

УДК 004.65(045)

Холявкіна Т.В., к.т.н. (Національний авіаційний університет)

## АДАПТАЦІЯ ПРОЦЕСОВ ОРГАНІЗАЦІИ ЗАПРОСОВ К БАЗЕ ДАННИХ

**Холявкіна Т.В. Адаптація процесів організації запитів до бази даних.** Описана розподілена інформаційно-обчислювальна система аналізу і обробки інформації про безпеку польотів. Задача оптимального розподілу запитів зведена до задачі квадратичного програмування з обмеженнями типу нерівностей.

**Ключові слова:** розподілена інформаційно-обчислювальна система, безпека польоту, оптимізація запитів, теорема Куна-Таккера, квадратичне програмування

**Холявкіна Т.В. Адаптация процессов организации запросов к базе данных.** Описана распределенная информационно-вычислительная система анализа и обработки информации о безопасности полетов. Задача оптимального распределения запросов сведена к задаче квадратичного программирования с ограничениями типа неравенств.

**Ключевые слова:** распределенная информационно-вычислительная система, безопасность полета, оптимизация запросов, теорема Куна-Таккера, квадратичное программирование.

**Holiavkina T.V. Adaptation of processes of organization of requests to database.** Here is considered a distributed data-processing analyzing system and processing information concerning flight safety. The task of optimal request distribution is reduced to task of quadratic programming with restrictions on types of inequality.

**Keywords:** distributed data-processing analyzing system, safety of flight, optimization of requests, Kuhn-Tucker theorem, quadratic programming

**Введение и постановка задачи.** Основываясь на “Положении о системе управления безопасностью полетов на авиационном транспорте”, авиакомпания создает базу данных с целью эффективного анализа полученной информации, в том числе по результатам расследования авиационных событий и добровольных извещений об опасных факторах.

Характерная особенность этой базы данных заключается в следующем. База данных должна иметь распределенный характер.

Благодаря этому, *во-первых*, обеспечивается более быстрый доступ к информации частного характера, необходимой для оперативного функционирования региональных органов авиакомпании, отдельных предприятий обработки полетной информации и др.

*Во-вторых*, повышается надежность хранения данных на терминальных узлах и промежуточных серверах сети.

По существу, рассматриваемая база данных представляет собой распределенную базу данных и знаний (РБД и З), поскольку, помимо задач хранения, решаются также задачи обработки данных [1]. Кроме того, осуществляется архивное копирование и страховочная перезапись данных с периодом, который выбирается по соображениям надежности хранения.

РБД и З является структурой критичного использования по соображениям безопасности и защиты информации. Утрата или другие нарушения целостности информации могут повлечь серьезные и непредсказуемые последствия, вплоть до нарушения работы всей авиационной транспортной инфраструктуры.

Распределенная БД безопасности полетов, по существу, представляет собой информационно-вычислительную систему (ИВС), состоящую из  $N$  устройств обработки и хранения,  $M$  клиентских приложений и иерархической системы управления базой данных. В

общем случае  $M \neq N$ . В каждом устройстве обработки и хранения (сервере) выделена область памяти объемом  $C_i$  ( $i=1, n$ ) только для хранения информации БД безопасности полетов (БП).

Кроме того, в сети могут иметься  $K$  серверов, играющих роль устройств промежуточного хранения данных системы управления безопасностью полетов. Данные из этих серверов используются для решения задач прогнозирования, локализации и обработки возникающих нештатных ситуаций.

Пусть вся информация БД БП представляет собой множество блоков

$$\{a_j\}, j=1, \dots, m, \quad (1)$$

где  $a_j$  объем каждого блока информации, причем не обязательно, чтобы эти объемы были одинаковы для каждого блока.

Каждый сервер может обратиться к любому другому серверу за любым числом блоков из множества  $\{a_j\}$  и получить необходимые данные. Время, которое затрачивается на получение этих данных, зависит от объема требуемой информации, состояния канала передачи (например, туннеля виртуальной частной сети), наличия и длины очереди, числа транзитных узлов коммуникации.

Кроме того, каждый сервер в пределах своего объема памяти  $C_i$  может иметь некоторое количество информации, для обращения к которой дополнительного времени не требуется.

Заявки на обслуживание, поступающие из системы управления БД (СУБД), содержат перечень служебной информации, некоторую необходимо использовать при обслуживании каждой задачи-заявки [2]. Каждая  $j$ -я заявка характеризуется множеством  $I$ , информационных блоков БД, используемых для обслуживания этой заявки ИВС.

Дисциплину обслуживания в штатной ситуации естественно связать со средним временем решения задачи на выделенном для этого сервере, в памяти которого имеется часть необходимой информации. Остальные части необходимой информации сервер получает от других источников.

В нештатной ситуации дисциплина обслуживания полностью определяется приоритетом вновь возникшей задачей. Очевидно, большинство (или даже все) задачи штатных ситуаций отодвигаются в очереди, пока не будет решена вновь возникшая задача [3].

Таким образом, цель адаптации (оптимизации, настройки) системы состоит в том, чтобы распределить ресурсы ИВС в соответствии со следующими правилами:

– дисциплиной обслуживания, т.е. правилом направления очередной заявки на один из серверов, руководствуясь при этом данными о параметрах требуемых блоков информации, состоянии памяти всех серверов сети и их загрузке;

– правилом распределения данных по серверам сети.

Зададим кодовую матрицу  $U = \|u_{ij}\|$ , элементами которой являются двоичные переменные, определяющие наличие ( $u_{ij} = 1$ ) или отсутствие ( $u_{ij} = 0$ ) в памяти  $i$ -го сервера  $j$ -го блока данных. Очевидно, что имеет место следующее ограничение, связанное с конечными размерами области памяти, выделяемой на каждом сервере для хранения данных БД БП:

$$\sum_{j=1}^m a_j u_{ij} \leq C_i, \quad i=1, \dots, m. \quad (2)$$

Пусть поток заявок на запросы данных образуется из  $K$  различных потоков, каждый из которых характеризуется своим дискретным распределением вероятностей использования блоков информации  $\{a_j\}$ :

$$P_l = (P_{l1}, \dots, P_{lm}), l = 1, \dots, k, \quad (3)$$

где  $P_{jl}$  – вероятность того, что при решении задачи  $l$ -го потока потребуется  $j$ -й блок информации.

Без потери общности можно ввести упрощающие правила нормировки:

$$\sum_{j=1}^m P_{jl} = 1, \quad (4)$$

хотя это и не обязательно, так как при запросе по той или иной задаче может потребоваться несколько блоков данных.

Определим вероятность попадания задачи  $l$ -го потока на  $i$ -й сервер при следующей простой и вполне логичной дисциплине обслуживания: задача направляется на решение на тот сервер, где часть общего объема необходимых данных больше всего. Выражение имеет следующий вид:

$$P_{li}(U) = \frac{\sum_{j=1}^m U_{ij} P_{lj}}{\sum_{i=1}^n \sum_{j=1}^m U_{ij} P_{lj}}. \quad (5)$$

Аналогично определяем вероятности того, что для решения задачи  $l$ -го потока, направленной на  $i$ -й сервер, найдется вся необходимая информация, и обращаться ко всей базе данных не придется:

$$\bar{P}_{li} = P_{li}(U).$$

Примем средние интенсивности потоков запросов для решаемых задач  $\lambda_1, \dots, \lambda_k$  и средние интенсивности обслуживания (доставки заявок и решения задачи)  $\mu_1, \dots, \mu_n$ .

Тогда простейшая задача оптимального размещения блоков данных (1) на серверах, при котором минимизируется общая интенсивность запросов к БД:

$$Q(U) = \sum_{i=1}^n \sum_{l=1}^k \lambda_l [1 - \bar{P}_{li}(U)] P_{li}(U) \rightarrow \min, U \in S, \quad (6)$$

где

$$S: \begin{cases} \sum_{l=1}^k \lambda_l P_{lj}(V) < \mu_i \\ \sum_{i=1}^m a_i u_{ij} \leq C_i, \quad i = 1, \dots, n. \end{cases} \quad (7)$$

$$V^T = \| V_{11}, \dots, V_{n1}, \dots, V_{nm} \|. \quad (8)$$

Таким образом, функционал  $Q(U)$  характеризует среднюю интенсивность запросов всех серверов сети.

Ограничения (7) связаны с пропускной способностью системы: при нарушении хотя бы одного из них очередь неограниченно растет. Необходимо также учитывать ограничения области памяти на каждом сервере.

Полученная задача является задачей стохастического программирования с булевыми переменными большой размерности  $n \times m$ . Для условий усреднения  $\lambda_i$  и  $\mu_i$  на интервале наблюдения можно, используя результаты теоремы Куна-Таккера [4...6], усреднить стохастический квазиградиент вида (6) и, таким образом, свести задачи к задаче квадратичного программирования. Таким образом, задача сводится к минимизации квадратичной целевой функции с нелинейными ограничениями вида (7).

Условия Куна-Таккера сформированы для общей задачи нелинейного программирования с ограничениями, как в виде равенств, так и в виде неравенств. Рассмотрим эти условия для нашей задачи с ограничениями только в виде неравенств: минимизировать функцию  $Q(U)$  при ограничениях вида (7).

Запишем условия Куна-Таккера

$$\tilde{\nabla} Q(U) - \sum_i \xi_i \tilde{\nabla} g_i(U) = 0, \quad (9)$$

где  $\tilde{\nabla} Q(U) = E \left[ \frac{Q(U) - Q(U_{n-1})}{\Delta_s} \right]$  – стохастический квазиградиент;

$E$  – символ математического ожидания;

$\Delta_s$  – величина шага квазиградиента между последовательными значениями  $U$  на  $(n-1)$ -м и  $n$ -м шагах;

$\xi_i \geq 0$  – множитель, смысл которого определим ниже;

$\tilde{\nabla} g_i(U)$  – стохастический квазиградиент функции  $g_i(U) = \mu_i - \sum_{l=1}^k \lambda_l P_{li}(U)$ :

$$\tilde{\nabla} g_i(U) = E \frac{g_i(U_n) - g_i(U_{n-1})}{\Delta_s}.$$

При таком выборе функции  $g_i(U)$  ограничения (7) принимают вид:

$$g_i(U) = \mu_i - \sum_{l=1}^k \lambda_l P_{li}(U) \geq 0. \quad (10)$$

Правомерность такого преобразования вытекает из того, что все переменные  $\lambda_i$ ,  $\lambda_l$ ,  $P_{li}$  – неотрицательные величины.

Условия оптимальности записываются как

$$\tilde{\nabla} Q(U) - \sum_j \xi_j g_j(U) = 0, \quad (11)$$

$$g_j(U) = 0, \quad j = 1, \dots, J. \quad (12)$$

Множитель  $\xi_j$  в данной постановке является неопределенным множителем Лагранжа, соответствующим  $j$ -му ограничению. Он представляет собой значение неявной функции стоимости, отражающей изменение минимального значения целевой функции. Подбор значения  $\xi_j$  осуществляется таким образом, чтобы координата точки безусловного минимума удовлетворяла условию  $g_j(U) = 0$ .

Это легко сделать, если, рассматривая  $\xi_j$  как независимую переменную, найти безусловный минимум функции  $Q(U)$  без учета влияния второго слагаемого в выражении (11), а затем выбрать значение  $\xi_j$ , при котором выполняется равенство в этом выражении целиком. Другими словами, мы погружаем исходную задачу минимизации (6) в задачу минимизации большей размерности (11).

### **Выводы**

1. Задача оптимального распределения запросов сведена к задаче квадратичного программирования с ограничениями типа неравенств.
2. Для определения условий необходимости и достаточности решения модифицированы условия теоремы Куна-Таккера с учетом дискретности и случайного характера квазиградиента.

### **Литература**

1. Ребекка М. Райордан. Основы реляционных баз данных / Ребекка М. Райордан. – М.: Издатель: Русская редакция, 2001. – 384 с.
2. Холявкина Т.В. Организация запросов в распределенной системе управления безопасностью полетов / Т.В. Холявкина // Проблеми інформатизації та управління : зб. наук. пр. – К.: НАУ, 2009. – Вип.4(28). – С. 140-143.
3. Гнеденко Б.В. Введение в теорию массового обслуживания / Б.В. Гнеденко, И.Н. Коваленко. – [2-е изд.]. – М.: Наука, 1987. – 336 с.
4. Аоки М. Введение в методы оптимизации / М. Аоки. – М.: Наука, 1977. – 344 с.
5. Реклейник Г. Оптимизация в технике: в 2-х кн., кн.2 / Г. Реклейник, А. Рейвинзрок, К. Рэгсдел; пер. с англ. – М.: Мир, 1986. – 320 с.
6. Башарин Г.П. Анализ очередей в вычислительных сетях. Теория и методы расчета / Г.П. Башарин, П.П. Бочаров, Я.А. Коган. – М.: Наука, 1989. – 335 с.