

Жуков И. А., д-р техн. наук,
Иванкевич А. В., канд. техн. наук,
Салим Аль Шибани,
Аль-Сурики Ибрагим

МЕТОД И ПРОГРАММНЫЕ СРЕДСТВА ОПТИМИЗАЦИИ РАСПРЕДЕЛЕНИЯ ИНФОРМАЦИОННЫХ ФАЙЛОВ В КОМПЬЮТЕРНЫХ СЕТЯХ

Институт компьютерных технологий Национального авиационного университета

Предложен метод оптимального распределения информационных файлов по узлам компьютерной сети при передаче больших объемов данных. Исследованы особенности повышения производительности универсальных мультисервисных сетей при использовании программных средств оптимизации загрузки узлов сетей.

Введение

В настоящее время объемы данных, хранимых и обрабатываемых в различных сетевых информационных приложениях, продолжают возрастать. Однако данные имеют реальную ценность лишь в том случае, если они получены в приемлемые сроки. Поэтому необходимо не только наращивать емкость носителей, но и повышать скорость доступа к ним.

Активно ведутся работы по созданию конвергированных или мультисервисных сетей, способных обеспечить транспорт любых данных, речи, видео и т.д. [1]. На сегодняшний день можно говорить о реализации отдельных фрагментов мультисервисных сетей, обеспечивающих, в основном, предоставление некоторого пакета услуг, а не о мультисервисных сетях в полном их понимании [1].

Вследствие совершенствования аппаратно-программных средств транспортировки больших объемов информации в сетях, что позволяет ускорить передачу и обработку данных в различных информационных системах, создание коммутационных систем, предназначенных для анализа состояния сети в каждый момент времени и оптимальной транспортировки данных является предметом прикладных и теоретических исследований.

Системы, выполняющие даже простейшую оптимизацию распределения

передачи данных по каналам сети, остаются чрезвычайно дорогими, а эффективность использования возможностей универсальных коммутаторов при передаче больших объемов мультимедийной информации по нескольким каналам одновременно – сравнительно низкой. Кроме того, на некоторых этапах передачи из-за неравномерного распределения потоков данных происходят заторы пакетов в отдельных узлах компьютерной сети, что снижает быстродействие сети в целом.

Решению вопросов оптимального распределения информационных файлов по узлам компьютерной сети посвящены исследования [2, 3], однако большинство из них базируется на теории очередей. При этом используются сложные модели, для которых точных методов реализации не существует или вводятся нереальные ограничения. Известно, что для сложных топологических структур сети построить реальные модели на основе теории очередей невозможно [4]. Поэтому разработка методов рационального использования информационных ресурсов в компьютерной сети является актуальной задачей.

Оптимизация загрузки узлов сети. Обеспечение многопользовательского доступа к информационным ресурсам, хранящимся в виде распределенной

базы данных (РБД), предполагает рациональное размещение файлов базы данных в узлах компьютерной сети.

Известные подходы к распределению информационных ресурсов обычно сводятся к формулировке оптимизационных задач, относящихся к классу задач целочисленного математического программирования. Существующие математические модели отличаются видом целевой функции и совокупностью ограничений, учитываемых при поиске оптимального варианта [4,5]. В большинстве случаев оптимизационные задачи решаются приближенными методами.

Для поиска оптимальных решений практическое применение находит известный программный продукт *Matlab* [5], в состав которого входит комплекс модулей оптимизации. Эти модули могут использоваться для линейных и нелинейных целевых функций при разных видах ограничений. В универсальной процедуре поиска решения реализованы симплекс-метод, алгоритм нелинейной оптимизации *Generalized Reduced Gradient*, метод «ветвей и границ» для линейных и целочисленных задач оптимизации с ограничениями и др.

После нахождения оптимального решения предоставляются возможности анализа устойчивости и чувствительности модели. Для оценки поведения системы в различных ситуациях можно провести параметрический анализ, причем изменение любого параметра с некоторым шагом может быть сохранено в виде сценария. Предусмотрены встроенные инструменты корректировки сценариев, а также формирования итогового отчета с результатами многовариантного исследования. В каждой новой матрице, отражающей загрузку узлов, предусмотрены дополнительные возможности обеспечения эффективной работы пользователей с *Matlab*.

Оценка возможностей оптимизации на базе *Matlab* применительно к задаче эффективной организации РБД осуществ-

лялась при выборе рационального размещения файлов.

Модель задачи оптимизации распределения информационных файлов. В случае использования критерия минимума затрат, обусловленных передачей данных между узлами компьютерной сети, математическая модель будет иметь вид [4, 6]:

$$C = C_1 + C_2 = \sum_{i=1}^n \sum_{j=1}^m \lambda_{ij} (\alpha_{ij} + \beta_{ij}) \min_{s: x_{ij}=1} d_{si} + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^n \lambda'_{ij} \gamma_{ij} d_{si} x_{sj} \rightarrow \min \quad (1)$$

при ограничениях

$$\sum_{i=1}^n x_{ij} \geq 1; \quad j = 1, 2, \dots, m; \quad (2)$$

$$\sum_{j=1}^m L_j x_{ij} \leq V_j; \quad j = 1, 2, \dots, n; \quad (3)$$

$$x_{ij} \in \{0 \cup 1\}; \quad i = 1, 2, \dots, m; \\ j = 1, 2, \dots, n, \quad (4)$$

где n – число узлов компьютерной сети;

m – число независимых файлов, входящих в состав РБД;

V_i – объем памяти для размещения файлов локальной базы данных i -го узла, $i=1, 2, \dots, n$;

L_j – размер j -го файла, $j=1, 2, \dots, m$;

λ_{ij} – интенсивность запросов, возникающих в i -м узле и требующих выборки данных из j -го файла;

α_{ij} – объем данных для запроса, инициированного в i -м узле к j -му файлу;

β_{ij} – объем данных, пересылаемых в i -й узел при ответе на запрос к j -му файлу;

λ'_{ij} и γ_{ij} – интенсивность сообщений, направляемых из i -го узла для корректировки содержимого всех копий j -го файла, и объем такого сообщения;

d_{si} – затраты на передачу единицы информации между s -м и i -м узлами ($s=1, 2, \dots, n$; $d_{ss}=0$).

Решением сформулированной задачи являются значения переменных $\{x_{ij}\}$, которые задают оптимальный вариант распределения файлов:

$$x_{ij} = \begin{cases} 1, & \text{если } j\text{-й файл обрабатывается в } i\text{-м узле;} \\ 0, & \text{в противном случае.} \end{cases}$$

В соответствии с условием (2) общее число копий для каждого файла может быть произвольным с учетом ограничений (3) на ресурсы памяти для локальных БД в узлах компьютерной сети.

Целевая функция (1) общих затрат на пересылку данных при функционировании информационной системы в течение единицы времени, включает две составляющие: величину C_1 , обусловленную данными, которые передаются в процессе выполнения запросов; C_2 — затраты на пересылку корректирующих сообщений. Формула для C_1 записана в предположении, что при обслуживании запроса к j -му файлу всегда используется копия этого файла из s -го узла ($x_{sj}=1$), которому соответствуют минимальные затраты на пересылку единицы ответной информации в i -й узел, где возникает запрос.

Как видно из соотношения (1), составляющая C_1 характеризуется нелинейной зависимостью от переменных $\{x_{ij}\}$, что усложняет задачу поиска оптимального распределения файлов. Для преобразования целевой функции к линейному виду введем дополнительные двоичные переменные $\{z_{ijk}; i, k=1; j=1, \dots, m\}$, значения которых определяются следующим образом:

$$z_{ijk} = \begin{cases} 1, & \text{если запрос из } i\text{-го узла к } j\text{-му файлу} \\ & \text{будет направлен в } k\text{-й узел;} \\ 0, & \text{в противном случае.} \end{cases}$$

Переменные $\{x_{ij}\}$ и $\{z_{ijk}\}$ связаны между собой условиями

$$x_{ij} + \sum_{k=1}^n z_{ijk} = 1; z_{ijk} \leq x_{jk}. \quad (5)$$

При этом, если j -й файл хранится в i -м узле, то запросы к этому файлу, возникающие в i -м узле, никуда не посылаются ($\sum_{k=1}^n z_{ijk} = 0$ при $x_{ij} = 1$).

В случае отсутствия j -го файла в i -м узле все запросы к этому файлу из i -го узла будут направляться в один из других узлов, в котором хранится требуемый файл ($\sum_{k=1}^n z_{ijk} = 1$ при $x_{ij}=0$ и одновременно для некоторого значения k должно выполняться условие $z_{ijk} x_{kj}=1$).

Обозначив $\theta_{ij} = \alpha_{ij} + \beta_{ij}$, получим оптимизационную задачу из класса задач линейного математического программирования с булевыми переменными:

$$C^* = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^n \lambda_{ij} \theta_{ij} d_{ki} z_{ijk} + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^n \lambda'_{ij} \gamma_{ij} d_{is} x_{sj} \rightarrow \min$$

при ограничениях (2)-(5).

Пример решения задачи оптимального распределения файлов в сети. Рассмотрим компьютерную сеть, состоящую из пяти узлов ($n=5$). Ограничения на размер локальных БД в этих узлах определяются вектором $\{V_j\}=\{16;15;14;11;10\}$. При этом используются соответствующие условные единицы измерения. Расстояния между узлами имеют значения, представленные матрицей

$$D = \|d_{is}\|_{n \times n} = \begin{vmatrix} 0 & 1 & 2 & 4 & 6 \\ 1 & 0 & 1 & 3 & 5 \\ 2 & 1 & 0 & 2 & 4 \\ 4 & 3 & 2 & 0 & 2 \\ 6 & 5 & 4 & 2 & 0 \end{vmatrix}.$$

Предположим, что РБД содержит девять файлов ($m=9$), а размеры этих фай-

лов принимают значения $\{L_i\}=\{6; 3; 2; 4; 3; 1; 5; 3; 4\}$. Потoki данных, передаваемых в процессе функционирования информационной системы, характеризуются матрицами

$$A = \|\lambda_{ij}\|_{n \times m} = \begin{pmatrix} 6 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 \end{pmatrix};$$

$$A' = \|\lambda'_{ij}\|_{n \times m} = \begin{pmatrix} 1 & 1 & 2 & 1 & 0 & 1 & 0 & 2 & 0 \\ 2 & 2 & 0 & 2 & 1 & 1 & 1 & 0 & 2 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 2 \\ 2 & 1 & 2 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 3 & 2 & 1 & 1 & 1 & 0 \end{pmatrix};$$

$$\Theta = \|\theta_{ij}\|_{n \times m} = \begin{pmatrix} 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 3 & 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 3 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 5 & 0 & 2 \end{pmatrix};$$

$$G = \|\gamma_{ij}\|_{n \times m} = \begin{pmatrix} 4 & 3 & 1 & 0 & 0 & 1 & 1 & 2 & 1 \\ 2 & 3 & 0 & 2 & 1 & 1 & 3 & 0 & 1 \\ 3 & 2 & 0 & 2 & 1 & 1 & 2 & 0 & 2 \\ 3 & 1 & 2 & 3 & 0 & 0 & 2 & 2 & 0 \\ 4 & 0 & 0 & 1 & 2 & 2 & 4 & 0 & 0 \end{pmatrix}.$$

Результаты вычислений представлены в виде диаграмм на рисунке. На диаграммах по оси у показано число файлов, по x – номера узлов сети. Полученные оптимальные варианты организации РБД сопоставляются по обобщенным характеристикам, которые оценивают число файлов (в разрезе по узлам сети и типам файлов), а также общие затраты памяти на хранение файлов и степень загруженно-

сти памяти для отдельных узлов компьютерной сети:

$$F_1(i) = \sum_{j=1}^m x_{ij}; F_2(j) = \sum_{i=1}^n x_{ij};$$

$$F_3(i) = \sum_{j=1}^m x_{ij} L_j; F_4(i) = \frac{1}{V_i} \sum_{j=1}^m x_{ij} L_j.$$

Исходные данные варианта 2 отличаются увеличением на порядок (по сравнению с вариантом 1) объема сообщений, пересылаемых при обработке каждого запроса (элементы матрицы θ). В этих условиях с целью снижения общего объема трафика, передаваемого по компьютерной сети, становится целесообразно предусмотреть дополнительные копии для 66,7% файлов (6 из 9). В итоге общие затраты памяти на хранение файлов РБД возрастают в 1,6 раза, а средний уровень загруженности памяти для одного узла компьютерной сети достигает 73%. Вместе с тем, можно ожидать, что такая стратегия дублирования и распределения отдельных фрагментов РБД позволит снизить время реакции информационной системы, работающей в компьютерной сети, за счет возможности параллельной обработки однотипных запросов и локальной доступности большей части данных.

Выводы. Предложен метод балансировки загрузки узлов сети РБД, предназначенной для обработки больших и сверхбольших объемов данных. Результаты расчетов свидетельствуют о возможности значительного увеличения скорости обработки данных в РБД большого объема за счет использования механизмов оптимизации загрузки узлов сети. Применение предложенного метода позволяет повысить производительность и снизить время реакции информационной системы, работающей с РБД.

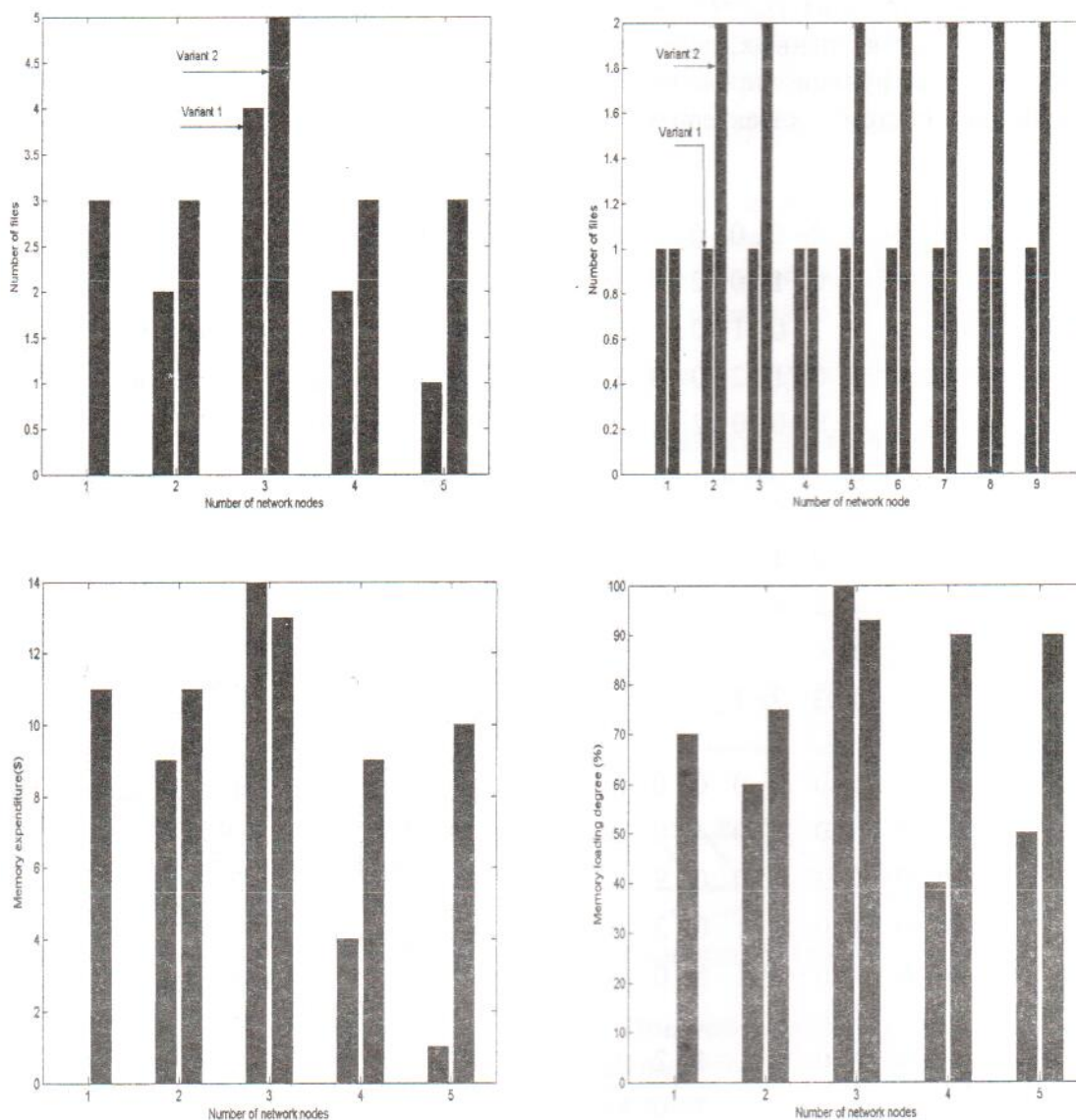


Рис. Диаграммы загрузки узлов компьютерной сети РБД

Список литературы

1. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы. – С.Пб.: Питер, 2004. – 864 с.

2. Аббасов А. М. Оптимизация размещения информационных баз с копиями в сети ЭВМ // Автоматика и вычислительная техника. – 1988. – №4. – С. 71-75.

3. Янбых Г. Ф., Бобер В. И., Бокоев Т. И. Оптимизация размещения файлов и каналов передачи данных в сети ЭВМ // Автоматика и вычислительная техника. – 1984. – №4. – С. 25-29.

4. Цегелик Г. Г. Системы распределенных баз данных. – Львов: Свит, 1990. – 168 с.

5. Кетков Ю. Л., Кетков А. Ю., Шульц М. М. MATLAB 7: программирование, численные методы. – С.Пб.: ВHV-Петербург, 2005. – 752 с.

6. Жуков И. А., Зайченко Ю. П., Печурин Н. К. Модель распределения информационных ресурсов в компьютерных сетях // Проблеми інформатизації та управління. – К.: НАУ, 2005. – Вип. 3 (14). – С. 9-14.