

# РАЗМЕЩЕНИЕ РЕПЛИК МАССИВОВ ДАННЫХ В НЕНАДЕЖНЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМАХ, ОПТИМАЛЬНОЕ ПО КРИТЕРИЮ МИНИМУМА ВРЕМЕНИ ДОСТУПА К ДАННЫМ

Микрин Е.А.<sup>1</sup>, Сомов С.К.<sup>2</sup>

<sup>1</sup> ПАО «Ракетно-космическая корпорация «Энергия» им. С.П. Королева»,  
г. Королев, ул. Ленина, 4а  
Eugeny.Mikrin@rsce.ru,

<sup>2</sup> Институт проблем управления им. В.А. Трапезникова РАН,  
Россия, г. Москва, ул. Профсоюзная д.65, стр.1.  
ssomov2016@ipu.ru

*Аннотация:* Рассматривается проблема оптимального размещения реплик взаимосвязанных массивов данных в распределенной системе. Каналы связи в системе ненадежны. В качестве критерия оптимизации применяется минимум времени ответа системы на запросы к данным. Предлагается эвристический алгоритм решения задачи оптимизации.

Ключевые слова: распределенная система обработки данных, ненадежные каналы связи, оптимальное распределение реплик массивов данных.

## Введение

При проектировании распределенных систем обработки данных (РСОД) необходимо выполнить различные требования к их характеристикам. Наиболее важные требования следующие: 1) приемлемое время обработки запросов к данным системы, 2) высокий уровень сохранности информации, 3) требуемый уровень надежности работы системы. Использование в РСОД репликации массивов данных системы позволяет выполнить эти требования. Задачи поиска оптимального распределения реплик по узлам сети обладают большой вычислительной сложностью (NP). Поэтому для их решения используются различные методы, которые уменьшают их вычислительную сложность [1-5]. В докладе рассматривается задача оптимального размещения реплик взаимосвязанных массивов данных по узлам РСОД с ненадежными каналами связи. В качестве критерия оптимизации используется минимум среднего времени обработки запросов.

## 1 Формальная модель распределенной системы обработки данных

В узлах РСОД выполняются несколько прикладных процессов (задач) разных типов. Частота решения задач задана. Каждая задача генерирует некоторый трафик информационных запросов и запросов на модификацию данных в репликах. Информационный запрос маршрутизируется в ближайший узел с необходимой для его обработки репликой. Запрос на модификацию данных массива данных маршрутизируется во все узлы с репликами этого массива.

Компьютерная сеть, на базе которой работает РСОД, состоит из  $N$  узлов. Топология сети задана взвешенным графом  $G(N, E, W)$ , для дуг которого определены их длины. РСОД использует  $M$  массивов данных. Сеть передачи состоит из  $\Phi$  ненадежных каналов связи с заданными для них пропускными способностями  $\varphi$ . При передаче единицы данных по каналу связи  $\varphi$ -м ошибка может возникнуть с вероятностью  $P(\varphi)$ . Успешная доставка сообщения подтверждается квитанцией АСК (ACKnowledgement). Пусть  $T$  – это время ожидания узлом-отправителем квитанции АСК от узла-адресата. Если за время  $T$  узел-отправитель не получил квитанцию АСК, то сообщение передается повторно. Полагаем, что размер квитанции АСК незначителен, она передается за пренебрежимо малое время, и вероятность ошибки при ее передаче равна нулю.

В системе решается несколько задач, принадлежащих одному из  $J$  типов. Частота решения задач определяется матрицей  $F^* = \|f_{nj}^*\|$  ( $n = \overline{1, N}; j = \overline{1, J}$ ), где  $f_{nj}^*$  - частота решения задачи  $j$ -го типа в

n-м узле. Матрица  $E^* = \|e^*_{jm}\|$ , задает частоту  $e^*_{jm}$  генерации j-й задачей j-го информационных запросов к m-му массиву данных. Матрица  $U^* = \|u^*_{jm}\|$  определяет частоты  $u^*_{jm}$  запросов на модификацию данных. Полагаем, что время  $T_{pr}$  обработки любого запроса в любом узле одинаково.

Между всеми парами вершин графа  $G$  определяются кратчайшие пути посредством одного из известных алгоритмов (например, алгоритмы Флойда — Уоршелла, Беллмана — Форда, [6]). В итоге для графа  $G$  определяется матрица  $SP = (sp_{nk})_{N \times N}$  кратчайших путей, где  $sp_{nk}$  — это длина кратчайшего пути/маршрута между узлами n и k.

Элементы матрицы  $A = \|a_{nm}\|$ , ( $a_{nm} \in \{0,1\}$ ) определяют распределение реплик по узлам системы. Элемент  $a_{nm}$  равен 1, если в n-м узле размещена реплика m-го массива. Каждый массив данных может иметь несколько реплик, размещенных в разных узлах.

Матрица  $B = \|b_{nkm}\|$ , размерности  $N \times N \times M$ , задает «близость» узлов сети. Элемент  $b_{nkm} \in \{0,1\}$ , и  $b_{nkm} = 1$ , если в узле k размещена реплика массива данных m, а узел k является ближайшим узлом для узла n.

Заданы значения: средней длины  $\bar{l}_m^q$  информационного запроса к массиву m-го типа, и  $\bar{l}_m^r$  — средняя длина ответа на этот запрос,  $\bar{l}_m^q$  — средней длины  $\bar{l}_m^q$  запроса на модификацию m-го массива, и  $\bar{l}_m^r$  — средней длины ответа на этот запрос.

## 2 Формулировка оптимизационной задачи размещения реплик в узлах системы.

Необходимо найти распределение  $A$  реплик  $M$  массивов данных в узлах распределенной системы такое, чтобы обеспечивалось минимальное значение функционала  $F(A)$  задачи:  $F(A) = \tilde{T} + T_{pr} + \tilde{T}$

здесь:  $\tilde{T}$  — среднее время передачи запроса пользователя в узел с репликой массива данных,  $T_{pr}$  — среднее время обработки запроса,  $\tilde{T}$  — среднее время передачи ответа на запрос.

Таким образом, необходимо найти минимум функционала  $F(A)$ :  $\min F(A) = \min(\tilde{T} + T_{pr} + \tilde{T})$

В задаче используются ограничения:

ограничение на затраты  $OP_{cost}$  на функционирование системы:  $OP_{cost} \leq COST_{MAX}$

ограничение  $RN_{MAX_n}$  на максимальное количество реплик в узлах системы.

$$\sum_{m=1}^M a_{nm} \leq RN_{MAX_n}, \quad n = \overline{1, N}$$

## 3 Величина средней задержки сообщений в распределенной системе

Значение средней задержки сообщений в системе определяется с учетом ненадежности каналов связи. В каждом из двух произвольных узлов сети  $i$  и  $k$  могут решаться задачи пользователей и храниться реплики. Информационные запросы адресуются из узла  $i$  в узел  $k$ , если узел  $k$  это ближайший и он имеет реплику нужного массива.

Во всей сети в единицу времени генерируется трафик  $Y$  сообщений, величина которого равна:

$$Y = \sum_{i=1}^N \sum_{k=1}^N (y_{ik}^q + y_{ik}^r)$$

Средняя длина  $1/\mu$  сообщений, передаваемых по каналам сети, равна:

$$1/\mu = Y^{-1} \left\{ \sum_{i=1}^N \sum_{k=1}^N \sum_{j=1}^N f_{ij}^* \left[ \sum_{m=1}^M a_{km} * d^1 + \sum_{m=1}^M a_{im} * d^2 \right] \right\},$$

$$d^1 = (\bar{l}_m^q * e^*_{jm} * sp_{ik} + \bar{l}_m^q * u^*_{jm}), \quad d^2 = \bar{l}_m^q * e^*_{jm} * sp_{ki} + \bar{l}_m^r * u^*_{jm}$$

В распределенных системах с ненадежными каналами связи с вероятностью  $q^* = 1 - \rho^*$  при передаче единицы данных в  $\varphi$ -м канале может возникнуть ошибка. Вероятность  $q_\varphi = 1 - \rho_\varphi$  возникновения ошибки в  $\varphi$ -м канале при передаче сообщения со средней длиной  $l = \mu^{-1}$  равна:

$$q_\varphi = 1 - \rho_\varphi = 1 - (1 - q^*)^l = 1 - (\rho^*)^l$$

Фактическая пропускная способность  $C'_\varphi$  ненадежного канала связи равна:

$$C'_\varphi = \mu^{-1}/t'_\varphi, \quad \text{где } t'_\varphi = \rho_\varphi T_\varphi^* + (t_{ACK} + \rho_\varphi T_\varphi^*) q_\varphi (1 - q_\varphi)^{-1}, \quad \text{где } q_\varphi = 1 - (\rho_\varphi^*)^l$$

Среднее время  $T'_\varphi$  задержки сообщения в  $\varphi$ -м ненадежном канале связи будет равно:

$$T'_\varphi = 1 / (\mu C'_\varphi - \lambda_\varphi^*)$$

Тогда, в системе с ненадежными каналами величина  $\tilde{T}$  среднего времени задержки сообщения при передаче по каналам сети равна:

$$\tilde{T} = \sum_i \sum_k \tilde{Z}_{ik}$$

Здесь  $\tilde{Z}_{ik}$  – значение средней задержки сообщения, передаваемого по ненадежным каналам связи сети из узла  $i$  в узел  $k$ .

#### 4 Алгоритм решения задачи

Для решения сформулированной задачи, имеющей большую вычислительную сложность, предложен эвристический алгоритм, подробно описанный в докладе. Алгоритм реализован на языке программирования C++ (MS Visual Studio). Алгоритм используется в автоматизированном комплексе поддержки принятия решений по обеспечению сохранности данных в распределенных системах. Использование алгоритма позволяет улучшать время реакции системы на запросы на 5-8% и сокращать издержки на эксплуатацию системы на 8-10%.

#### Литература

1. Чернышев Г.А. Обзор подходов к организации физического уровня в СУБД // Труды СПИИРАН. 2013. - Санкт-Петербург, 2013. Вып. 1(24). – С. 222 – 275.
2. Azzam Sleit and oth. A Dynamic Object Fragmentation and Replication Algorithm In Distributed Database Systems // American Journal of Applied Sciences. 2007. Vol. 4 (8). P.613-618.
3. Loukopoulos T., Ahmad I., Papadias D. An Overview of Data Replication on the Internet // Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN.02). 2002. 6 p.
4. Sahoo J., Salahuddin M.A., Glitho R. A Survey on Replica Server Placement Algorithms for Content Delivery Networks // IEEE Communications Surveys & Tutorials. 2016. – P. 30. URL: <https://arxiv.org/ftp/arxiv/papers/1611/1611.01729.pdf> (дата обращения 2018-07-06).
5. Mansouri N. Adaptive data replication strategy in cloud computing for performance improvement // Frontiers of Computer Science. 2016. — Vol. 10, N 5. — P. 925-935.
6. Cormen T.H., et al. Introduction to Algorithms, Third Edition // The MIT Press, 2009. — 1313 p.