

УДК 004.04

¹Жуков И. А., д-р техн. наук,
²Зайченко Ю. П., д-р техн. наук,
¹Печурин Н. К., д-р техн. наук

МОДЕЛЬ РАСПРЕДЕЛЕНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ В КОМПЬЮТЕРНЫХ СЕТЯХ

¹Институт компьютерных технологий Национального авиационного университета
²Национальный технический университет Украины «КПИ»

Описана простая математическая модель задачи размещения информационных файлов по узлам компьютерной сети с топологией общей шины и соображения по ее обобщению. В качестве критерия оптимальности модели использована минимальная сумма средних времен обслуживания запросов, поступающих в течение единицы времени. Приведено описание двух алгоритмов анализа математической модели.

Введение

Задача оптимального распределения информационных ресурсов в компьютерных сетях в настоящее время снова становится объектом внимания исследователей. Одной из причин этого является бурное внедрение компьютерных сетей, основанных на беспроводных системах передачи данных (*Wi-Fi*), которые характеризуются крайне ограниченным ресурсом пропускной способности. Оптимизация распределения информационных ресурсов позволяет улучшить технико-экономические показатели компьютерной сети, например, – среднее время реакции на запрос пользователя или затраты на достижение заданных показателей качества.

Для оценки эффективности компьютерной сети в целом или ее части можно использовать систему показателей, основанную на наборе показателей качества сервиса эталонной модели ВОС/МОС [1], расширенном показателями, характеризующими стоимость достижения соответствующих показателей для каждого из семи уровней модели взаимодействия открытых систем. Именно использование такой системы показателей дает возможность оценить эффективность (а не лишь уровень качества сервиса) компьютерной сети в целом или ее части. Здесь можно поставить задачу распределения информационных ресурсов как многокритериальную или воспользоваться классическим критерием «затраты/производительность». В обоих случаях перечень показателей эффективности должен включать

оценки стоимости обеспечения соответствующих показателей качества *QoS* [2].

Несмотря на сказанное, рассмотрим сначала некоторые показатели *QoS*, которые могут использоваться непосредственно в качестве критериев эффективности для задачи распределения информационных ресурсов.

На трех верхних уровнях модели ВОС этими показателями могут быть задержка передачи, задержка установления соединения, задержка завершения соединения; на четырех нижних уровнях – транзитная задержка, пропускная способность, скорость передачи. Наиболее часто используются следующие обобщения перечисленных характеристик [3]:

– время реакции – интервал между возникновением запроса пользователя к сетевой службе и получением ответа;

– скорость передачи данных или пропускная способность – объем данных, передаваемых сетью или ее частью в единицу времени;

– задержка передачи – интервал между моментом поступления данных на вход сетевого устройства и появлением данных на его выходе.

Время реакции, таким образом, – это интегральная оценка уровня пользователя (прикладной уровень), скорость передачи и задержка передачи представляют чисто сетевыми параметрами (нижние уровни модели ВОС).

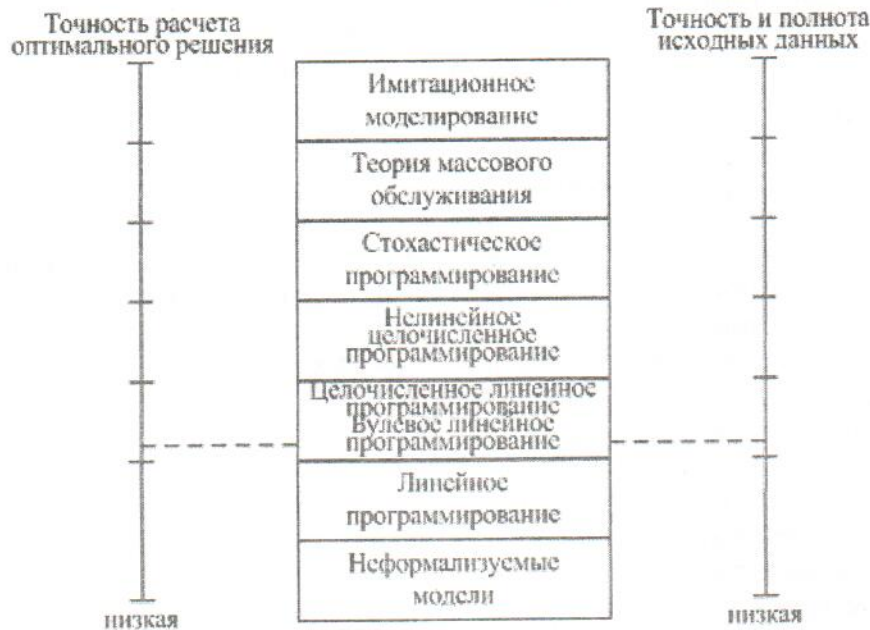
В работах [4–8] были изучены вопросы оптимизации РБД по различным критериям оптимальности. При этом рас-

смаивались алгоритмы размещения информационных файлов по узлам компьютерной сети. Предполагалось, что каждый узел компьютерной сети состоит из компьютера, терминальных устройств и аппаратуры передачи данных; линия или канал связи – из аппаратуры передачи данных, промежуточной аппаратуры и физической среды передачи. В качестве критериев оптимальности рассматривались средний объем данных, пересылаемых по каналам связи в единицу времени, общая стоимость

трафика сети, общее время обработки запросов и др.

Задача оптимального размещения файлов ниже будет рассмотрена как однокритериальная.

После решения вопроса с выбором критерия (критериев) оптимальности необходимо определиться с типом используемых моделей. Тип модели, даже если априори предполагается использование математических моделей, необходимо увязывать с доступностью, полнотой, точностью исходных данных (см. схему ниже).



На начальных этапах жизненного цикла компьютерной сети исходные данные для построения математических моделей характеризуются большой степенью неопределенности [9], которая уменьшается по мере продвижения к этапу реализации. Неопределенность исходных данных выражается как в их (данных) недоступности, так и в стохастическом характере процессов, на основании анализа которых эти данные формируются. Неопределенность исходных данных, которая выражена в вероятностном характере последних, традиционно преодолевается использованием оценок математического ожидания соответствующих величин. Пользоваться этим приемом можно тогда, когда алгоритм анализа (оптимизационной) модели характеризуется «линейными» преобразованиями исходных данных. В противном случае сущест-

вует опасность получения смещенных оценок эффективности распределения информационных ресурсов [10].

Алгоритмы решения задачи оптимального размещения файлов ниже будут рассматриваться как «линейные».

Математическая модель. Рассмотрим задачу оптимального распределения m независимых информационных файлов F_i ($i=1, 2, \dots, m$) РБД на n -узловой компьютерной сети с шинной топологией на основе общего подхода [6]. Топология шины, однотипность линий связи и их небольшая длина в локальных компьютерных сетях позволяют сделать предположение о том, что время пересылки не зависит от того, в каком узле запрос инициирован и в какой узел передан.

Используются следующие обозначения: V_i – объем файла F_i ; U_j – объем

памяти узла K_j для размещения файлов; z – число типов поисковых запросов; λ_{kij} – интенсивность поисковых запросов k -типа к файлу F_i из узла K_j ; c – число типов корректировок; λ'_{lij} – интенсивность корректировок l -типа файла F_i из узла K_j ; T_{kij} – время обработки поискового запроса k -типа к файлу F_i в узле K_j ; $T_{ki}^{(1)}$ – время пересылки поискового запроса k -типа к файлу F_i ; $T_{ki}^{(2)}$ – время пересылки ответа на поисковый запрос k -типа к файлу F_i ; T'_{lij} – время обработки корректировки l -типа файла F_i в узле K_j ; T'_{li} – время пересылки корректировки l -типа файла F_i ;

$$x_{ij} = \begin{cases} 1, & \text{если файл } F_i \text{ находится в узле } K_j; \\ 0, & \text{в противном случае} \end{cases}$$

В качестве критерия оптимальности использовано значение суммы времен, необходимых для обслуживания запросов, поступивших в сеть в течение единицы времени. Такую модель можно использовать для загрузки файлами узлов с более производительными компьютерами сети. Здесь предполагается, что время пересылки запроса по каналам связи зависит только от типа запроса и от файла, к которому адресован запрос. Под запросом подразумевается поисковый или корректирующий запрос к данным, хранящимся в РБД. Запросы обслуживаются в K_j узле ($j=1,2,\dots,n$) в порядке их поступления.

Время пересылки информации из узла K_j при выполнении поискового запроса k -типа к файлу F_i определяется как $(T_{ki}^{(1)} + T_{ki}^{(2)})(1 - x_{ij})$. Тогда значение суммы времен пересылки информации по каналам связи между узлами компьютерной сети при выполнении поисковых запросов, поступивших в сеть в течение единицы времени, равно

$$T = \sum_{k=1}^z \sum_{i=1}^m \sum_{j=1}^n \lambda_{kij} (T_{ki}^{(1)} + T_{ki}^{(2)}) (1 - x_{ij})$$

Если $a_{ij} = \sum_{k=1}^z \lambda_{kij} (T_{ki}^{(1)} + T_{ki}^{(2)})$, то

$$T = T_0 - \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ij}, \text{ где } T_0 \equiv \sum_{i=1}^m \sum_{j=1}^n a_{ij}.$$

Сумма времен пересылки данных из узла K_j при выполнении корректировки

l -типа файла F_i равно $\sum_{\substack{p=1 \\ p \neq j}}^n T'_{li} x_{ip}$. Сумма

времен пересылки данных по каналам связи между узлами при выполнении корректирующих запросов, поступивших в сеть в течение единицы времени, определяется как

$$T' = \sum_{l=1}^r \sum_{i=1}^m \sum_{j=1}^n \sum_{\substack{p=1 \\ p \neq j}}^n \lambda'_{lij} T'_{li} x_{ip}.$$

Если $a'_{ij} = \sum_{l=1}^r \lambda'_{lij} T'_{li}$, то

$$T' = \sum_{i=1}^m \sum_{j=1}^n a'_{ij} \sum_{\substack{p=1 \\ p \neq j}}^n x_{ip} = \sum_{i=1}^m \sum_{j=1}^n x_{ij} \sum_{\substack{p=1 \\ p \neq j}}^n a'_{ip} = \sum_{i=1}^m \sum_{j=1}^n \hat{a}_{ij} x_{ij},$$

где $\hat{a}_{ij} = \sum_{\substack{p=1 \\ p \neq j}}^n a'_{ip}$.

Справедлива формула

$$T_{kij} x_{ij} + \hat{t}_{kij} (1 - x_{ij}) = \hat{t}_{kij} + (T_{kij} - \hat{t}_{kij}) x_{ij},$$

где $\hat{t}_{kij} = \frac{1}{n-1} \sum_{\substack{p=1 \\ p \neq j}}^n T_{kip}$.

Тогда сумма времен обработки всех поисковых запросов, поступивших в сеть в течение единицы времени, определяется как

$$t = t_0 + \sum_{i=1}^m \sum_{j=1}^n U_{ij} x_{ij},$$

где $t_0 \equiv \sum_{k=1}^z \sum_{i=1}^m \sum_{j=1}^n \lambda_{kij} \hat{t}_{kij}$,

$$U_{ij} = \sum_{k=1}^z \lambda_{kij} (t_{kij} - \hat{t}_{kij})$$

Сумма времен выполнения всех запросов на корректировку, поступивших в сеть в течение единицы времени, определяется как

$$t' = \sum_{l=1}^r \sum_{i=1}^m \sum_{j=1}^n \sum_{q=1}^n \lambda'_{lij} T'_{liq} x_{iq} = \sum_{l=1}^r \sum_{i=1}^m \sum_{j=1}^n \sum_{q=1}^n \lambda'_{liq} T'_{lij} x_{ij}.$$

Если положить $U'_{ij} = \sum_{l=1}^r \sum_{q=1}^n \lambda'_{liq} T'_{lij}$, то

$$t' = \sum_{i=1}^m \sum_{j=1}^n U'_{ij} x_{ij}.$$

Задача оптимального распределения файлов по узлам компьютерной сети по критерию минимальной суммы времен, необходимых для обслуживания запросов, поступавших в сеть в течение единицы времени, решается с использованием оптимизационной модели вида [11, 12]:

$$f(X) = f_0 + \sum_{i=1}^m \sum_{j=1}^n C_{ij} x_{ij} \rightarrow \min \quad (1)$$

при ограничениях

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

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

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

где $f_0 \equiv T_0 + t_0$, $C_{ij} = -a_{ij} + \hat{a}_{ij} + U_{ij} + U'_{ij}$.

Задача (1)–(4) относится к классу задач целочисленного линейного математического программирования с булевыми переменными.

Необходимо отметить, что использование в качестве критерия оптимальности минимизацию времени, необходимого для обслуживания запросов, приведет к модели, кардинально отличающейся от описанной выше и не классифицируемой как линейная.

Алгоритмическая реализация. Эвристический алгоритм. Для решения представленной задачи разработаны ряд алгоритмов [12]. Опишем двухэтапный эвристический алгоритм исследования представленной выше модели. На первом этапе алгоритма находится начальное распределение файлов, которое будет оптимальным, если не учитывать условия (3). На втором этапе проводится перераспределение файлов, если для начального распределения существует хотя бы один индекс j , при котором условие (3) не выполняется. На каждом шаге второго этапа осуществляется перераспределение одного файла из переполненного узла таким образом, чтобы достичь минимального увеличения целевой функции. Выполнение второго этапа продолжается до тех пор, пока не будет найдено

распределение, удовлетворяющее условию (3). Ниже дано описание алгоритма.

Этап определения начального распределения файлов по узлам сети.

1. Сформировать матрицу целевой функции $C(m, n)$, где $C = \{C_{ij}\}_{mn}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).

2. Если для $i \exists C_{ij} < 0$, то

$$x_{ij}^* = \begin{cases} 1, & C_{ij} < 0; \\ 0, & C_{ij} \geq 0. \end{cases}$$

3. Если для $i \forall C_{ij} \geq 0$, то найти

$\min_{1 \leq j \leq n} C_{ij}$. Пусть $\min_{1 \leq j \leq n} C_{ij} = C_{ji}$. Тогда

$$x_{ij}^* = \begin{cases} 1, & j = j_i; \\ 0, & j \neq j_i. \end{cases}$$

Этап перераспределения файлов.

1. Сформировать вектор признаков $E = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$, где $\epsilon_j = 0$ ($j = 1, 2, \dots, n$).

В процессе работы алгоритма, после перераспределения файлов из некоторого переполненного узла K_j соответствующему элементу ϵ_j вектора E присваивается значение 1 и узел считается закрытым для перераспределения.

2. Для всех индексов j , для которых $\epsilon_j = 0$, проверить выполнение условия (3). Если оно выполняется, то работа алгоритма заканчивается. Если для некоторого $j = r$ имеем $\sum_{i=1}^m L_i x_{ir} > U_r$, то перейти на шаг 3.

3. Если $\exists z \neq r$ такой, что $\sum_{i=1}^m L_i x_{ir} \leq U_z$, $\sum_{i=1}^m L_i x_{iz} \leq U_r$, то поменять местами память r -го и z -го узлов и возвратиться на шаг 2. В противном случае перейти на шаг 4.

4. Для тех i , для которых $\exists j \neq r$ такое, что $x_{ij} = 1$, найти $\min_i (-C_{ir})$. Пусть $\min_i (-C_{ir}) = -C_{kr}$. Для тех i , для которых $x_{ij} = \begin{cases} 1, & j = r, \\ 0, & j \neq r, \end{cases}$ найти $\min_j (C_{ij} - C_{ir})$, где минимум берется по тем индексам $j \neq r$, для которых $\epsilon_j = 0$. Пусть

$\min_j(C_{ij} - C_{ir}) = C_{ij} - C_{ir}$. Тогда найти $\min_i(C_{ij} - C_{ir})$. Пусть $\min_i(C_{ij} - C_{ir}) = C_{ij} - C_{ir}$. Если $\min(-C_{kr}, C_{ij} - C_{ir}) = -C_{kr}$, тогда в матрице X полагаем $x_{kr} = 0$, что означает удаление файла F_k из узла K_r . Если $\min(-C_{kr}, C_{ij} - C_{ir}) = C_{ij} - C_{ir}$, тогда в матрице X полагаем $x_{ir} = 0$, $x_{ij} = 1$. Это означает, что файл F_i из узла K_r перераспределен в узел K_j . Такому перераспределению соответствует минимальное увеличение значения целевой функции.

5. Проверить условие (3) для $j = r$. При выполнении условия элементу ϵ_r вектора E присвоить значение 1 и перейти на шаг 2. Если условие не выполняется, то перейти на шаг 3.

Проведенные статистические испытания показывают, что для реальных задач полученное решение незначительно отличается от оптимального.

Точный алгоритм оптимального распределения файлов. Для решения задачи (1)-(4) можно предложить также и точный алгоритм, использующий метод последовательного анализа и отсева вариантов (ПАВ).

Как известно, метод ПАВ включает две процедуры отсева: отсев по ограничениям W_1 и отсев по критерию W_2 [7].

Рассмотрим его реализацию.

Пусть уже проведено s итераций и построено частичное множество вариантов $*_{\Delta} = \{x_{ij\Delta}\}$. Обозначим через $*_{\Delta 0}$ - подмножество переменных $x_{ij}^{(s)} = 0$; $*_{\Delta 1}$ - подмножество переменных $x_{ij}^{(s)} = 1$; а через $*_{\Delta}$ - подмножество оставшихся переменных, значения которых еще не фиксированы, т. е. $x_{ij} \in \{0,1\}$ если $x_{ij} \in *_{\Delta}$.

Опишем $(\Delta+1)$ -ю итерацию.

Процедура W_1 .

Процедура отсева для переменной x_{ij} по ограничению (2) запишется в виде

$$x_{ij} < 1 - \max_{k \neq j} x_{ik} = 1 - \sum_{k \neq j} x_{ik} - \sum_{r \neq j} x_{ir} \quad (5)$$

$x_{ik} \in *_{\Delta 1} \quad x_{ir} \in *_{\Delta}$

Отсеивается значение $\overline{x_{ij}} = 0$, если правая часть больше нуля, и нет отсева, если она меньше нуля.

Процедуру отсева (5) повторяем со всеми переменными x_{ij} , $j = \overline{1, n}$ и $i = \overline{1, m}$. Далее переходим к процедуре отсева по ограничению (3).

$$L_i x_{ij} > U_j - \min_{k \neq i} \sum L_k x_{ikj} = U_j - \sum L_k \quad (6)$$

$k : x_{kj} \in *_{\Delta 1}$
 $k \neq i$

Отсеивается $\overline{x_{ij}} = 1$, если правая часть (6) больше нуля, но меньше L_i . Если же (6) меньше нуля, то нет допустимых вариантов $*_{\Delta} = 0$.

Указанную процедуру повторяем со всеми ограничениями $j = \overline{1, n}$. В результате получим сокращенное подмножество $*_{\Delta}(1)$. Переходим к процедуре отсева по величине критерия W_2 .

Процедура W_2 .

Задаемся порогом отсева $f_{\Delta+1}^*$.

$$\text{Для } \Delta = 1 \quad f_1^* = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n C_{ij}$$

На итерации $\Delta+1$ величину порога выбираем так:

1) $f_{\Delta+1}^* = \frac{1}{2}(f_{\Delta}^* + f_{\Delta \min})$ - для сужения множества вариантов;

2) $f_{\Delta+1}^* = \frac{1}{2}(f_{\Delta}^* + f_{\Delta \max})$ - для расширения множества вариантов;

3) $f_{\Delta+1}^* = \frac{1}{2}(f_{\Delta-1}^* + f_{\Delta}^*)$ - для сужения после расширения и наоборот.

Отсев для переменной x_{ij} выглядит так:

$$C_{ij} x_{ij} \rightarrow f_{\Delta} - (f_0 + \min_i \sum_j C_{rj} x_{rj}) = f_{\Delta} - f_0 - \sum_{k_j : x_{kj} \in *_{\Delta 1}} C_{kj} \quad (7)$$

$k \neq 0$

Отсеивается значение $\overline{x_{ij}} = 1$, если правая часть меньше C_{ij} , но больше нуля.

Нет отсева, если правая часть больше C_{ij} .

И, наконец, нет допустимых вариантов, если правая часть (7) меньше нуля.

Процедуру отсева (7) повторяем со всеми ограничениями (3).

Если отсева нет, то снова переходим на процедуру W_2 , выбрав $f_{\Delta+1}^*$ для сужения множества вариантов. Если же нет допустимых вариантов, то снова на процедуру W_2 , выбрав $f_{\Delta+1}^*$ из условия расширения вариантов.

В случае если отсев есть и есть допустимые варианты, то идем на процедуру W_1 . Переходим к $(\Delta + 2)$ -й итерации.

Последовательность итераций повторяем до тех пор, пока на некоторой итерации k не получим также сокращенные подмножества вариантов $*_k$ из которых искомым можно выбрать путем простого перебора вариантов.

Выводы

1. Критерии для оценки эффективности распределения информационных ресурсов в компьютерной сети следует выбирать из набора качества сервиса эталонной модели ВОС/МОС, расширенном показателями, характеризующими стоимость достижения соответствующего уровня качества.

2. Простая оптимизационная модель задачи размещения файлов РБД в компьютерной сети с топологией общей шины может применяться в тех случаях, когда исходные данные характеризуются большой степенью неопределенности.

3. Эвристические или точные алгоритмы решения даже простой, описанной выше задачи, могут дать смещенные оценки эффективности распределения информационных ресурсов.

4. На последних этапах жизненного цикла компьютерной сети при моделировании целесообразно применять более точные, нежели описанная, модели (модели стохастического программирования, теории массового обслуживания, имитационные и др.).

Список литературы

1. Зайцев С. С., Кравцунов М. И., Ротанов С. В. Сервис открытых информационно-вычислительных сетей: Справочник. – М.: Радио и связь, 1990. – 240 с.

2. Печурин Н. К. и др. Диалоговая система проектирования систем телеобработки ДИСТЕЛ // В кн. Каталог диалоговых систем: Материалы по математическому обеспечению. – К.: ИК АН УССР, 1986. – 232 с.

3. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 2-е изд. – С.Пб.: Питер, 2003. – 864 с.

4. Клименко С. А., Стрекалов А. А. Декомпозиционный алгоритм оптимизации размещения информационных ресурсов в сетях ЭВМ с зональной структурой // Автоматика и вычислительная техника. – 1988. – №6. – С.77–81.

5. Ладыженский Ю. В., Бельков Д. В. Рациональное размещение файлов распределенной базы данных в вычислительной сети с произвольной топологией // 36. наукових праць ДДТУ. Серія “Проблеми моделювання та автоматизації проектування динамічних систем”. – Вип.6. – Донецьк: ДДТУ, 1999. – С.44–49.

6. Цегелик Г. Г. Системы распределенных баз данных. – Львов: Світ, 1990. – 168 с.

7. Зайченко Ю. П., Гонта Ю. В. Структурная оптимизация сетей ЭВМ. – К.: Техніка, 1986. – 168 с.

8. Янбух Г. Ф. Оптимизация размещения вычислительных комплексов, программ и файлов в сети ЭВМ // Автоматика и вычислительная техника. – 1984. – №5. – С. 14–20.

9. Бандуренко В. О., Печурин М. К. Модели прийняття рішень при проектуванні систем телеобробки даних в умовах невизначеності // Вісн. Київ. політехн. інту. Техн. кібернетика. – Вип. 10. – К.: Вища школа, 1986. – С. 148–153.

10. Печурин М. К., Кондратова Л. П. Аналіз похибок алгоритмів оптимізації топології інформаційної мережі // Наукові вісті. – К.: НТУУ “КПІ”, 2001. – №3.

11. Лэссон Л. Оптимизация больших систем. – М.: Наука, 1975. – 432 с.

12. Таха Х. Введение в исследование операций, 7-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 912 с.

13. Зайченко Ю. П. Комп'ютерні мережі. – К.: Слово, 2003. – 256 с.