

Сомов С.К.

Адаптивный алгоритм размещения массивов данных в распределенных системах, имеющих динамические параметры

Аннотация: В работе предложен эвристический алгоритм размещения массивов данных в распределенных системах, параметры которых динамично изменяются. Цель работы алгоритма заключается в повышении эффективности работы и степени доступности ресурсов распределенных систем. Данная цель достигается за счет динамического перераспределения массивов данных, в зависимости от изменения трафика запросов к массивам в узлах сети. Предложен эвристический алгоритм решения данной задачи, работающий периодически или по запросу. Критерием оптимальности схемы перераспределения является минимум затрат на функционирование системы.

Ключевые слова: распределенные системы, распределение массивов данных, доступность данных

Эффективность работы распределенных систем во многом зависит от того, насколько оптимально размещены в узлах сети массивы данных, используемые системой. К сожалению, алгоритмы поиска оптимального размещения массивов по узлам сетей, особенно крупномасштабных и имеющих динамично изменяемые параметры работы, имеют большую вычислительную сложность.

В работе предлагается эвристический алгоритм для размещения массивов данных в распределенной системе, который может адаптироваться к меняющимся с течением времени параметрам системы. Используется формальная модель динамического не избыточного перераспределения массивов данных в узлах сети, учитывающего изменения трафика запросов к данным, генерируемого в узлах системы.

В модели предполагается, что топология сети определена и задано первоначальное размещение массивов по узлам сети. Алгоритм рассчитывает не избыточное перераспределение массивов с учетом изменений в трафике запросов в узлах сети. Критерием оптимальности перераспределения массивов служит минимум затрат на функционирование системы.

Распределенная система работает на базе компьютерной сети из M узлов: $N = \{N_1, \dots, N_M\}$. Каждому узлу сети сопоставлен максимальный объем памяти S_m и максимальное количество массивов NA_m , которые можно разместить в памяти узла. Система использует множество N массивов данных: $A = \{A_1, \dots, A_N\}$. Известен объем VA_n каждого массива. Распределение массивов в сети задано матрицей $X = \{x_{nm}\}$, где $x_{nm} = 1$, если массив A_n размещен в узле N_m , и равен 0 в противном случае. В каждом узле системы может выполняться от 1 до K разных процессов из множества $P = \{P_1, \dots, P_K\}$. Частота выполнения процессов в узлах задана матрицей $PFR = \{pfr_{km}\}$. Заданы матрицы $ENF = \{ENF_{kn}\}$ и $UPF = \{UPF_{kn}\}$ с частотами генерации в процессах запросов на извлечение и изменение данных.

Задана матрица $CC = \{C_{ij}\}$ стоимости C_{ij} передачи единицы данных между парами смежных узлов. На основе матрицы смежности сети и матрицы CC , с использованием любого известного алгоритма поиска кратчайших путей в графе, строится матрица $DTC = \{DTC_{ij}\}$, $i, j = \overline{1, M}$ затрат на передачу данных между узлами сети по кратчайшим путям.

В процессе работы алгоритма должны выполняться следующие ограничения.

Каждый узел m сети не должен хранить массивов больше, чем максимальное количество массивов, заданное для данного узла:

$$\sum_{n=1}^N x_{nm} \leq NA_m, \quad 1 \leq m \leq M$$

Распределение массивов X должно быть таково, что каждый массив A_n должен быть распределен в одном из узлов сети N_m :

$$\sum_{m=1}^M x_{nm} \geq 1, \quad 1 \leq n \leq N$$

Суммарный объем всех массивов, размещенных в узле сети, не должен превышать максимальный объем, разрешенный для этого узла:

$$\sum_{n=1}^N x_{nm} * VA_n \leq S_m, \quad 1 \leq m \leq M$$

В предлагаемом алгоритме используется понятие «вес массива данных» $DAW(N_j, A_i)$, значение которого равно сумме затрат на передачу всех запросов к массиву A_i , размещенному в узле N_j , и рассчитывается по формуле:

$$DAW(N_j, A_i) = \sum_{m=1}^M QF_{mji} * DTC_{mj}$$

Здесь:

$$QF_{mji} = \sum_{k=1}^K pfr_{km} * (ENF_{ki} + UPF_{ki})$$

Переменная QT_{mji} это общее количество запросов, возникающих в m -м узле сети к массиву A_i , размещенному в узле N_j .

Значение веса массива данных DAW используется в ситуации, когда в нескольких узлах генерируются запросы к одному массиву. В этом случае массив размещается в узле, имеющем наибольшее значение веса DAW .

Затраты на функционирование системы включают в себя стоимость обработки запросов и использования каналов связи.

Во время работы алгоритма вычисляются значения нескольких вспомогательных таблиц (например затраты на обработку запросов на модификацию данных, затраты на использование каналов связи). Данные этих таблиц используются алгоритмом для выбора наилучшего варианта перераспределения массивов в узлах сети.

Для каждого узла сети задан лимит на максимальное количество запросов, которые узел может обработать в единицу времени.

С некоторой заданной частотой в системе производится подсчет запросов, поступающих в узлы с размещенными в них массивами. Если в каких-то узлах этот лимит превышен, то для этих узлов выполняется описанный выше алгоритм и производится перераспределение размещенных в этих узлах массивов данных.

Данная работа подготовлена в рамках программы Президиума РАН № 30 (7) «Теория и технологии многоуровневого децентрализованного группового управления в условиях конфликта и кооперации».

Литература:

1. Микрин Е.А., Сомов С.К. Обзор моделей и методов обеспечения сохранности данных в распределенных системах обработки данных//Информационные технологии и вычислительные системы. – 2017. – № 4. – С. 5-28.

2. Кульба В.В., Сомов С.К. Проблема оптимального размещения файлов данных в крупномасштабных ненадежных распределенных системах / Материалы 10-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD'2017, Москва). Т. 2. – М.: ИПУ РАН, 2017. – С. 272-275.

Муромцев В.В., Муромцева А.В.

Проблемы коммуникаций в цифровом информационном пространстве

Аннотация: Рассматриваются проблемы цифровизации страны. Представлены основные проблемы коммуникаций, стоящие на пути цифровизации.

Ключевые слова: коммуникации, цифровизация, проблемы взаимодействия, человек-человек, человек-машина, машина-человек

Современный этап развития СЭС характеризуется активным и даже агрессивным процессом цифровизации общества. Этот процесс охватывает практически все сферы деятельности людей и, если информатизация в основном представляла собой создание информационной макроструктуры и не претендовала на обязательное её использование, то цифровизация является процессом обязательным для внедрения, без которого дальнейшее развитие общества невозможно, или, по крайней мере, существенно замедленно. Кроме того, само существование человека в цифровом обществе станет невозможным без знания цифровых технологий.

Тематика статьи связана с библейским сказанием о строителях Вавилонской башни. Как Вы помните, проблема заключалась в том, что строители с некоторого момента перестали понимать друг друга, и проект не был реализован.