

ВЫБОР ОБЪЕМА ПАМЯТИ КОМПЬЮТЕРОВ ДЛЯ ОПТИМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ФАЙЛОВ В КОМПЬЮТЕРНОЙ СЕТИ

Ладыженский Ю.В.

Кафедра ПМиИ ДонНТУ

ladyzhen@pmi.donetsk.ua

Бельков Д.В.

Кафедра ВМиП ДонНТУ

Abstract

Ladyzhensky Y.V., Belkov D.V. The choice of computer memory volume for optimal file distribution in the computer network. An important practical task, arising up on the stage of design a computer network, which consists in the choice of computer memory volume for optimal file distribution in the computer network, is decided in article.

Введение

При проектировании распределенной системы необходимо стремиться к рациональному размещению файлов по узлам компьютерной сети. Если задача рационального размещения файлов не будет решена, то это может привести к перегрузке каналов связи и неэффективной работе распределенной системы. Поэтому одной из целей проектирования компьютерной сети является синтез оптимального плана размещения m файлов по n узлам сети с учетом особенностей функционирования распределенной системы [1]. В работах [2,3] показано, что рациональное размещение файлов может быть получено жадными методами. Однако влияние объема памяти узлов на качество размещения не исследовано.

Если память узлов не ограничена, то размещение файла i не зависит от размещения файла $(i-1)$ и матрица размещений представляет собой систему независимых векторов (матроид). Известно [4], что строго точное решение задачи на матроиде определяется с помощью жадного метода. Если память узлов ограничена, то жадный метод находит лишь приближенное решение задачи размещения файлов. Относительную погрешность метода можно вычислить по формуле:

$$Q \leq (M - A) / M, \text{ где}$$

M - максимально возможное решение задачи (решение задачи на матроиде), A - приближенное решение задачи, получаемое жадным методом. Чрезмерно большой запас памяти для размещения файлов в узлах экономически невыгоден. Поэтому важной задачей, возникающей при проектировании распределенных систем, является выбор оптимального объема памяти узлов компьютерной сети, который обеспечивает минимальную погрешность решения задачи распределения файлов по

узлам. В данной статье предложены формулы для расчета объема памяти компьютеров, который обеспечивает решение задачи распределения файлов по узлам с заданной погрешностью.

Формулы для расчета объема памяти компьютеров

Пусть объемы файлов и узлов являются целыми величинами. Обозначим: m - число файлов, n - число узлов, V_i - объем файла i , B_j - объем узла j , $X_{ij} = 1$, если файл i размещен в узел j , иначе - $X_{ij} = 0$,

$$U_j = \sum_{i=1}^m V_i X_{ij}, \quad SV = \sum_{i=1}^m V_i, \quad SB = \sum_{j=1}^n B_j.$$

Распределение m файлов по n узлам возможно, если $\sum_{j=1}^n B_j \geq \sum_{i=1}^m V_i$.

Аналогичным свойством в теории комбинаторного анализа обладают разбиения чисел [5]. Разбиением числа SV на m частей является сумма

$$\sum_{i=1}^m V_i, \quad \text{разбиением числа } SB \text{ на } n \text{ частей является сумма } \sum_{j=1}^n B_j.$$

Разбиение SV вложимо в разбиение SB , если части разбиения SV можно так сгруппировать в n групп (каждая часть входит ровно в одну группу), что после сложения всех частей каждой группы, получится n чисел U_1, U_2, \dots, U_n таких, что $U_j \leq B_j, j=1, 2, \dots, n$.

Таким образом, условие возможности распределения m файлов по n узлам совпадает с условием вложения разбиения SV в разбиение SB . Поэтому для оценки качества распределения файлов можно использовать экстремальные свойства разбиений чисел, доказанные в работе [6]:

Утверждение 1. Если SB_{opt} - наименьшее SB , при котором каждое разбиение SV на m частей, вложимо в каждое разбиение SB на n частей, то

$$SB_{opt} = \max\{n \cdot SV - mn + 1, SV\} \quad (1)$$

Эта формула позволяет определить минимальную суммарную память узлов, обеспечивающую строго точное решение задач распределения m файлов по n узлам.

Утверждение 2. Если SB_p - наибольшее SB , при котором для каждого разбиения SV на m частей, найдется разбиение, которое не вложимо в разбиение SB на n частей, то

$$SB_p = (SV \lfloor SV / m \rfloor + n - 1) \times \lfloor SV / m \rfloor \quad (2)$$

Эта формула позволяет определить максимальную суммарную память узлов, которая не обеспечивает решение задач распределения m файлов по n узлам.

Пусть все узлы имеют одинаковый объем $B = SB/n$, $B_p = SB_p/n$, $B_{opt} = SB_{opt}/n$.

Согласно утверждениям 1 и 2, зависимость относительной погрешности распределения файлов от памяти узлов имеет вид, показанный на рисунке 1. Если $B \geq B_{opt}$, то задачи распределения файлов принадлежат оптимальной зоне - любая задача решается строго точно ($Q=0$). Если $B \leq B_p$, то задачи распределения файлов принадлежат запретной зоне - не всякая задача может быть решена ($Q=1$). Если $B_p < B < B_{opt}$, то задачи распределения файлов принадлежат переходной зоне - любая задача может быть решена приближенно ($0 < Q < 1$).

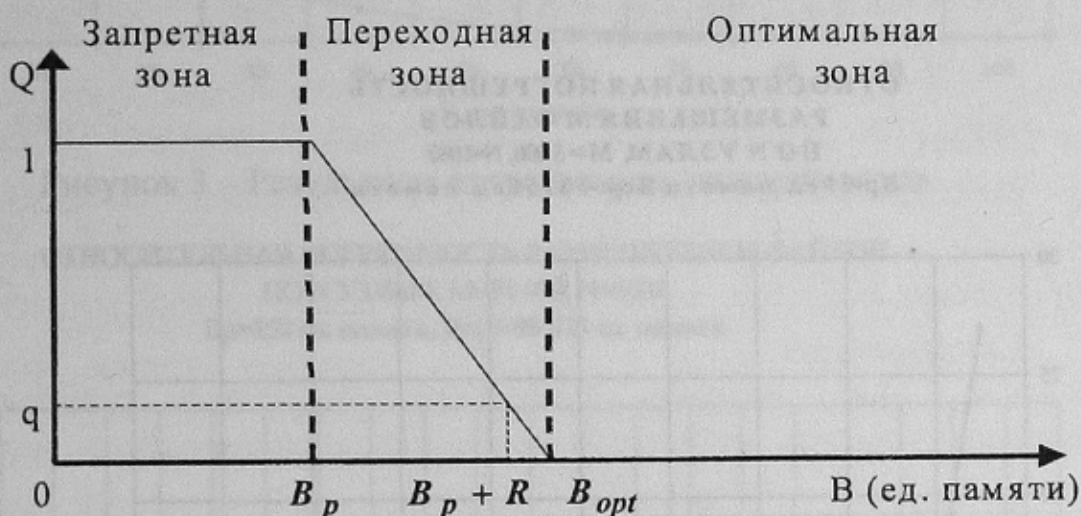


Рисунок 1 - Зависимость относительной погрешности распределения файлов от объема памяти узлов

Пусть необходимо распределить файлы так, чтобы относительная погрешность распределения не превышала допустимой величины q . Для решения этой задачи нужно найти такое значение R , что память узлов $B_q = SB_p/n + R$ обеспечивает распределение файлов с погрешностью не более q . Решение задачи следует из построений, показанных на рисунке 1:

$$R = (1 - q) \cdot (SB_{opt} - SB_p) / n$$

$$B_q = SB_p / n + (1 - q) \cdot (SB_{opt} - SB_p) / n$$

Результаты вычислительных экспериментов

Для проверки полученных теоретических результатов проведены вычислительные эксперименты. В каждой серии экспериментов решены задачи распределения m файлов по n узлам жадным методом А2, предложенным в работе [7].

На первом этапе метода определяются те узлы, в которые файл помещается по размеру. На втором этапе, среди найденных узлов определяется узел с наибольшей интенсивностью запросов, и файл размещается в этот узел. Временная сложность метода - $O(mn)$.

Результаты вычислительных экспериментов подтверждают зависимость относительной погрешности распределения файлов от памяти узлов, показанную на рисунке 1. На рисунках 2 – 4 показано, что в переходной зоне относительная погрешность решения снижается с увеличением размера памяти компьютеров.

ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТЬ
РАЗМЕЩЕНИЯ ФАЙЛОВ
ПО N УЗЛАМ, $M=5000$, $N=1000$
 $V_p=19$ ед. памяти, $V_{opt}=10156$ ед. памяти

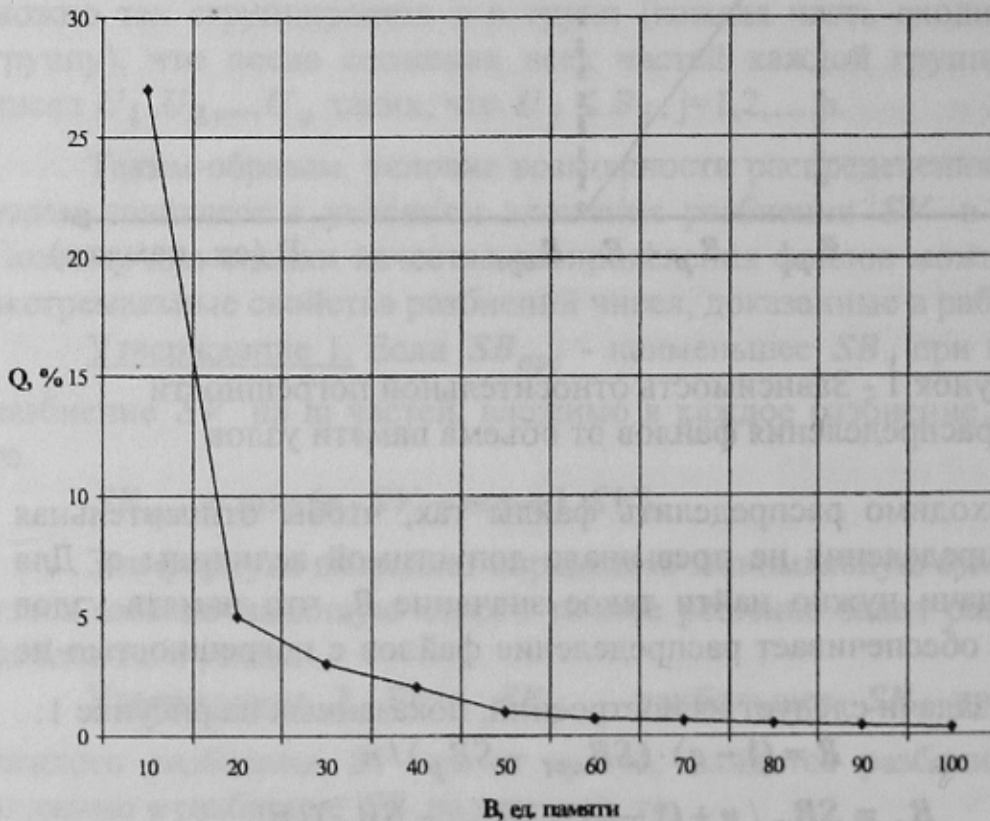


Рисунок 2 – Результаты первой серии экспериментов

ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТЬ РАЗМЕЩЕНИЯ М ФАЙЛОВ

ПО N УЗЛАМ, M=10 000, N=1000

Vp=34 ед. памяти, Vopt=20 573 ед. памяти

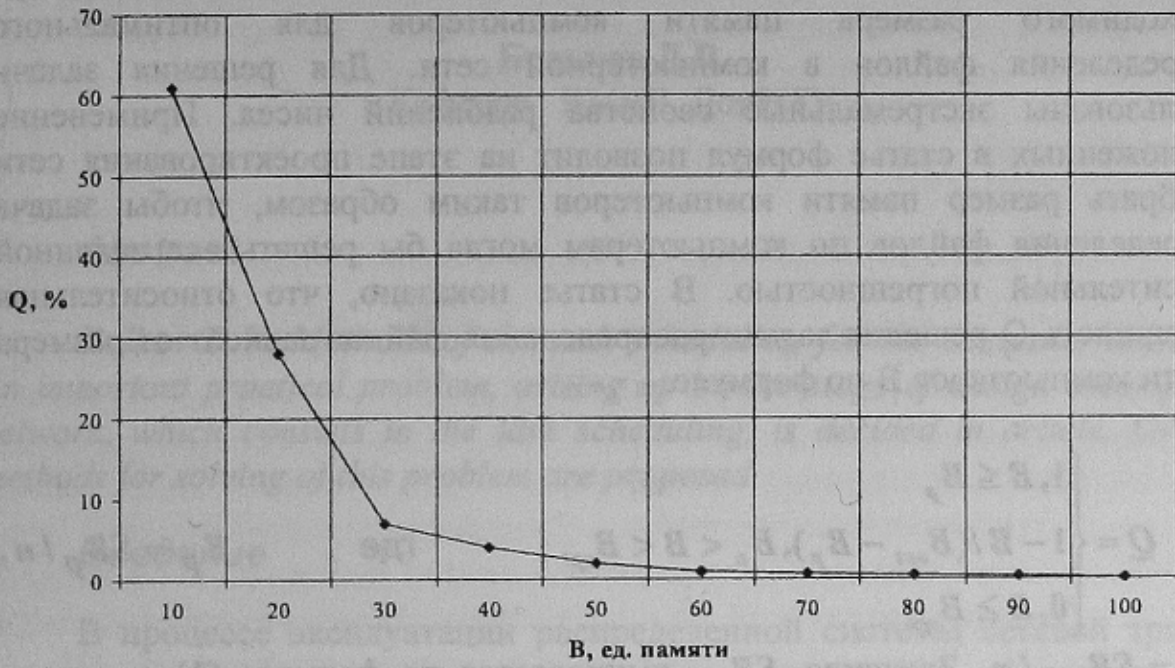


Рисунок 3 – Результаты второй серии экспериментов

ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТЬ РАЗМЕЩЕНИЯ М ФАЙЛОВ

ПО N УЗЛАМ, M=50 000, N=1000

Vp=152 ед. памяти, Vopt=99 023 ед. памяти

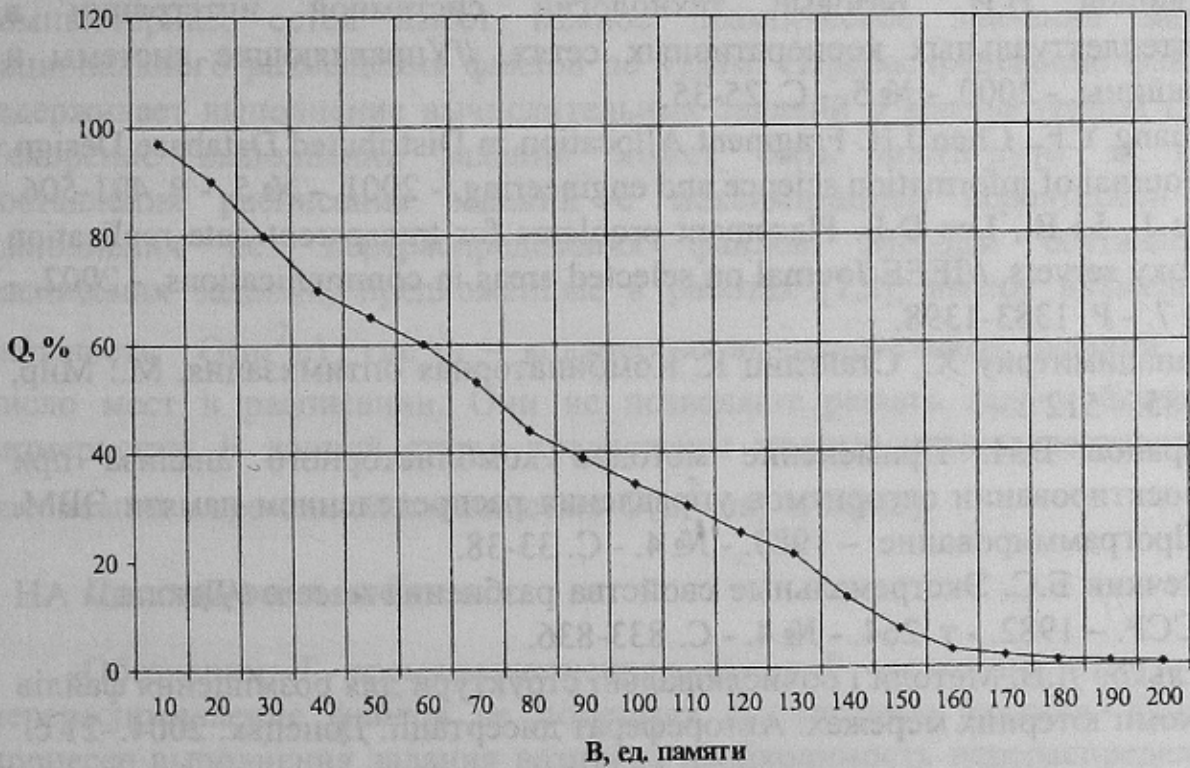


Рисунок 4 – Результаты третьей серии экспериментов

Заключение

В статье решена важная практическая задача, возникающая на этапе проектирования распределенных систем, которая заключается в выборе необходимого размера памяти компьютеров для оптимального распределения файлов в компьютерной сети. Для решения задачи использованы экстремальные свойства разбиений чисел. Применение предложенных в статье формул позволит на этапе проектирования сети подобрать размер памяти компьютеров таким образом, чтобы задача распределения файлов по компьютерам могла бы решаться с заданной относительной погрешностью. В статье показано, что относительная погрешность Q решения задачи распределения файлов зависит от размера памяти компьютеров B по формуле:

$$Q = \begin{cases} 1, & B \leq B_p \\ 1 - B / (B_{opt} - B_p), & B_p < B < B_{opt} \\ 0, & B \geq B_{opt} \end{cases}, \quad \text{где} \quad B_p = SB_p / n,$$

$B_{opt} = SB_{opt} / n$. Значение SB_{opt} вычисляется по формуле (1), значение SB_p вычисляется по формуле (2). Результаты вычислительных экспериментов подтверждают зависимость относительной погрешности распределения файлов от памяти компьютеров, полученную теоретически.

Литература

1. Алишов Н.И. Базовые технологии системной интеграции в интеллектуальных корпоративных сетях. //Управляющие системы и машины. - 2000. - № 5. - С. 25-35.
2. Huang Y.F., Chen J.H. Fragment Allocation in Distributed Database Design. //Journal of information science and engineering. - 2001. - № 5. - P. 491-506.
3. Xu J., Li B., Lee D.L. Placement problems for transparent data replication proxy servers. //IEEE Journal on selected areas in communications. - 2002. - № 7. - P. 1383-1398.
4. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. М.: Мир, 1985. - 512 с.
5. Баранов В.И. Применение методов комбинаторного анализа при проектировании алгоритмов управления распределением памяти ЭВМ. //Программирование. - 1985. - № 4. - С. 33-38.
6. Стечкин Б.С. Экстремальные свойства разбиения чисел. //Доклады АН СССР. - 1982. - т. 264. - № 4. - С. 833-836.
7. Бельков Д.В. Методи і обчислювальні структури для розміщення файлів в комп'ютерних мережах. Автореферат дисертації. Донецьк: 2004.-21 с.