресурса (узел 4) отлична от единицы. Пусть в информационном ресурсе узла 19 произошло превышение порогового значения. Полностью аналогично пре-

дыдущему случаю в основной таблице соседних ресурсов каталога 19 статусу неисправного компьютера 4 будет также присвоено значение «BAD». Далее происходит формирование запроса на изменение статуса неисправного элемента сети у соседних с ним узлов. Правила пересылки запросов указаны в [14, 15]. В итоге, как показано на рис. 2, все узлы, соседние четвертому, получили значение «BAD» в соответствующих основных таблицах информационных ресурсов. Стоит заметить, что целесообразность данного дополнения подходит только для промышленных гридов, в которых предъявляются высокие требования к производительности ресурсов и надежности каналов связи. Таким образом, неисправные узлы будут исключены из дальнейшей обработки заданий грида.

Необходимо указать способы возобновления дальнейшей обработки заданий проблемным узлом грида. В случае ручного управления системой администратор домена генерирует запрос к соседним узлам с тэгом «ALIVE». При получении данного запроса узел, являющийся соседним у инициатора запроса, очищает статус в основной таблице. При работе сети в автоматическом режиме соседние узлы периодически посылают эталонные задания неисправному узлу. В случае, если возвращаемое значение корректное, происходит удаление статуса «BAD» из основной таблицы каталога.

Для детектирования ситуации потери соединения с узлом необходимо использовать стандартные средства – многочисленные аналоги команды ping протокола Internet Control Message Protocol (ICMP). Таким образом, если ини-

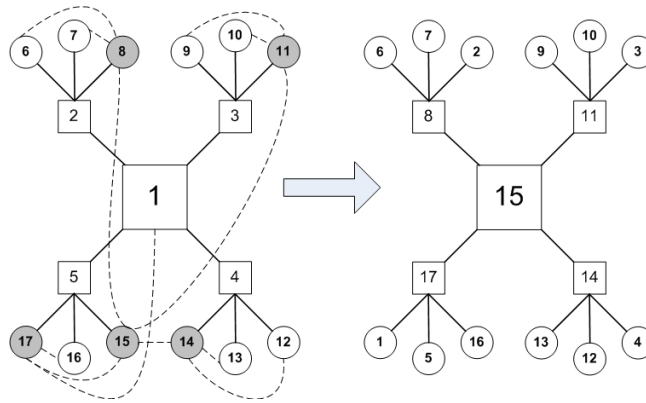


Рис. 3. Модификация грида с заменой всех неграничных узлов

циатор запроса не получает ответа на L сообщений, то он отсылает сообщение администратору домена, и, если это предусмотрено, переходит на резервные каналы связи. Если узел недоступен несмотря на переключение каналов, то инициатор определяет аварийное завершение работы ресурса. В случае, если степень вершин последнего равна единице, то родительский узел должен указать в статусе основной таблицы, что узел выключен: «N/A». Данная авария никак не влияет на общую производительность системы при значительном числе узлов. Представляет интерес ситуация, когда выходит из строя центральный элемент сети. В отличие от традиционного грида система не завершит работу при аварии на информационном ресурсе, но распадется на d сегментов (где d – степень вершины центрального узла), что резко снизит производительность грида. В случае, если произошла аварийная остановка узла, степень вершины которого не равна единице, то при наличии технической возможности необходимо реконфигурировать грид таким образом, чтобы все узлы, которые не принадлежат последнему уровню LDF сети обязательно имели степень вершины узла, равную единице. Известно, что:

$$\begin{cases} \sim \frac{1}{d} \text{ вершин имеют степень } d, \\ \sim 1 - \frac{1}{d} \text{ вершин имеют степень } 1. \end{cases} \quad (1)$$

Количество узлов LDF сети с полностью заполненными уровнями

$$L = \sum_{n=2}^N ((d-1)^{n-2} d) + 1. \quad (2)$$

Нетрудно заметить, что существует множество вариантов реконфигурации грида, однако в силу географической удаленности отдельных узлов и политики виртуальной организации есть ненулевая вероятность, что текущий вариант будет единственным. В общем случае модификация грида показана на рис. 3.

Должны существовать резервные каналы связи, обозначенные пунктиром на рис. 3, необходимые для трансформации сети.

Заключение

В данной работе был рассмотрен децентрализованный грид, построенный с помощью фракталов с ограниченной степенью вершин. Предложены методики обнаружения различных видов неисправностей в сегментах сети. Разработан динамический адаптивный алгоритм управления отказами. Предложен алгоритм трансформации сети для сохранения работоспособности грида.

Список литературы

1. Foster I., Kesselman C. The Grid: Blueprint for a New Computing Infrastructure (2nd Edition). San Mateo, CA: Morgan Kaufmann, 2004
2. Foster I., Kesselman C., Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations // Inter. J. of High Performance Computing Applications. 2001. № 15. P. 200–222.
3. Haynos M. Perspectives on grid: Grid computing – Next-generation distributed computing. IBM Developer Works, 2004.
4. Foster I., Kesselman C., Tuecke S. The physiology of the grid. An Open Grid Services Architecture for Distributed Systems Integration // Global Grid Forum. 2002.
5. Czajkowski K., Fitzgerald S., Foster I., Kesselman C. Grid Information Services for Distributed Resource Sharing // Proceedings of the 10-th IEEE International Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press. 2001.
6. Tierney B., Aydt R., Gunter D. A grid monitoring architecture, 2000. URL: <http://www.ogf.org/documents/GFD.7.pdf> (дата обращения 01.09.2010).
7. Yeong W., Howes T., Kille S., Lightweight Directory Access Protocol // Network Working Group, 1995. URL: <http://www.ietf.org/rfc/rfc1777.txt> (дата обращения 01.09.2010).

8. Коваленко В.Н., Корягин Д.А. Организация ресурсов грид // ИПМ им. М. В. Келдыша РАН. 2004, № 63.
9. Aiello W., Chung F. R. K., Lu L. A random graph model for massive graphs // ACM Symposium on Theory of Computing (STOC). 2000. P. 171–180.
10. Albert R., Barabasi A.-L. Statistical mechanics of complex networks // Rev. Modern Phys. 2002. № 74. P. 47–97.
11. Erdos P., Renyi A. The evolution of random graphs // Publications of the Mathematical Institute of the Hungarian Academy of Sciences. 1960. № 5. P. 17–61.
12. Жаринов И. В., Крылов В. В. Конструирование графов с минимальной средней длиной пути // Вестник ИжГТУ. 2008. № 4.
13. Watts D., Strogatz S. Collective dynamics of small-world networks // Nature. 1998. № 363. P. 202–204.
14. Анисимов С. А., Зыбин В. А., Крылов В. В. Протокол поиска ресурсов в децентрализованных гридах, обладающих свойствами тесного мира // Труды НГТУ. 2009. Т. 80. № 1. С. 13–19.
15. Анисимов С. А. Распределение нагрузки в децентрализованных гридах, обладающих свойствами тесного мира // Вестник ННГУ (в печ.)

FAULT MANAGEMENT IN DECENTRALIZED GRIDS WITH SMALL-WORLD CHARACTERISTICS

S.A. Anisimov

A decentralized grid built using fractals with a limited degree of vertices is considered. A technique of failure detection in the network segments is proposed. Different fault management algorithms have been developed. The selection of the algorithm depends on the type of the grid, technical capabilities and support rules set out in the virtual organization.

Keywords: small-world network, decentralized grid, fault management.