

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2019, №6, Том 11 / 2019, No 6, Vol 11 <https://esj.today/issue-6-2019.html>

URL статьи: <https://esj.today/PDF/59ITVN619.pdf>

Ссылка для цитирования этой статьи:

Сомов С.К. Пример задачи размещения реплик массивов данных в ненадежных распределенных системах и алгоритм ее решения // Вестник Евразийской науки, 2019 №6, <https://esj.today/PDF/59ITVN619.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:

Somov S.K. (2019). An example of the task of allocation the replicas of data in unreliable distributed systems and an algorithm of its solution. *The Eurasian Scientific Journal*, [online] 6(11). Available at: <https://esj.today/PDF/59ITVN619.pdf> (in Russian)

Работа выполнена при финансовой поддержке:

- РФФИ в рамках научного проекта № 1607-00245 «Модели и методы управления региональной безопасностью на основе сценарного подхода»

- программы Президиума РАН №30 «Теория и технологии многоуровневого децентрализованного группового управления в условиях конфликта и кооперации»

УДК 519.1; 519.2

ГРНТИ 20.23.17; 50.39.15

Сомов Сергей Константинович

ФГБУН «Институт проблем управления им. В.А. Трапезникова Российской академии наук», Москва, Россия
Старший научный сотрудник
Кандидат технических наук
E-mail: ssomov2009@gmail.com

Пример задачи размещения реплик массивов данных в ненадежных распределенных системах и алгоритм ее решения

Аннотация. В работе рассматривается один из методов повышения эффективности работы распределенных систем обработки данных – оптимальное размещение в узлах системы реплик массивов данных, используемых в системе. Данный метод позволяет обеспечить высокий уровень сохранности информации и высокую надежность работы системы, а также минимизировать затраты на ее эксплуатацию. Данный метод широко используется на практике при создании распределенных информационных систем различного назначения и масштаба.

Выполнена формулировка задачи оптимального распределения реплик нескольких взаимосвязанных массивов данных в узлах распределенной системы с ненадежными каналами связи, а также приводится эвристический алгоритм решения данной задачи. В качестве критерия оптимизации в задаче используется минимум затрат на эксплуатацию системы.

В начале работы приведены статистические данные о размере убытков, которые несут компании в случае простоя информационных систем, вызванных негативными факторами разной природы. Дан краткий обзор работ по тематике размещения данных в распределенных системах.

Далее в работе представлена формальная модель распределенной системы с ненадежными каналами, в узлах которой выполняется множество прикладных процессов, генерирующих информационные запросы к репликам массивов данных и запросы на изменение данных в репликах. Приводится формулировка задачи оптимального размещения реплик в узлах распределенной системы.

Приведен перечень шагов и формулы для определения величины средней задержки сообщений в каналах связи системы с учетом ненадежности каналов связи, а так же формулы для вычисления затрат на эксплуатацию распределенной системы с размещенными в ее узлах репликами массивов данных.

Так как сформулированная задача обладает большой вычислительной сложностью, в работе представлен эвристический алгоритм для поиска оптимального размещения реплик множества массивов данных, используемых в системе.

В заключении работы перечислено несколько перспективных направлений для продолжения исследований в области использования информационной избыточности в виде реплик массивов данных для повышения эффективности работы распределенных систем.

Ключевые слова: распределенные системы; компьютерные сети; репликация данных; размещение реплик; эвристический алгоритм; минимум затрат на эксплуатацию системы; сохранность информации; надежность работы

Введение

К распределенным системам обработки данных (РСОД) различного масштаба и назначения предъявляются различные требования. Наиболее важными из этих требований являются следующие: высокий уровень сохранности информации, высокая надежность работы распределенных систем, время ответов на запросы, не превышающее заданного значения, и минимальные затраты на эксплуатацию системы. Эффективным методом реализации этих требований является репликация данных, используемых в системе, и оптимальное размещение по узлам распределенной системы реплик массивов данных [1–3].

Актуальность использования методов репликации данных в распределенных системах определяется тем, что данные методы не только повышают быстродействие таких систем, но и обеспечивают высокий уровень их надежности.

Методы репликации широко используются, например, в распределенных базах данных (РБД), когда необходимо оптимальным образом разместить таблицы данных РБД, либо горизонтальные или вертикальные фрагменты таких таблиц [4]. В GRID системах репликация данных используется для повышения доступности, отказоустойчивости, балансировки нагрузки и масштабируемости системы при одновременном снижении нагрузки на каналы связи и сокращения времени выполнения заданий [5; 6].

Недостаточная надежность распределенных систем может привести к тому, что под воздействием негативных факторов разной природы система может дать сбой, что в свою очередь вызовет простой в ее работе той или иной длительности. Причинами, которые могут вызывать сбои в работе распределенных систем могут, например, быть отказы аппаратных средств, перебои в электропитании, некорректная работа программного обеспечения (ошибки в программном коде), некорректные действия персонала или пользователей системы, проникновение в систему компьютерных вирусов [1].

Простои в работе распределенных систем вызывают прямые и косвенные убытки компаний и организаций, использующих в своей работе сервисы данных систем. В частности, в ходе исследования, которое в 2016 году было выполнено компанией DELL, было опрошено 2200 специалистов из 18-ти стран, принимающих решения в области ИТ¹. Часть из заданных

¹ Dell EMC Global Data Protection Index II // [Электронный ресурс] – URL: <http://www.emc.com/microsites/emc-global-data-protection-index/index.htm> (дата обращения 20-03-2019).

специалистам вопросов касалась причин негативных инцидентов в информационных системах и размеров убытков компаний за час простоя, вызванного инцидентами с информационными системами. В результате опроса специалистов выяснилось, что наиболее частыми причинами внеплановых простоев ИТ-систем являются следующие: сбои аппаратуры (45 % от общего числа сбоев), сбои в электроснабжении (35 %) и программные ошибки (34 %). Величина убытков за час внепланового простоя у большинства компаний (58 %) находится в диапазоне от 50 000 до 500 000 долларов США. У пятнадцати процентов компаний размер убытков превышает 500 000 долларов (11 %) или даже имеет величину более миллиона долларов (4 %).

В работе приведена формулировка задачи оптимального распределения реплик нескольких взаимосвязанных массивов данных в узлах распределенной системы с ненадежными каналами связи, а также эвристический алгоритм решения данной задачи. В качестве критерия оптимизации в задаче используется минимум затрат на эксплуатацию системы.

1. Обзор работ по тематике

Задача оптимального размещения данных по узлам сети имеет давнюю историю. Проблема размещения реплик в некотором смысле является наследницей задачи размещения файлов на узлах распределенной компьютерной сети. Первой работой, в которой была сформулирована такая задача, была работа [7]. В этой работе автором (W.W. Chu) была представлена формальная модель, позволяющая минимизировать общие эксплуатационные расходы системы (многомашинного комплекса) при ограничении времени отклика на запрос и емкости памяти для хранения файлов при фиксированном количестве копий каждого файла.

В работе [8] автор модифицировал модель распределения файлов, предложенную в работе [7] и смягчил требование о фиксированном количестве копий. В этой же работе автором подчеркнута разница в обработке транзакций чтения и обновления данных. В [9] было доказано, что сформулированная в работе [7] задача является NP-полной, поэтому поиск ее оптимального решения невозможен из-за большой вычислительной сложности. В работе [7] авторами была рассмотрена проблема размещения файлов для типичных приложений распределенных баз данных с простой моделью выполнения транзакций.

В работе [10] доказано, что проблема размещения фрагментов в распределенной базе данных существенно отличается от проблемы размещения файлов. В этой же работе автором был предложен метод распределения данных, предназначенный для минимизации общих затрат на обмен данными во время выполнения множества транзакций.

Примерно в это же время стали появляться работы, посвященные проблемам фрагментации баз данных, как, например, работа [11].

Выше даны ссылки на работы, которые были опубликованы довольно давно. Современные работы по данной тематике широко представлены на сайтах различных издательств и на других Интернет-ресурсах. В частности, обширный перечень работ по проблеме фрагментации и распределении фрагментов таблиц распределенных баз данных приведен в работе [4].

Цель данной статьи – представить модель распределенной системы, использующей каналы связи с ненулевой вероятностью сбоя при передаче сообщений, сформулировать задачу оптимального размещения реплик в узлах такой системы и предложить алгоритм решения сформулированной задачи.

Основное содержание данной работы состоит из нескольких разделов. В следующем разделе представлена модель распределенной системы с ненадежными каналами связи. Затем,

в разделе 3, дается формулировка задачи поиска оптимального распределения реплик в представленной распределенной системе. В четвертом разделе приведены формулы для определения величины средней задержки сообщений в каналах связи системы с учетом ненадежности каналов связи. В пятом разделе приведена последовательность шагов для вычисления затрат на эксплуатацию распределенной системы, в узлах которой размещены реплики, а каналы связи ненадежны. В шестом разделе представлен эвристический алгоритм решения сформулированной в статье задачи.

2. Модель распределенной системы с ненадежными каналами связи

Предположим, что имеется распределенная система обработки данных, которая функционирует на базе компьютерной сети из N узлов. Ее топология представлена взвешенным графом $G = (X, \Gamma)$. Заданы длины дуг графа. Система использует M ($m = \overline{1, M}$) массивов данных разного типа. Сеть передачи данных сети состоит из Φ ($\varphi = \overline{1, \Phi}$) каналов связи с заданными пропускными способностями C_φ . В сети передачи данных используется метод коммутации сообщений. В каналах связи сети передачи данных с некоторой вероятностью могут возникать ошибки/сбои при передаче данных. Обозначим как $q_\varphi^* = 1 - \rho_\varphi^*$, ($\varphi = \overline{1, \Phi}$) вероятность возникновения ошибки в φ -м канале при передаче единицы данных по каналу единичной длины. Успешность доставки сообщений в узлы-адресаты подтверждается передачей в узлы-отправители квитанций АСК (ACKnowledgement). Пусть t_{ACK} – время ожидания узлом-отправителем квитанции АСК о доставке сообщения в узел-получатель. Если по истечении времени ожидания узлом-отправителем квитанция не получена узлом-отправителем, производится повторная передача не доставленного сообщения.

В узлах РСОД (пользовательские сервера) выполняется несколько прикладных процессов (задач) различного типа. В одном узле может одновременно выполняться несколько задач. Частота решения задач в узлах сети задана матрицей $F^* = \|f_{nj}^*\|$ ($n = \overline{1, N}$; $j = \overline{1, J}$), где f_{nj}^* это частота решения в n -м узле задачи j -го типа. Каждая задача в процессе решения может генерировать некоторое количество информационных запросов к нескольким массивам данных и некоторое количество запросов на модификацию массивов данных. Матрица $E^* = \|e_{jm}^*\|$, определяет частоту e_{jm}^* – генерации j -й задачей информационных запросов к массиву m . Матрица $U^* = \|u_{jm}^*\|$ задает частоту u_{jm}^* генерации запросов на модификацию m -го массива. Предположим, что время обработки любого запроса в любом узле системы постоянно и равно T_{pr} . Запрос на модификацию массива данных адресуется во все узлы с репликами этого массива. Информационные запросы адресуются в ближайшие узлы с репликами запрашиваемого массива данных. Кратчайшие пути между всеми парами вершин графа G определяются с помощью одного из известных алгоритмов поиска кратчайших путей на взвешенных графах (например, алгоритмы Беллмана-Форда, Флойда-Уоршелла [12]). Итогом работы такого алгоритма является матрица $SP = (sp_{nk})_{N \times N}$ кратчайших путей в графе, где элемент sp_{nk} равен длине кратчайшего пути между узлами n и k .

Распределение реплик описывается матрицей $A = \|a_{nm}\|$, в которой $a_{nm} \in \{0,1\}$, и $a_{nm} = 1$, если в n -м узле сети размещена реплика m -го массива данных. Каждый массив может иметь несколько реплик, размещенных в разных узлах сети. На основе матриц SP и A вычисляются элементы матрицы «близости» узлов $B = \|b_{nkm}\|$ размерности $N \times N \times M$, в которой элемент $b_{nkm} \in \{0,1\}$, и $b_{nkm} = 1$, если в узле k размещена реплика массива данных m , и узел k является ближайшим узлом для n . Задана средняя длина \bar{l}_m^q информационного запроса к массиву m -го типа, и средняя длина \bar{l}_m^r ответа на запрос. Также определена средняя длина \bar{l}_m^q запроса на модификацию массива данных m и средняя длина \bar{l}_m^r ответа на этот запрос.

3. Формулировка задачи оптимального размещения реплик

Необходимо для РСОД найти такое распределение A реплик M массивов данных по узлам системы, чтобы обеспечить минимальное значение функционала $F(A)$ задачи. Функционал $F(A)$ это величина затрат на эксплуатацию системы.

Таким образом, необходимо найти минимум функционала:

$$\min F(A) = \min OP_{cost}(A) = \min(Cost_{dtr} + Cost_{st} + Cost_{EU}) \quad (1)$$

Здесь:

$OP_{cost}(A)$ – затраты на эксплуатацию системы,

$Cost_{dtr}$ – затраты на обмен данными между узлами сети,

$Cost_{st}$ – стоимость хранения реплик в узлах системы,

$Cost_{EU}$ – стоимость обработки запросов, генерируемых в системе.

В оптимизационной задаче используются два ограничения:

- ограничение RN_{MAX_n} на максимальное количество реплик, размещенных в одном узле системы:

$$\sum_{m=1}^M a_{nm} \leq RN_{MAX_n}, n = \overline{1, N} \quad (2)$$

- ограничение T_{MAX} на среднее время обработки запросов:

$$T = (T_{enq} + T_{pr} + T_{rep}) \leq T_{MAX} \quad (3)$$

Здесь: T_{enq} среднее время передачи запроса от пользовательского сервера в узел системы с репликой необходимого массива данных (сервер реплик), а T_{rep} среднее время передачи ответа на запрос от сервера реплик до пользовательского сервера. T_{pr} – среднее время обработки запроса в узле сети (сервер реплик).

4. Величина средней задержки сообщений в системе

Определим величину среднего времени передачи сообщений (запросов и ответов на них) в распределенной системе с ненадежными каналами связи.

Рассмотрим трафик сообщений между двумя произвольными узлами i и k сети (квитанции АСК не рассматриваем). В соответствии с заданной моделью системы количество генерируемых в узле i информационных запросов и запросов на модификацию данных, адресованных в узел k , равно:

$$y_{ik}^q = \sum_{j=1}^J f_{ij}^* * \sum_{m=1}^M (e_{jm}^* * b_{ikm} + a_{km} * u_{jm}^*)$$

За то же время узел i отправляет в узел k ответы на запросы с частотой:

$$y_{ik}^r = \sum_{j=1}^J f_{kj}^* * \sum_{m=1}^M (e_{jm}^* * b_{kim} + a_{im} * u_{jm}^*)$$

Во всей сети в единицу времени генерируется трафик сообщений Y , равный:

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

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

$$1/\mu = Y^{-1} \left\{ \sum_{i=1}^N \sum_{k=1}^N \sum_{j=1}^N f_{ij}^* \left[\sum_{m=1}^M a_{km} * d^1 + \sum_{m=1}^M a_{im} * d^2 \right] \right\}, \quad (4)$$

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

Определим величину T^* среднего времени передачи сообщений, используя результаты работ [13; 14]. Согласно данным результатам, при передаче сообщения по очередному каналу связи оно имеет длину, распределенную в соответствии с показательным законом. Время передачи сообщений по каналам также независимая случайная величина. Предположим, что обработка сообщений происходит в соответствии с дисциплиной *FCFS* [13; 15]. В этом случае каждый φ -й канал можно рассматривать как систему массового обслуживания (СМО) вида $M|M|1|\infty$. Такая СМО обладает пуассоновским входным потоком со средним значением γ_{ik} сообщений в секунду и показательным законом распределения времени обслуживания со средним значением μ^{-1}/C_φ . Тогда средняя задержка сообщения равна:

$$T^* = \sum_{i=1}^N \sum_{k=1}^N \frac{\gamma_{ik}}{\gamma} Z_{ik} \quad (5)$$

Здесь Z_{ik} – средняя задержка сообщения, передаваемого из узла i в узел k по пути sp_{ik} .

Канал связи φ включен в путь sp_{ik} , если сообщения, передаваемые по этому пути, используют данный канал. Тогда нам задано множество из Φ матриц $X^\varphi = \|x_{ik}^\varphi\|$, ($\varphi = \overline{1, \Phi}$, $i, k = \overline{1, N}$), в которых элемент $x_{ik}^\varphi = 1$, если канал связи φ входит в путь sp_{ik} . Через канал φ проходит поток сообщений с интенсивностью:

$$\lambda_\varphi^* = \sum_{i=1}^N \sum_{k=1}^N \gamma_{ik} x_{ik}^\varphi \quad (6)$$

Обозначим через T_φ^* среднее время ожидания сообщением в очереди на обслуживание и на передачу его по φ -му каналу. Тогда величина Z_{ik} будет равна сумме времен T_φ^* обслуживания сообщения во всех каналах маршрута sp_{ik} :

$$Z_{ik} = \sum_{\varphi=1}^{\Phi} x_{ik}^\varphi * T_\varphi^*$$

С учетом формулы (5) средняя задержка сообщения в сети равна:

$$T^* = \sum_{\varphi=1}^{\Phi} \frac{T_\varphi^*}{Y} \sum_{i=1}^N \sum_{k=1}^N \gamma_{ik} x_{ik}^\varphi$$

Используя формулу (6), получим:

$$T^* = \sum_{\varphi=1}^{\Phi} \frac{\lambda_\varphi^*}{Y} T_\varphi^* \quad (7)$$

Согласно результатам, изложенным в [13] величина T_φ^* средней задержки сообщений в надежном канале φ равна:

$$T_\varphi^* = 1 / (\mu C_\varphi - \lambda_\varphi^*) \quad (8)$$

В распределенных сетях с ненадежными каналами связи в φ -м канале с вероятностью $q_\varphi^* = 1 - \rho_\varphi^*$ может возникнуть ошибка при передаче единицы данных. В этом случае производится повторная передача сообщения.

Вероятность возникновения ошибки $q_\varphi = 1 - \rho_\varphi$ при передаче сообщения со средней длиной $l = \mu^{-1}$ единиц данных по φ -му каналу равна:

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

Следовательно, среднее время t'_φ передачи сообщения длиной $l = \mu^{-1}$ по ненадежному φ -му каналу:

$$t'_\varphi = \rho_\varphi T_\varphi^* + (t_{ACK} + \rho_\varphi T_\varphi^*) q_\varphi (1 - q_\varphi)^{-1}, \text{ где } q_\varphi = 1 - (\rho_\varphi^*)^l \quad (9)$$

Пусть C'_φ это фактическая пропускная способность ненадежного φ -го канала. Значение среднего времени t'_φ передачи сообщения по ненадежному каналу равно отношению его длины к фактической пропускной способности канала, т. е. $t'_\varphi = \mu^{-1} / C'_\varphi$. Тогда фактическая пропускная способность ненадежного канала равна $C'_\varphi = \mu^{-1} / t'_\varphi$, где t'_φ определяется по формуле (9). Среднее время T'_φ задержки сообщения в φ -м ненадежном канале связи будет равно $T'_\varphi = 1 / (\mu C'_\varphi - \lambda_\varphi^*)$.

Тогда, для РСОД с ненадежными каналами связи величина \tilde{T} среднего времени задержки сообщения в каналах сети вычисляется по следующей формуле:

$$\tilde{T} = \sum_{i=1}^N \sum_{k=1}^N \frac{\gamma_{ik}}{\gamma} \tilde{Z}_{ik}, \text{ где } \tilde{Z}_{ik} = \sum_{\varphi=1}^{\Phi} x_{ik}^\varphi * T'_\varphi \quad (10)$$

5. Величина затрат на эксплуатацию распределенной системы

Определим величину затрат на эксплуатацию распределенной системы. Будем использовать следующие обозначения:

- $dst(x_i, x_k)$ – длина кратчайшего пути между узлами x_i и x_k , равная значению элемента sp_{ik} ;
- dtc – стоимость передачи единицы данных по пути единичной длины;
- v^e, v^u – средний объем данных, передаваемых по каналам связи при обработке информационного запроса и запроса на модификацию данных соответственно;
- $S_cst_m(x_k)$ – стоимость хранения одной реплики массива данных в узле x_k ;
- $E_cst_m(x_k), U_cst(x_k)$ – стоимость обработки информационного запроса и запроса на модификацию данных реплики m -го массива, размещенной в узле x_k ;
- $W = \{w_1, \dots, w_k, \dots, w_N\}$ – вектор «весов» узлов графа. Вес w_k узла k равен среднему объему данных, которыми узел k обменивается с другими узлами системы при обработке запросов, сгенерированных в этом узле за единицу времени, т. е.:

$$w_k = \sum_{j=1}^J f_{kj}^* \left(\sum_{m=1}^M e_{jm}^* v^e + \sum_{m=1}^M u_{jm}^* v^u \right)$$

Используя матрицу A и множество X , получим M множеств X_m^r ($m = \overline{1, M}$) с номерами тех узлов из множества X , в которых размещены реплики M массивов данных.

Пусть $d^m(x_n, X_m^r)$ это минимальное расстояние от узла x_n из множества X до одного из узлов множества X_m^r :

$$d^m(x_n, X_m^r) = \min_{x_j \in X_m^r} \text{dst}(x_n, x_j), m = \overline{1, M}$$

Для каждого множества X_m^r определим передаточное число σ , т. е. затраты на обмен данными между узлами из множества X_m^r и узлами из множества X при обработке запросов в узлах множества X_m^r :

$$\sigma(X_m^r) = \sum_{n=1}^I d^m(x_n, X_m^r) * \text{drc} * w_n$$

Суммарная стоимость OP_cost затрат на эксплуатацию РСОД равна:

$$OP_cost = Cost_{dtr} + Cost_{st} + Cost_{EU}$$

Здесь:

$$Cost_{dtr} = \sum_{m=1}^M \sigma(X_m^r); Cost_{st} = \sum_{n=1}^N \sum_{m=1}^M S_{cst_m}(x_k) * a_{nm};$$

$$Cost_{EU} = \sum_{n=1}^N \sum_{j=1}^J f_{nj}^* * \sum_{m=1}^M \left(e_{jm}^* * \sum_{k=1}^N b_{nkm} * E_{cst_m}(x_k) + u_{jm}^* * \sum_{k=1}^N a_{km} * U_{cst}(x_k) \right)$$

6. Алгоритм решения оптимизационной задачи

Сформулированная выше задача (1)–(3) имеет большую вычислительную сложность, и для ее решения предлагается следующий эвристический алгоритм:

Шаг 1. Определяем частоту fr_m всех запросов к m -му массиву данных ($m = \overline{1, M}$), возникающих во всех узлах системы:

$$fr_m = \sum_{n=1}^N f_{nj}^* (v_{jm}^* + u_{jm}^*); m = \overline{1, M}$$

Шаг 2. Выстраиваем номера массивов данных в порядке убывания соответствующих им частот fr_m ($m = \overline{1, M}$).

Шаг 3. Строим вектор $VM = \langle vm_k \rangle, k = \overline{1, M}$. Элемент вектора vm_1 это номер массива, к которому возникает наибольшее количество запросов. Элемент vm_M содержит номер массива данных с наименьшим количеством запросов.

Шаг 4. Определяем значения параметров P_{min} P_{max} количества узлов сети, в которых могут быть размещены реплики.

Шаг 5. В цикле по количеству p узлов с репликами ($p = \overline{P_{min}, P_{max}}$), выполняем шаги с 6 по 11:

Шаг 6. В цикле выбираем номера vm_k ($k = \overline{1, M}$) массивов из элементов вектора VM . Для каждого выбранного массива данных выполняются действия:

Случайным образом из X выбираются p узлов для распределения реплик массива m . Номера этих узлов составляют множество $X_p^m = \{x_{pi}^m | i = \overline{1, p}\}$. Элемент x_{pi}^m этого множества содержит номер узла сети, в котором размещается одна реплика m -го массива. Номера этих узлов запоминаются во множестве X_m « m -протестированных» узлов. Остальные номера из X , не вошедшие в X_p^m , запоминаются во множестве \bar{X}_m «не протестированных» узлов. Т. е. $\bar{X}_m = \{X \setminus X_p^m\}$.

Шаг 7. Присваиваем “1” тем элементам a_{nm}^p m -го столбца матрицы A^p , у которого номер n строки матрицы равен номеру x_{pi}^m узла с репликой.

Тем самым, матрица $A^p = \|a_{nm}^p\|$ задает распределение p реплик каждого из M массивов. Элемент $a_{nm}^p \in \{0, 1\}$ этой матрицы равен “1”, если в узле n размещена реплика m -го массива.

Шаг 8. Для матрицы $A^p = \|a_{nm}^p\|$ выполняем следующие действия:

Определяем значение $F(A^p)$ в соответствии с (1).

Запоминаем полученное значение: $F_{min}^p = F(A^p)$.

Шаг 9. В цикле выбираем номера vm_k ($k = \overline{1, M}$) массивов из вектора VM .

Для каждого очередного m -го массива выполняются следующие операции:

Из множества \bar{X}_m «не протестированных» узлов случайным образом выбираем номер узла и запоминаем его в x_i^{m*} .

Если «не протестированных» узлов нет, то возвращаемся к шагу 9.

Шаг 9.1. В цикле каждую вершину x_{pi}^m из $X_p^m = \{x_{pi}^m | i = \overline{1, p}\}$ заменяем на вершину x_i^{m*} .

В результате получаем новое множество X_p^{m*} , в котором одна из вершин заменена на x_i^{m*} .

Копируем матрицу A^p в матрицу A^{p*} .

Так же, как и на шаге 7 присваиваем “1” элементам a_{nm}^{p*} m -го столбца матрицы A^{p*} в соответствии со значениями элементов из X_p^{m*} .

Подсчитываем значение функционала $F(A^{p*})$ для матрицы A^{p*} .

Если $F(A^{p*}) \geq F_{min}^p$ то возвращаемся к Шагу 9.1.

Для полученного нового распределения реплик A^{p*} выполняем проверку ограничений задачи (3) и (4).

Если хотя бы одно ограничение не выполнено, возвращаемся на Шаг 9.1.

Запоминаем текущее значение функционала и распределение реплик:

$$F_{min}^p = F(A^{p*}), A^p = A^{p*}$$

Если цикл по X_p^m не завершен, то возврат на Шаг 9.1.

Иначе возвращаемся на Шаг 9.

Шаг 10. Если просмотрены не все элементы вектора VM , то возврат на Шаг 6.

Шаг 11. Если цикл по p реплик не завершен ($p < P_{max}$), то возврат на Шаг 5.

Шаг 12. Поиск распределения реплик закончен.

Найдено распределение реплик A^p , близкое к оптимальному распределению, которое обеспечивает минимальные затраты на эксплуатацию системы, равные F_{min}^p .

Конец работы алгоритма.

Заключение

Задачи обеспечения высокого уровня сохранности информации в распределенных системах приобретают особую актуальность в случае систем, имеющих ненадежные каналы связи. Одним из эффективных методов решения данных задач является создание и оптимальное размещение в узлах системы реплик массивов данных. Данный метод широко используется на практике при создании распределенных информационных систем различного назначения и масштаба. В работе приведена формулировка задачи оптимального размещения реплик нескольких массивов данных в узлах распределенной системы, имеющей ненадежные каналы связи. В силу большой вычислительной сложности задач данного класса для решения сформулированной задачи предложен эвристический алгоритм. При этом предполагается, что в узлах системы с некоторой частотой выполняются прикладные процессы (задачи пользователей), каждый из которых генерирует множество запросов к репликам массивов данных. Генерируются запросы двух типов: информационные запросы (операции чтения данных реплик) и запросы на модификацию данных в репликах. Поиск оптимального распределения реплик выполняется с учетом характеристик генерируемых запросов в узлах сети запросов и затрат на передачу между узлами сети запросов и ответов на них.

Дальнейшее развитие исследований в рассматриваемой предметной области возможно в следующих перспективных направлениях:

Оптимальное распределение реплик в распределенных системах с динамически изменяющимся трафиком запросов в узлах системы.

Оптимальное распределение реплик в мобильных средах.

Применение методов кластеризации графов для сокращения размерности задач данного класса на сильно и слабо связанных графах большой размерности.

Итоги работ в данных направлениях позволят решать широкий круг теоретических и практических задач при проектировании и создании распределенных систем обработки данных, работающих надежно, с минимально возможными затратами, обеспечивающих высокий уровень сохранности информации и максимально возможной реакцией на возникающие запросы.

ЛИТЕРАТУРА

1. Сомов С.К. Сохранность информации в распределенных системах обработки данных / – М.: ИПУ РАН. – 2019. – 254 с.
2. Сомов С.К. Репликация данных как инструмент повышения надежности функционирования распределенных систем // журнал «Информационные технологии и вычислительные системы», 2018, – №3. – с. 69–79.
3. Микрин Е.А., Сомов С.К. Обзор моделей и методов обеспечения сохранности данных в распределенных системах обработки данных // журнал «Информационные технологии и вычислительные системы», 2017, – №4. – с. 5–28.
4. Чернышев Г.А. Обзор подходов к организации физического уровня в СУБД // Труды СПИИРАН. 2013. – Санкт-Петербург. 2013, – Вып. 1(24). – с. 222–275.
5. Tos U. and oth. Dynamic replication strategies in data grid systems: a survey. – The Journal of Supercomputing. 2015, – Volume 71, – Issue 11. – pp. 4116–4140.
6. Sourì A., Rahmani A. Survey for Replica Placement Techniques in Data Grid Environment // I.J. Modern Education and Computer Science, 2014, № 5. – pp. 46–51.
7. Chu W.W, File Allocation in a Multiple Computer System // IEEE Transactions on Computers, 1969, – Vol. C-18, – № 10. – p. 885–889.
8. Casey R.G. Allocations of copies of a file in an Information Network // AFIPS Conference Proceedings, 1972, – Vol. 40. – p. 617–625.
9. Eswaran K.P. Placement of records in a file and file allocation in a computer // IFIP Congress, 1974. – p. 304–307.
10. Ceri S., Navathe S., Wiederhold G. Distribution design of logical database schemas // IEEE Transactions on Software Engineering, 1983. – Vol. 9. – p. 487–504.
11. Apers P.M.G. Data allocation in distributed databases // ACM Transactions on Database Systems, 1988, – Vol. – 13, – №. 3. – pp. 263–304.
12. Кормен Т. и др. Алгоритмы: построение и анализ, 2-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 1296 с.
13. Kleinrock L. Queueing systems. Volume II: Computer applications // New York: Wiley, 1976. – 569 p.
14. Kleinrock L. Communication Nets: Stochastic Message Flow and Design // McGraw-Hill, 1964. – 220 p.
15. Hamdy A. Taha. Operations Research: An introduction. 8-th Edition // Pearson Prentice. Hall., 2007. – 838 p.

Somov Sergey Konstantinovich

V.A. Trapeznikov institute of control sciences of Russian academy of sciences, Moscow, Russia
ssomov2009@gmail.com

An example of the task of allocation the replicas of data in unreliable distributed systems and an algorithm of its solution

Abstract. The paper considers one of the methods for increasing the efficiency of distributed data processing systems – the optimal placement of replicas of the data arrays used in the system. This method allows ensuring a high level of information integrity and high reliability of the system, as well as minimizing the cost of its operation. This method is widely used in practice when creating distributed information systems for various purposes and scale.

The formulation of the problem of optimal allocation of replicas of several data arrays in nodes of a distributed system with unreliable communication channels is completed, and a heuristic algorithm for solving this problem is also given. As a criterion for optimization, the task uses the minimum cost of the system operating.

At the beginning of the paper, statistical data on the amount of losses incurred by companies in the event of downtime of information systems caused by negative factors of different nature are presented.

At the beginning of the work, statistics are presented on the amount of losses that companies incur in the event of a downtime of information systems caused by negative factors of various nature.

A brief review of the work on the topic of data placement in distributed systems is given.

Further, the paper presents a formal model of a distributed system with unreliable channels. At the nodes of this system, many application processes are performed that generate information requests to replicas of data arrays and requests to change data in replicas. The formulation of the problem of optimal placement of replicas in the nodes of a distributed system is given.

A list of steps and formulas for determining the average message delay in the communication channels of the system, taking into account the unreliability of communication channels, is given, as well as formulas for calculating the costs of operating a distributed system with replicas of data arrays located in its nodes.

Since the formulated problem has great computational complexity, a heuristic algorithm is presented in the work for finding the optimal placement of replicas of the set of data arrays used in the system.

In conclusion, several promising areas are listed for continuing research in the use of information redundancy in the form of replicas of data arrays to increase the efficiency of distributed systems.

Keywords: distributed systems; computer networks; data replication; replica placement; heuristic algorithm; minimum system maintenance costs; information integrity; system reliability