

УДК 519.716.325.

**Использование логических алгоритмов для оптимального построения
Грид - систем.**

Димитриченко Дмитрий Петрович

Учреждение Российской академии наук институт прикладной математики Кабардино-Балкарского научного центра РАН

Email: dimdp@rambler.ru

Концепция grid появилась еще в конце 60-х годов и как огромное число других интеллектуальных инноваций ее рождение и развитие многие годы было связано с фундаментальными научными исследованиями.

Проблема использования свободных ресурсов компьютеров была особенно актуальна 25-35 лет назад, когда для построения серьезных математических моделей были нужны мощности суперкомпьютерных центров, которых тогда в мире было совсем не много. А технически задача решалась довольно просто, поскольку речь шла о разовых вычислительных заданиях без очень жестких требований по их реализации.

Сам же термин "грид" возник в середине 90-х годов и стал обозначать распределенную вычислительную инфраструктуру для сложных инженерных и научных расчетов. Постоянный прогресс в этой области, а также, настойчивый поиск решений все новых классов задач привели к расширению этого понятия. В настоящее время концепция grid (название предложено по аналогии с электрическими сетями - electric power grid) состоит прежде всего в глобальной интеграции компьютерных ресурсов.

Одно из ранних (1998 г.) определений grid, данное Карлом Кессельманом и Яном Фостером (которых считают отцами Грид-технологии), гласит: «Вычислительный grid является программно-аппаратной инфраструктурой, которая обеспечивает надежный, совместимый, повсеместный и недорогой доступ к вычислительным ресурсам большой мощности». Затем, в 2000 году, к этому определению добавилось «координированное распределение ресурсов и решение проблем в динамических виртуальных организациях».

По предложению Яна Фостера, систему можно называть grid, если она:

1. Координирует ресурсы, которые не контролируются централизованно. Например, объединяются компьютерные системы, находящиеся в разных организациях или разных административных единицах одной компании. Система должна решать вопросы политики доступа, безопасности, оплаты услуг и т. д., которые возникают в гетерогенных системах. В ином случае можно говорить о локально управляющейся системе;

2. Задействует открытые, стандартные протоколы и интерфейсы общего назначения. Эти протоколы и интерфейсы используются для решения таких базовых вопросов, как аутентификация, авторизация, поиск ресурсов и доступ к ним. Если протоколы не являются стандартными и открытыми, система является специализированной в отношении приложения;

3. Предоставляет новое качество. Объединяя различные ресурсы, grid позволяет предоставить новый уровень сервиса с точки зрения времени отклика, пропускной способности, доступности, безопасности и т. д. Иными словами, польза от всей системы существенно больше, чем от простой суммы составляющих ее частей.

Основными характеризующими свойствами Грид-систем являются следующие:

1. масштабы вычислительного ресурса (объем памяти, количество процессоров), многократно превосходят ресурсы отдельного компьютера или одного вычислительного комплекса;

2. гетерогенность среды; в ее состав могут входить компьютеры различной мощности, работающие под управлением различных операционных систем и собранные на различной элементной базе;

3. пространственное (географическое) распределение информационно-вычислительного ресурса;

4. объединение ресурсов, которые не могут управляться централизованно (в случае, если они не принадлежат одной организации);
5. использование стандартных, открытых, общедоступных протоколов и интерфейсов;
6. обеспечение информационной безопасности.

Для наиболее полного раскрытия всех возможностей Грид-систем необходимо решить следующие вопросы:

1. объединение разнородных систем [1];
2. совместное использование данных [2];
3. динамическое выделение ресурсов [3];
4. переносимость приложений в гетерогенной среде [1-3];
5. обеспечение информационной безопасности.

Специфики Грид-систем определяются, прежде всего, характером самой вычислительной среды на которой они строятся, представляя собой динамическое образование, с изменяющимся в течение некоторого периода времени количеством активных вычислительных узлов (например, за счет нештатного сбоя, планового выключения или подключения новых вычислительных мощностей), так и их качественных характеристик (изменение вычислительной нагрузки одного узла).

Очевидно, что свойства самой вычислительной среды определяются характеристиками сетевых топологических структур, образующих саму эту среду, в рамках которой будет функционировать проектируемая Грид-система.

Из выше перечисленных требований, предъявляемых к вычислительным компьютерным сетям, как основе для построения Грид-систем, наиболее существенными являются следующие:

- Надёжность;
- Стоимость;
- Пропускная способность.

Так как именно эти характеристики отражают наиболее важные аспекты функционирования вычислительных сетей любого масштаба, соответственно:

- Интегральный параметр, характеризующий устойчивую работоспособность системы в целом;
- Экономический фактор, актуальный для заказчика (проектировщика);
- Качество обслуживания абонентов проектируемой сети.

Эти характеристики мы и положим в основу критериев оценки основных сетевых топологических структур [4].

При этом под базовыми топологиями мы будем понимать совокупность следующих топологий:

- полная ячеистая топология;
- кольцевая топология;
- топология “Звезда”;
- линейная топология.

Поскольку сетевая топологическая структура не всегда соответствует одной из рассмотренных выше базовых структур, то определим еще один класс: «смешанная сетевая топологическая структура».

Такой подкласс будет необходим для анализа сетей, отличных от базовых (что является достаточно распространенным в практике сетевого строительства).

Для дальнейшего анализа мы выделили пять основных сетевых топологических структур, используемых для построения Грид-систем. При этом нами выделены наиболее актуальные критерии надежности, стоимости и пропускной способности. Надежность позволит Грид-системе сохранять устойчивую работоспособность. Пропускная способность обеспечит наилучшую передачу пакетов данных (актуальных для каждого из активных узлов вычислительного процесса). Критерий стоимости является, отчасти, критерием менее важным,

поскольку Грид-системы строятся на основе уже функционирующих (ранее приобретенных) сетей. Однако если речь идет о стоимости обслуживания сетевой составляющей такой системы в целом, или же о покупке нового сетевого оборудования для построения Грид-системы, то этот критерий окажется полезным.

Способ получения оценочных функций для каждого из критериев был изложен в [4]. В рамках данной работы существенным является не способ вычисления оценочных функций а факт их наличия, что делает возможным **Анализ оценок при помощи логических методов и алгоритмов.**

Эффективным методом моделирования свойств объектов в слабо формализуемых областях знаний является метод описания объектов при помощи переменных логических предикатов [5].

Такая интерпретация значений логических предикатов позволяет рассматривать значение логической функции, как описание совокупности автоматных состояний системы «анализируемое множество объектов». Например, цвета: «синий», «зеленый» и «красный» в рамках такого подхода будут выражать лишь состояние, в котором находится объект по характеристике «цвет». При этом не утверждается, что «красный» самый лучший цвет, а «синий» или «зеленый» являются самыми худшими.

Таким образом, предлагаемая логическая модель описания объектов в терминах переменных логических предикатов является более универсальной. Значение соответствующей логической функции при конечной интерпретации является автоматом, характеризующим наиболее подходящие (близкие к запрошенной совокупности состояний) объекты [6].

Для целей многокритериального анализа рассматриваемых объектов (в нашем случае сетевых топологических структур) мы будем использовать логические алгоритмы.

Поскольку, в данном случае, мы располагаем конечным числом m анализируемых объектов, каждый из которых характеризуется набором из n наиболее существенных (актуальных) для целей анализа признаков.

Приведем общую постановку задачи:

Пусть W – множество рассматриваемых объектов, а x_1, x_2, \dots, x_n – предикаты значности k_i каждый, характеризующие свойства (признаки) объектов.

Пусть оценка объекта по каждому из n критериев (признаков), сопоставлена с некоторым значением логической переменной соответствующей значности. Тогда всю совокупность из m рассматриваемых объектов можно охарактеризовать наборами из n логических переменных.

Основой для построения системы логических функций, описывающих множество из m рассматриваемых объектов, служит база данных (БД), представленная следующей таблицей 1:

Таблица 1.

x_1	x_2	...	x_n	W
$x_1(w_1)$	$x_2(w_1)$...	$x_n(w_1)$	w_1
$x_1(w_2)$	$x_2(w_2)$...	$x_n(w_2)$	w_2
.....
$x_1(w_m)$	$x_2(w_m)$...	$x_n(w_m)$	w_m

В этой таблице каждый соответствующий признак $x_i(w_j)$ в общем случае кодируется предикатом k_i -значности, $k_i \geq 2$, $i = 1, \dots, n$, $j = 1, \dots, m$. Таблицу 1 будем называть обучающей БД или обучающей выборкой (ОВ).

Функция, построенная по такой ОВ, способом, предложенным в [5,6], будет иметь следующий вид

$$f(x) = \bigwedge_{j=1}^m (\bigwedge_{i=1}^n x_i(w_j) \rightarrow w_j) = \bigwedge_{j=1}^m (\bigvee_{i=1}^n \overline{x_i(w_j)} \vee w_j).$$

Указанный вид функции следует из известного логического тождества: $a \rightarrow b = \overline{a} \vee b$, где a конъюнкция характеристик (признаков), определяющих объект, а b — предикат, равный единице, когда w_j становится равным соответствующему определяемому объекту. В рамках данной статьи такие предикаты будем называть объектными предикатами, а дизъюнкты, содержащие такие предикаты- продукционными.

Для простоты обозначим критерии надежности, стоимости и пропускной способности буквами a, b, c соответственно.

А анализируемые топологии: "полноячеистую", "звезду", "кольцо", "линейную", "смешанную", - буквами F "Full", S "Star", R "Ring", L "Line", M "Mix".

Для учета плотности значений оценочных функций, полученных в [4], интервал $[0, \dots, 1]$ (для каждого из критериев) разбивается на интервалы переменной длины в соответствии с расстояниями между значениями оценок. Такое разбиение порождает для трех критериев логики различных значностей: для $A = 4$, для $B = 3$, для $C = 5$.

Соответствующие значения логических предикатов приведены в таблице 2.

Таблица 2

Топология	Надежность (0-3)	Стоимость (0-2)	Проп. способность (0-4)
Полноячеистая	3	0	4
Звезда	2	1	1
Кольцо	0	1	0
Линейная	1	2	2
Смешанная	2	1	3

Отметим, что таблица 2 целиком определяется значениями оценочных функций, которые в общем случае считаются произвольно заданными.

Как отмечалось выше, наиболее существенными для построения Грид-систем являются критерии надежности и пропускной способности. Построим логическую функцию, характеризующую сетевые топологические структуры по двум этим критериям. Логическая функция Q_f будет иметь следующий вид:

$$Q_f = A^0 C^1 \vee A^0 C^2 \vee A^0 C^3 \vee A^0 C^4 \vee A^0 R \vee A^1 C^0 \vee A^1 C^1 \vee A^1 C^3 \vee A^1 C^4 \vee \\ \vee A^1 L \vee A^2 C^0 \vee A^2 C^2 \vee A^2 C^4 \vee A^2 S M \vee A^3 C^0 \vee A^3 C^1 \vee A^3 C^2 \vee A^3 C^3 \vee A^3 F \vee \\ \vee C^0 R \vee C^1 S \vee C^2 L \vee C^3 M \vee C^4 F \vee F S R L M$$

Однако для анализа полученных данных достаточно исследовать продукционную часть этой функции, исключив дизъюнкты свободных знаний [7]:

$$Q_p = A^2 S M \vee A^0 R \vee A^1 L \vee A^3 F \vee C^0 R \vee C^1 S \vee C^2 L \vee C^3 M \vee C^4 F \vee F S R L M$$

Поскольку исследуемые топологические структуры практически все различны, то, в данном случае, выявлен только один класс: топологии «звезда» и «смешанная» совпадают по критерию топологической надежности. Остальные дизъюнкты составляют множество логических аксиом.

При этом интерпретация некоторых групп дизъюнктов будет следующей:

1. Наличие в вычисленной логической функции дизъюнктов, относящихся к свободным знаниям свидетельствует о том, что в базе знаний отсутствуют объекты с заданным значением соответствующих параметров.

2. Дизъюнкт-аксиома свидетельствует о том, что в базе знаний существует единственный объект, у которого текущее значение рассматриваемого параметра соответствует заданному во входном векторе.

3. Наличие в вычисленной функции классов обозначает, что в базе знаний присутствуют несколько объектов, в которых значение соответствующего параметра равно задаваемому во входном векторе.

4. Дизъюнкт, являющийся конъюнкцией всех, содержащихся в БД объектов, говорит о том, что игнорирование свойств объектов делает их равнозначными, т.е. всегда есть возможность сделать выбор "любой из объектов, находящихся в базе знаний".

Для выбора оптимальной топологической структуры достаточно в продукционную часть полученной логической функции подставить запрашиваемые (актуальные для заказчика) характеристики.

Если, например, значения логических переменных интерпретировать, как величины предложенных выше оценок и вычислить полученную функцию при максимальных значениях переменных A и B останутся только те дизъюнкты, у которых $A = 3$ и $B = 4$: $Q_p = F \vee F \vee FSRLM$. Таким образом, мы получили одну полноячеистую топологию.

Это, на самом деле, отражает тот факт, что без учета стоимости (или каких-либо иных затрат) самой надежной и оптимальной по быстродействию является полноячеистая топология.

Если необходимо учесть влияние на выбор оптимальной топологии всех трех критериев, то надо вычислить логическую функцию, целиком опираясь на всю совокупность данных таблицы 2.

Полученная при таком разбиении продукционная часть логической функции H имеет следующий вид:

$$H = A^0 R \vee A^1 L \vee A^2 SM \vee A^3 F \vee B^0 F \vee B^1 SRM \vee \\ \vee B^2 L \vee C^0 R \vee C^1 S \vee C^2 L \vee C^3 M \vee C^4 F \vee FSRLM$$

Дальнейший отбор множества оптимальных топологий может быть осуществлен при помощи анализа частот присутствия искомого идентификатора (сетевой топологической структуры) в вычисленной логической функции.

В качестве примера вычислим логическую функцию H при следующих значениях переменных $A = 2$, $B = 1$, $C = 3$, поскольку нас интересует сетевая структура с надежностью выше среднего, недорогая по стоимости, с приемлемой пропускной способностью.

В данном случае мы имеем следующие соотношения: "S, M, S, R, M, M".

Отметим, что объектный дизъюнкт прикладного значения не имеет и при частотном анализе значений логической функции не учитывается.

Соотношение частот будет следующим:

$$R = 1/6, S = 2/6, M = 3/6.$$

Таким образом, при заданных условиях:

- § наиболее предпочтительной является смешанная топология;
- § менее предпочтительна топология "звезда";
- § кольцевая топология является менее всего предпочтительной.

Литература

1. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. – М.: Наука, 2006. – 410 с.

2. Сырцев А.В., Тимофеев А.В. Модели и методы маршрутизации потоков данных в телекоммуникационных системах с изменяющейся динамикой. М.: Новые технологии, 2006, 85 с.

3. Тимофеев А.В. Мульти-агентные системы управления региональными телекоммуникационными сетями. – Материалы международной конференции «Моделирование устойчивого регионального развития (МУРР–2007)», Изд. КБНЦ РАН, Нальчик, 2007, с. 45–50.

4. Дмитриченко Д.П. Получение комплексной оценки при сравнении сетевых топологических структур // Материалы международного конгресса студентов, аспирантов и молодых ученых «Перспектива 2007». С.175-177.

5. Лютикова Л.А., Тимофеев А.В., Стурев В.В., Йоцов В.И Развитие и применение многозначных логик и сетевых потоков в интеллектуальных системах. // Труды СПИИРАН, вып. 2, 2005. С. 114–126.

6. Лютикова Л.А. Моделирование и минимизация баз знаний в терминах многозначной логики предикатов. Нальчик. – Препринт, 2006.

7. Дмитриченко Д.П. К вопросу об интерпретации дизъюнктов переменнзначных логических функций, построенных по базе знаний сетевых топологических структур. // Вестник ВГТУ. 2008. Том 4. № 4. С. 81–85.