
Разработка модуля оптимального размещения информационных ресурсов на узлах вычислительной сети: описание реализуемых методов и структур данных

В.А. Евсин, Н.А. Тихонов, С.П. Воробьев

*Южно-Российский государственный политехнический университет (НПИ)
имени М.И. Платова*

Аннотация: В данной статье рассматривается проблематика оптимального размещения информационных ресурсов на узлах вычислительной сети. Представлены основные методы, используемые в ходе решения данной проблемы. В частности, рассмотрен метод случайного размещения ресурсов, оптимизация размещения ресурсов с использованием метода ветвей и границ, а также оптимизация размещения ресурсов с использованием генетического алгоритма. Для данных методов определена структура входных и выходных данных, кроме того для представленных методов продемонстрирована внутренняя структура размещения ресурсов. Ключевым аспектом рассмотрения в данной статье является постановка задачи и моделирование процесса ее решения с использованием представленных алгоритмов. Как итог, представлено тестирование разработанного модуля на входных данных и анализ перспектив использования модуля.

Ключевые слова: вычислительная сеть, информационный ресурс, случайное размещение ресурсов, симплекс-метод, метод ветвей и границ, генетический алгоритм, проектирование, моделирование.

В ходе увеличения сложности и комплексности вычислительных сетей стала возникать необходимость в оптимальном размещении информационных ресурсов на узлах вычислительной сети, подробнее о которых можно прочесть в [1,2]. Одним из наиболее значимых показателей оптимального размещения ресурсов является стоимость размещения ресурса на сервере, другим важным аспектом оптимального размещения ресурсов является средняя стоимость доставки информационного ресурса до сервера, следовательно, задача оптимизации в общем виде может быть сформулирована следующим образом:

$$\sum_{i=1}^n \sum_{j=1}^m (TS_{ij} * K_1 + CS_i * VR_j * K_2) * B_{ij} \rightarrow \min, \quad (1)$$

где i – индекс сервера в кластере, n – количество серверов в кластере, $i = \overline{1, n}$,
 j – индекс информационного ресурса в массиве, m – количество

информационных ресурсов, $j = \overline{1, m}$, TS_{ij} – средняя стоимость доставки j -го информационного ресурса на i -й сервер, K_1 – коэффициент значимости стоимости доставки ресурса, CS_i – стоимость размещения единицы объема информационного ресурса на i -м сервере, VR_j – объем j -го информационного ресурса, K_2 – коэффициент значимости стоимости размещения ресурсов, B_{ij} – бинарное значение размещения информационного ресурса на сервере. Данная целевая функция должна выполняться при следующих ограничениях:

$$\left\{ \begin{array}{l} \sum_j VR_j * B_{ij} \leq VS_i, i = \overline{1, n} \\ B_{ij} = \begin{cases} 1, \text{ если } j\text{-й информационный ресурс отправляется на } i\text{-й сервер} \\ 0, \text{ иначе} \end{cases} \end{array} \right. \quad (2)$$

где VS_i – объем i -го сервера, остальные параметры представлены выше. Подробнее о математических моделях в кластерных вычислительных системах в [3,4]. Для решения поставленной задачи могут быть использованы следующие методы:

- метод случайного размещения: данный метод предполагает отсутствие оптимизации при размещении ресурса на случайном сервере с условием ограничения по объему сервера;

- метод ветвей и границ: данный метод предполагает использование симплекс метода для оптимизации размещения ресурсов, а также использование метода ветвей и границ для отсека нецелочисленных параметров решения;

- генетические алгоритмы: данный метод предполагает поиск решений путем кроссовера, а также мутации хромосом, содержащих параметры решения. Структура данных входного параметра для методов размещения ресурсов данной системы имеет следующую форму:

$$DTD = \langle LS, LIR, KCPS, KCSS \rangle,$$

где $LS = \{SERV_i, i = \overline{1..n}\}$ – массив серверов вычислительной сети, $LIR = \{LIR_j, j = \overline{1..m}\}$ – массив информационных ресурсов, $KCPS$ - коэффициент значимости стоимости размещения на сервере, $KCSS$ - коэффициент значимости стоимости отправления данных на сервер.

Структура данных i -го сервера имеет следующую форму:

$$LS_i = \langle MCAP_i, COST_i, USCAP_i, LIR_i \rangle,$$

где $MCAP_i$ – максимальный объем i -го сервера, $COST_i$ – стоимость размещения на i -м сервере, $USCAP_i$ – используемый объем i -го сервера, LIR_i – список информационных ресурсов, которые размещены на i -м сервере. Структура данных j -го информационного ресурса имеет следующую форму:

$$LIR_j = \langle RID_j, VOL_j, HTCOST_j \rangle,$$

где RID_j – идентификатор j -го ресурса, VOL_j – объем j -го ресурса, $HTCOST_j$ – таблица стоимостей отправки j -го ресурса до серверов. Структура результирующих данных после размещения ресурсов имеет следующую форму:

$$PLSERV = \langle CODE, MES, LS, LIR \rangle,$$

где $CODE$ – код возвращаемого значения, MES – сообщение об ошибке, LS – массив серверов, LIR – массив информационных ресурсов. Диаграмма классов рассмотренных структур представлена на рис.1.

В ходе решения методом случайного размещения необходимо провести сортировку стека серверов по их объему, после чего необходимо заполнять сервера случайным образом, итерации прекращаются, когда будут заполнены все сервера или распределены все ресурсы.

В ходе решения задачи методом ветвей и границ, о котором в [5,6], необходимо определить минимум функции (1) при заданных ограничениях (2), для решения необходимо использовать симплекс метод, в ходе которого

определяется оптимальное решение без учета ограничения по целочисленности, подробнее о данном методе в [7,8].

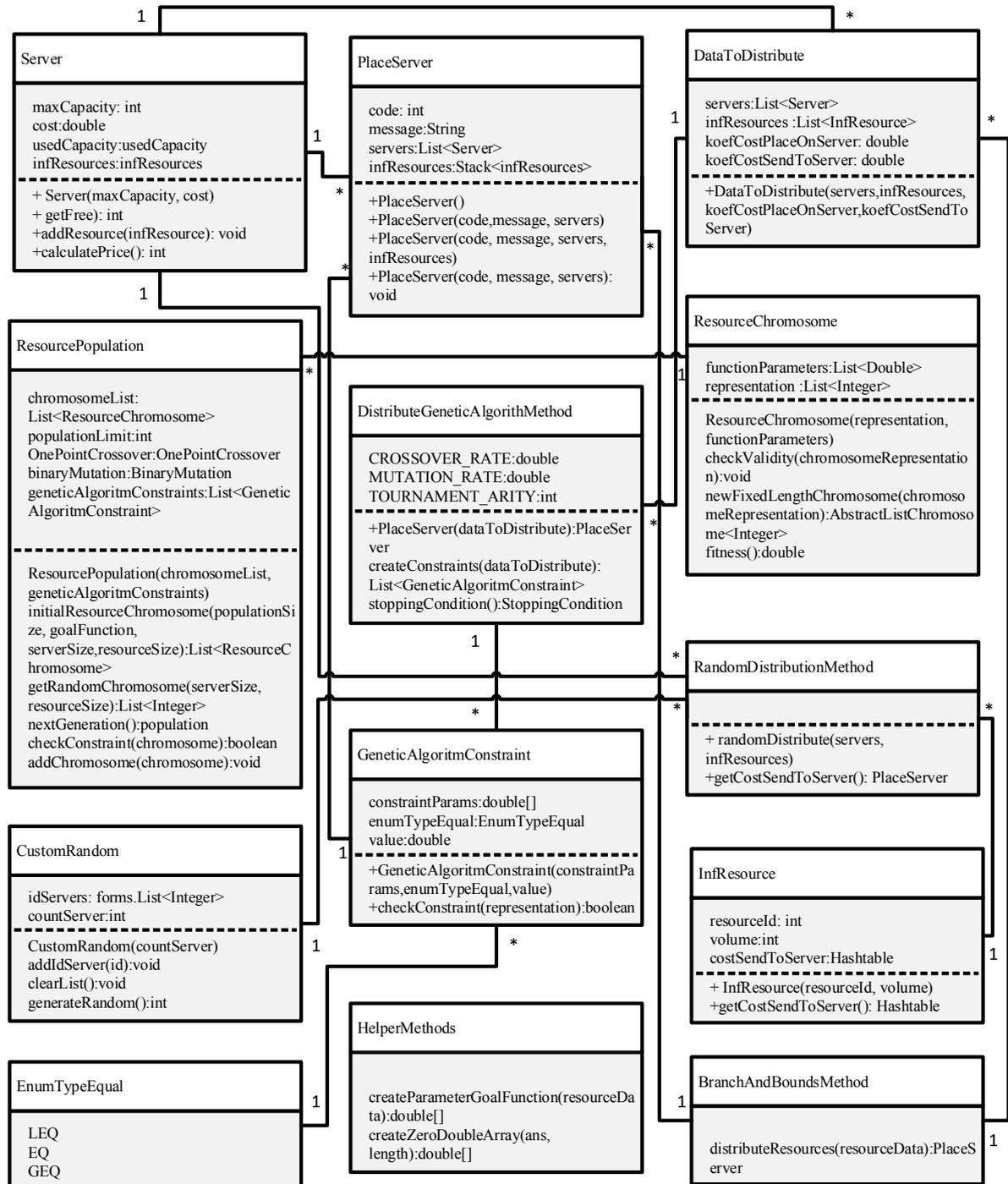


Рис. 1 – Диаграмма классов модуля оптимального размещения информационных ресурсов по узлам вычислительной сети

Для определения целочисленного решения необходимо для каждого нецелочисленного параметра разделить решение на 2 ветви:

- для первой ветви определить $x = 0$;
- для второй ветви определить $x \geq 1$, после чего необходимо произвести пересчет нижней границы решения и произвести выбор решения с наименьшей границей. В ходе итерационного пересчета параметров будет определено целочисленное решение.

Для решения генетическим алгоритмом, о котором можно прочесть в [9-11], необходимо определить начальную популяцию хромосом, которая бы удовлетворяла ограничениям (2). Хромосомы данной модели имеют следующую форму:

$$REPR_j = |b_i|, b_i = \begin{cases} 1, \text{ если } j\text{-й ресурс отправляется на } i\text{-й сервер} \\ 0, \text{ иначе} \end{cases}$$

Фитнес-функция данной модели имеет следующую форму:

$$F = \sum_j FP_j * REPR_j,$$

где FP_j – j -й параметр целевой функции, $REPR_j$ – значение генов j -й хромосомы. В качестве модели селекции используется модель элитарного отбора по критерию минимальной фитнес-функции хромосом. Для определения генов новых потомков используется модель точечного кроссовера. С вероятностью 30% у хромосом возникает мутация гена у рассматриваемой хромосомы. Каждая новая хромосома должна быть допустимой. Допустимость достигается путем выполнения критериев (2), в противном случае хромосома отбрасывается из общего набора. Каждая новая популяция имеет максимальный размер, равный 1.5 размера предыдущей популяции. Условие выхода из процесса селекции имеет следующую форму:

$$E = \begin{cases} 1, \text{ если } \frac{1}{m} \sum_j |F_j - \bar{F}| < \varepsilon \text{ или } ser \geq ter, \\ 0, \text{ иначе} \end{cases}$$

где F_j - наиболее приспособленная особь j -й эпохи, \bar{F} - усредненный показатель наиболее приспособленных особей за последние m -эпох, ε - максимальная ошибка модели, $сер$ - номер текущей эпохи, $тер$ - максимальное количество эпох.

Тестирование данных на 6 серверах и 12 информационных ресурсах генетическим алгоритмом показало результат, представленный на рис.2.

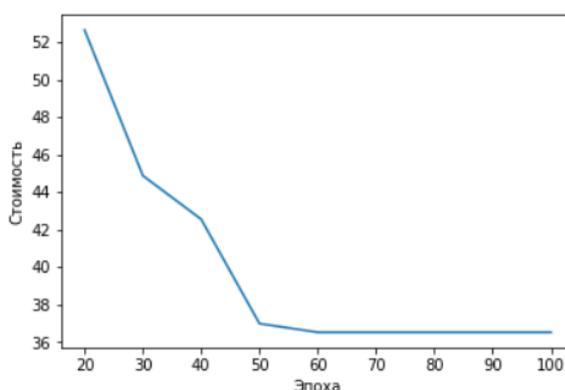


Рис. 2 – Динамика абсолютной стоимости

Дальнейшая разработка приложения будет направлена на разработку модуля моделирования среднего времени доступа к информационным ресурсам. Разработанное приложение может быть использовано предприятиями различных отраслей в целях оптимизации размещения ресурсов на узлах вычислительной сети.

Литература

1. Степанов А.Н. Архитектура вычислительных систем и компьютерных сетей. - СПб.: Питер, 2007. - 509 с.
2. Богданов А.В., Корхов В.В., Мареев В.В., Станкова Е.Н. Архитектуры и топологии многопроцессорных вычислительных систем. - М.: ИНТУИТ.РУ "Интернет-Университет Информационных технологий", 2004. - 172 с.
3. Евсин В.А., Литвяк Р.К. Математическая модель информационной системы высокой готовности на базе кластерной

архитектуры // Моделирование. Теория, методы и средства. - Новочеркасск: Лик, 2016. - С. 176-179.

4. Евсин В.А., Широбокова С.Н., Евсина В.А., Продан Е.А. Математическое моделирование распределенного реестра в сфере аренды недвижимости как сети массового обслуживания // Инженерный вестник Дона, 2018, №3 URL: ivdon.ru/ru/magazine/archive/n3y2018/5078

5. Черноморов Г.А. Теория принятия решений. - 3 изд. - Новочеркасск: Ред. журн. «Изв. вузов. Электромеханика», 2005. - 448 с.

6. Parmigiani G., Inoue L., Lopes, H. Decision Theory. Principles and Approaches. - New Jersey: John Wiley & Sons, 2009. – 405 p.

7. Банди Б. Основы линейного программирования. - М.: Радио и связь, 1989. - 176 с.

8. Васильев Ф.П. Методы оптимизации. - М.: МЦНМО, 2011. - 620с.

9. Haupt R.L., Haupt S.E. Practical Genetic Algorithms. - 2 edition. - New Jersey: John Wiley & Sons, 2004. – 272 p.

10. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. – М.: Горячая линия – Телеком, 2006. – 383 с.

11. Орловский Н.М. Решение задачи однокритериальной оптимизации процесса планирования действий экипажа Российского сегмента Международной космической станции на основе генетического алгоритма // Инженерный вестник Дона. 2013. №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1776

References

1. Stepanov A.N. Arkhitektura vychislitel'nykh sistem i komp'yuternykh setey [Architecture of computer systems and computer networks] SPb. Piter, 2007. 509 p.



2. Bogdanov A.V., Korhov V.V., Mareev V.V., Stankova E.N. Архитектуры и топологии многoproцессорных вычислительных систем [Architecture and topology of multiprocessor computing systems]. М.: INTUIT.RU "Internet-Universitet Informacionnyh tehnologij", 2004. 172 p.
3. Evsin V.A., Litvjak R.K. Modelirovanie. Teorija, metody i sredstva. Novoчеркасск: Lik, 2016. pp. 176-179.
4. Evsin V.A., Shirobokova S.N., Evsina V.A., Prodan E.A. Inženernyj vestnik Dona (Rus), 2018, №3 URL: ivdon.ru/ru/magazine/archive/n3y2018/5078.
5. Chernomorov G.A. Teorija prinjatija reshenij [Decision theory]. 3 izd. Novoчеркасск: Red. zhurn. «Izv. vuzov. Jelektromehanika», 2005. 448 p.
6. Parmigiani G., Inoue L., Lopes, H. Decision Theory. Principles and Approaches. New Jersey: John Wiley & Sons, 2009. 405 p.
7. Bandi B. Osnovy linejnogo programmirovaniya [Basics of linear programming]. М.: Radio i svjaz', 1989. 176 p.
8. Vasil'ev F.P. Metody optimizacii. [Optimization methods]. М.: MCNMO, 2011. 620 p.
9. Haupt R.L., Haupt S.E. Practical Genetic Algorithms. 2 edition. New Jersey: John Wiley & Sons, 2004. 272 p.
10. Rutkovskaja D., Pilin'skij M., Rutkovskij L. Nejronnye seti, geneticheskie algoritmy i nechetkie sistemy. [Neural networks, genetic algorithms and fuzzy systems] М.: Gorjachaja linija. Telekom, 2006. 383 p.
11. Orlovskij N.M. Inženernyj vestnik Dona (Rus), 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1776