

РАСПРЕДЕЛЕНИЕ ПРИОРИТЕТОВ ПАКЕТОВ И ОПТИМИЗАЦИЯ СТРУКТУРЫ ЦЕНТРАЛИЗОВАННЫХ ЛВС

Проблема оптимального распределения приоритетов сообщений и пропускных способностей магистральных каналов в централизованных локальных вычислительных сетях формулируется как задача математического программирования комбинированного типа, в которой оптимальному назначению приоритетов соответствуют целочисленные (булевы) переменные, а распределению пропускных способностей – непрерывные (континуальные) величины. Предлагается последовательный алгоритм распределения приоритетов и пропускных способностей и его программная реализация.

1. Постановка задачи. Существует ряд проблем рациональной организации производственных, распределительных и управляющих систем, связанных с необходимостью обработки информации, поступающей от удаленных пользователей по каналам связи в центральный вычислитель (сервер). К таким системам можно отнести, например, WEB-сервер или сервер баз данных. Эта система обработки обычно организована в виде централизованной локальной вычислительной сети (ЛВС) звездообразной структуры, в центре которой расположен сервер, как, например, в ЛВС типа Gigabit Ethernet [1]. Структура вычислительной сети этого типа изображена на рисунке 1.

Рисунок 1 – Структура централизованной ЛВС

Узел i , обслуживающий обычно группу пользователей, подключен к серверу через канал с пропускной



способностью C_i и направляет в этот канал через его буфер многомерный поток сообщений (пакетов) интенсивностью $\lambda_i = (\lambda_{i1}, \dots, \lambda_{is}, \dots, \lambda_{iN})$, где λ_{is} – интенсивность потока сообщений s -го типа. Длина (например, в байтах) сообщения s -го типа ($s = 1, \dots, N$) – случайная величина с известными средним I_s и вторым начальным моментом $I_s^{(2)}$; эта длина не зависит от номера i узла. Каждое такое сообщение предварительно помещается в s -ую секцию буфера канала и затем направляется в канал из соответствующей очереди в порядке FIFO.

По приходу в центр оно помещается в s -ую секцию буфера концентратора и оттуда также в порядке FIFO, поступает как заявка на обслуживание в сервер. Трудоемкость обслуживания этой заявки – случайная величина с известными средним θ_s и вторым начальным моментом $\theta_s^{(2)}$, ($s = 1, \dots, N$).

Порядок обслуживания (сканирования) буферов каналов и концентратора определяется дисциплиной относительных приоритетов, устанавливаемых с целью минимизации потерь. Потери оцениваются в соответствии с функцией

$$A = \sum_{i=1}^M \sum_{s=1}^N \alpha_s \lambda_{is} (w_s + w_{is}), \quad (1)$$

где α_s – потери (штраф) за единицу времени ожидания в очередях к каналу и к серверу одной заявки S -го типа, w_s – среднее время ожидания начала обслуживания этой заявки в процессоре, w_{is} – среднее время её ожидания в буфере i -го канала. Предполагается, что величины w_s и w_{is} определены для стационарного режима функционирования как сервера, так и всех каналов ЛВС, т.е. режима, в котором загрузка каждого из указанных элементов ЛВС не превышает единицы. Предполагается также, что потери заявок и сообщений по причине нехватки памяти для их размещения в буферах пренебрежительно малы.

Для аренды каналов ЛВС на некоторый период времени выделена известная сумма D , а также известны величины d_i , $i = 1, \dots, M$, где d_i – стоимость аренды канала i единичной пропускной способности на этот же период.

Необходимо минимизировать ожидаемые потери (1) путем оптимального назначения приоритетов сообщениям как при передаче их по каналам связи, так и при обработке в сервере, а также оптимального распределения пропускных способностей каналов ЛВС.

Отметим, что подобная задача рассматривалась нами в работе [2], однако там приоритеты сообщений предполагались фиксированными, и речь шла лишь о распределении пропускных способностей каналов.

2. Задача в терминах математического программирования. Введём следующие переменные: квадратные матрицы X и Y порядка N и вектор C размерности M , определив их элементы x_{sk} , y_{sk} и C_i следующим образом:

$x_{sk} = 1$, если заявке s -го типа присваивается k -й приоритет при обслуживании в процессоре сервера; $x_{sk} = 0$, иначе.

$y_{sk} = 1$, если заявке s -го типа присваивается k -й приоритет передачи по каналу ЛВС; $y_{sk} = 0$, иначе.

C_i – пропускная способность i -ого канала ($C_i \geq 0$).

Задача состоит в отыскании значений X, Y, C , минимизирующих потери

$$A(X, Y, C) = \sum_{i=1}^M \sum_{k=1}^N \alpha_k \lambda_{ik} (w_k(X) + w_{ik}(Y, C)) \quad (2)$$

при условиях:

$$\sum_{s=1}^N x_{sk} \leq N, \quad k = 1, \dots, N; \quad (3)$$

$$\sum_{k=1}^N x_{sk} = 1, \quad s = 1, \dots, N; \quad (4)$$

$$\sum_{s=1}^N y_{sk} \leq N, \quad k = 1, \dots, N; \quad (5)$$

$$\sum_{k=1}^N y_{sk} = 1, \quad s = 1, \dots, N; \quad (6)$$

$$x_{sk}, y_{sk} \in \{0, 1\}; \quad (7)$$

$$\sum_{i=1}^M d_i C_i \leq D; \quad (8)$$

$$C_i \geq 0, \quad i = 1, \dots, M. \quad (9)$$

Условия (3) – (6) означают, что каждому типу заявок (сообщений) присваивается свой приоритет. Условие (7) означает, что переменные X, Y целочисленные, а точнее – двоичные.

Из выражений (2)–(9) следует, что по своей математической постановке рассматриваемая оптимизационная задача является задачей математического программирования. Будем называть вектор (X, Y, C) , удовлетворяющий условиям (3) – (9), допустимым распределением. Допустимое распределение, минимизирующее функцию (2), назовём оптимальным распределением, имея в виду распределение приоритетов по типам заявок как в каналах передачи, так и в концентраторе, а также распределение пропускных способностей по каналам локальной вычислительной сети.

Постановка задачи допускает два различных варианта её решения.

В первом варианте пропускная способность каналов локальной вычислительной сети может принимать только дискретные значения. В этом случае задачу можно отнести к классу задач математического программирования, в котором все переменные – целочисленные. Для решения задачи можно применить методы комбинаторного поиска, такие как полный и граничный перебор. Метод полного перебора даёт оптимальный по установленному критерию результат, но при этом для получения решения требуется наибольшее время. Количество комбинаций имеет порядок $K^M \cdot N! \cdot N!$ и даже при умеренных K, M и N , представляющих практический интерес, астрономически велико (например, при $K=7, N=5, M=10$ оно составляет величину порядка 10^{12}).

Модификацией метода полного перебора является граничный перебор, который позволяет сократить количество рассматриваемых вариантов за счет применения различных критериев оценки текущего решения и отсека неперспективных вариантов. Основная проблема данного метода заключается в выборе способа оценки конечного результата по промежуточным данным, поскольку последний может быть и не строго оптимальным.

Второй вариант постановки задачи предполагает, что пропускные способности каналов ЛВС – непрерывные величины, и тогда имеем оптимизационную задачу математического программирования

комбинированного типа. Эта задача, вообще говоря, более сложна. Однако можно предположить, что оптимизацию пропускных способностей каналов ЛВС и относительных приоритетов различных типов заявок допустимо выполнять отдельно. В таком случае можно рассматривать две независимые задачи. Задача оптимизации пропускных способностей каналов ЛВС представляет собой поиск минимального значения функции при ограничении-равенстве, а потому является классической задачей Лагранжа [2]. Решение задачи назначения относительных приоритетов различным типам заявок зависит от характеристик системы массового обслуживания [3]. Особенность системы М/Г/1 заключается в том, что оптимальный порядок обслуживания не зависит от интенсивности входящих потоков и определяется лишь первыми двумя моментами длительности обслуживания и весовыми коэффициентами (штрафами). В представляемой здесь оптимизируемой модели предполагается, что все потоки, передаваемые по каналам, – простейшие.

3. Синтез относительных приоритетов. Рассмотрим принцип назначения приоритетов заявкам. Важной особенностью задачи является то, что каждый из M каналов может рассматриваться как система массового обслуживания типа М/Г/1 с относительными приоритетами. Каждый канал определяет время обслуживания заявки, как случайную величину с известными средним l_k / C_i и вторым начальным моментом обслуживания $l_k^{(2)} / C_i^2$ в системе i . Время обслуживания каждой заявки процессором сервера – случайная величина с известными средним ϑ_k и вторым начальным моментом $\vartheta_k^{(2)}$, $k = 1, \dots, N$.

Задача сводится, во-первых, к присвоению приоритетов типам заявок при поступлении их в процессор, а, во-вторых, к назначению приоритетов различным типам заявок в очереди к каналу. Другими словами, необходимо найти оптимальные квадратные матрицы X и Y порядка N , удовлетворяющие условиям (3) – (7).

Приоритеты различным типам заявок в очереди к процессору сервера должны быть назначены по правилу [3] $\alpha_j / \vartheta_j \geq \alpha_{j+1} / \vartheta_{j+1}$, $j = 1, \dots, N-1$, где α_j – потери (штраф) за единицу времени ожидания к процессору одной заявки j -го типа, ϑ_j – время обработки заявки j -го типа в процессоре, $\vartheta_j = \theta_j / B$, где B – быстродействие процессора.

Для каждого канала ЛВС необходимо установить свои приоритеты заявок так же, по правилу [3]:

$$\alpha_j / l_j / C_i \geq \alpha_{j+1} / l_{j+1} / C_i, \quad j = 1, \dots, N-1, \quad i = 1, \dots, M. \quad (10)$$

Как видно из (10), назначение приоритетов различным типам заявок в каналах передачи не зависит от номера канала и его пропускной способности. Таким образом, приоритеты однотипных заявок в различных каналах одинаковы.

4. Определение эффективности функционирования ЛВС. После расчета матриц X , Y и вектора C необходимо вычислить минимальные суммарные потери (2), где средние времена ожидания $w_k(X)$ и $w_{ik}(Y, C)$ заявок k -го приоритета определены как

$$w_k(X) = \frac{\sum_{s=1}^N \lambda_s \cdot \theta_s^{(2)}}{2 \cdot (B - \sum_{j=1}^k \sum_{s=1}^N \theta_s \cdot \lambda_s \cdot x_{sj}) \cdot (B - \sum_{j=k+1}^N \sum_{s=1}^N \lambda_s \cdot \theta_s \cdot x_{sj})}, \quad (11)$$

$$w_{ik}(Y, C) = \frac{\sum_{s=1}^N \lambda_{is} \cdot l_s^{(2)}}{2 \cdot (C_i - \sum_{j=1}^k \sum_{s=1}^N l_s \cdot \lambda_{is} \cdot y_{sj}) \cdot (C_i - \sum_{j=k+1}^N \sum_{s=1}^N \lambda_{is} \cdot l_s \cdot y_{sj})}, \quad (12)$$

где B – быстродействие процессора; λ_s – суммарная интенсивность потока заявок s -го типа, обслуживаемых процессором сервера.

5. Алгоритм оптимального распределения. Систематизируя результаты предыдущих разделов, можно предложить следующий алгоритм отыскания оптимального распределения пропускных способностей каналов ЛВС и относительных приоритетов заявок. В алгоритме существенно используется часть результатов изложенных нами в работе [2].

Начало

1. Ввод исходных данных:

M – количество групп станций;

N – количество типов сообщений (пакетов);

D – сумма, выделенная на аренду каналов ЛВС;

B – быстродействие процессора сервера;

$(\lambda_{ik})_N^N$ – матрица интенсивностей потоков сообщений;

$(l_1, \dots, l_N)$ – вектор средних длин сообщений;

$(l_1^{(2)}, \dots, l_N^{(2)})$ – вектор вторых начальных моментов длин сообщений;

$(\theta_1, \dots, \theta_N)$ – вектор средних трудоемкостей обработки сообщений на сервере;

$(\theta_1^{(2)}, \dots, \theta_N^{(2)})$ - вектор вторых начальных моментов трудоемкостей сообщений;

$(d_1, \dots, d_M)$ - вектор удельных стоимостей аренды каналов ЛВС;

$(\alpha_1, \dots, \alpha_N)$ - вектор штрафов за ожидание в очередях ЛВС.

2. Назначение относительных приоритетов сообщениям в соответствии с их типом, то есть формирование матриц X, Y : $x_{ii} = y_{ii} = 1$ ($i = 1, \dots, N$), $x_{ij} = y_{ij} = 0$ для $i \neq j$.

3. Вычисление значения минимально необходимых средств для аренды каналов:

$$D_0 = \sum_{i=1}^M d_i \sum_{k=1}^N \lambda_{ik} \cdot l_k.$$

4. Проверка условия $D > D_0$. Если оно не выполняется, то следует выдача сообщения «РЕШЕНИЯ НЕТ», и переход на конец.

5. Численное решение системы уравнений относительно C_i

$$\left. \begin{aligned} & \left[\sum_{k=1}^N \alpha_k \cdot \lambda_{ik} \cdot r_i \cdot \frac{a_{ik} + a_{ik-1} + 2 \cdot C_i}{(C_i - a_{ik})^2 \cdot (C_i - a_{ik-1})^2} \right] + \gamma \cdot d_i = 0, \\ & i = 1, \dots, M; \\ & \sum_{i=1}^M d_i \cdot C_i - D = 0, \end{aligned} \right\}$$

где γ - неопределенный множитель Лагранжа;

$$r_i = \sum_{k=1}^N \lambda_{ik} \cdot l_k^{(2)}, \quad i = 1, \dots, M;$$

$$a_{ik} = \sum_{j=1}^k \lambda_{ij} \cdot l_j, \quad i = 1, \dots, M; \quad k = 1, \dots, N,$$

с начальной точкой

$$C_i^0 = \sum_{k=1}^N (\lambda_{ik} \cdot l_k) + \frac{D - D_0}{\sum_{i=1}^M d_i}, \quad i = 1, \dots, M.$$

6. Вычисление отношений α_j / θ_j , $j = 1, \dots, N$ и сортировка их в порядке убывания.

7. Формирование оптимальной матрицы X по результатам сортировки вектора отношений α_j / θ_j : если заявке s -го типа присваивается k -й приоритет при обслуживании в процессоре сервера, то $x_{sk} = 1$, иначе $x_{sk} = 0$.

8. Вычисление отношений α_j / l_j , $j = 1, \dots, N$ и сортировка их в порядке убывания.

9. Формирование оптимальной матрицы Y по результатам сортировки вектора отношений α_j / l_j : если заявке s -го типа присваивается k -й приоритет передачи по каналу ЛВС, то $y_{sk} = 1$, иначе $y_{sk} = 0$.

10. Вычисление минимальных суммарных потерь $A(X, Y, C)$ в ЛВС по формуле (2) с учетом (11), (12).

11. Вывод A , а также оптимальных матриц X, Y и вектора C .

Конец

Для реализации предложенного алгоритма создана в среде *Delphi5* паскаль-программа, с помощью которой была решена серия задач с типичными для подобных ЛВС исходными данными [1]. Результаты решения показывают существенное уменьшение потерь при введении и/или перераспределении относительных приоритетов.

Дальнейшее уменьшение потерь возможно при введении абсолютных приоритетов для наиболее критичных потоков сообщений. Исследования в этом направлении планируются на ближайшую перспективу.

Библиографический список

1. Столлингс В. Компьютерные системы передачи данных / В. Столлингс. — М.: Изд-во "Вильямс", 2002. — 928 с.
2. Кирюхин В.В. Оптимальное распределение пропускных способностей каналов в централизованной ЛВС с приоритетами // Вестн. СевГТУ: Сб. науч. тр. — Севастополь, 2003. — Вып. 47. — С. 31–37.
3. Бронштейн О.И. Модели приоритетного обслуживания в информационно-вычислительных системах / О.И. Бронштейн, И.М. Духовный. — М.: Наука, 1976. — 220 с.

Поступила в редакцию 03.02.2005 г.